Multi-time and multi-moment rating and portfolio models with nonparametric frontier methods : overview, new proposals and empirical studies

Tiantian Ren

湖南大學 工商管理學院
BUSINESS SCHOOL OF HUNAN UNIVERSITY

Université de Lille

Business School of Hunan University

Multi-Time and Multi-Moment Rating and Portfolio Models with Nonparametric Frontier Methods: Overview, New Proposals and Empirical Studies

Modèles de notation et de portefeuille multi-temporels et multi-moments avec des méthodes de frontière non paramétriques: Aperçu, nouvelles propositions et études empiriques

Une thèse présentée pour le grade de Docteur en Sciences Économiques

Tiantian Ren

Soutenue le 15 novembre 2021

Membres du Jury

Rapporteurs :
Prof. Barbara Casu, City University London
Prof. Ignace Van de Woestyne, KU Leuven

Examinateurs :
Prof. Stéphane Vigeant, Université de Lille (Président du jury)
Prof. Walter Briec, Université de Perpignan
Associate Prof. Bing Xu, Edinburgh Business School

Directeurs :
Prof. Kristiaan Kerstens, CNRS-LEM, IESEG School of Management
Prof. Zhongbao Zhou, Business School of Hunan University
Associate Prof. Paolo Mazza, CNRS-LEM, IESEG School of Management

Acknowledgements

It has been a wonderful and amazing experience in my life to pursue a PhD. I am so grateful to have met many great people and to have rediscovered myself on this journey. It would not have been possible to do without the support and guidance that I received from them.

My deepest gratitude goes first and foremost to my supervisors, Prof. Dr. Zhongbao Zhou, Prof. Dr. Kristiaan H.J. Kerstens and Prof. Dr. Paolo Mazza. Under the guidance of Prof. Zhou, I started a wonderful academic journey. During the six years from my master to PhD, as the supervisor and also my friend, Prof. Zhou has assisted with the practical and theoretical aspects of my research, which has been instrumental, and has also encouraged me, provided opportunities, promoted me and my work, and has fostered growth in me as an early career academic. I am extremely grateful for having worked with him during these formative years. Prof. Kerstens is a supervisor with an endless interest and passion for his research. His dynamism, vision and sincerity have always been contagious and motivational for me, even during tough times in the Ph.D. pursuit. He has taught me the methodology to carry out the research and to present the research works as clearly as possible. It is a great privilege and honor to work and study under his guidance. I am sincerely thankful for what he has helped me on my research topic and my research career. His constant trust has also given me a lot of confidence and energy to pursue my future research. I am grateful to my co-supervisor, Prof. Mazza, for his ongoing support and patient guidance. His rich research experience on financial topics and professional feedbacks have greatly helped me to improve

this thesis, as well as the related papers. Together, special thanks to my great supervisors. Their attitudes towards both academics and life have deeply influenced me.

My sincerest thanks also go to Prof. Dr. Barbara Casu, Prof. Dr. Ignace Van De Woestyne, Prof. Dr. Stéphane Vigeant, Prof. Dr. Walter Briec and Prof. Dr. Bing Xu for their time and effort invested in reading my work as the jury members of my PhD defense, and for their constructive comments that helped to improve this thesis.

I gratefully acknowledge the funding received towards my PhD study in France from Business School of Hunan University. I am also grateful to IESEG School of Management for the excellent working conditions to realize my Ph.D thesis. Meanwhile, my sincere thanks go to my university of inscription, l'Université de Lille and l'Ecole Doctorale SESAM.

I would like to say thanks to my friends and research colleagues on both sides of Business School of Hunan University and IESEG School of Management. At the Business School of Hunan University, I appreciated the group of people who share the same interests: I have appreciated meeting the group of lovely friends at Business School of Hunan University (some are from other universities in China). Helu Xiao, Ling Lin, Yong Jiang, Cenjie Liu, Yanfei Bai, Qianying Jin, Ximei Zeng, Qing Liu, Shuanglin Zhou, Dongran Su and Xiaoqing Chen. They have granted me a great deal of accompany and encouragement during the whole journey of my PhD study. I also appreciate the great help from every colleague I met at IESEG, especially my PhD colleagues and the colleagues in Building C for the research experience and perspectives we shared with each other. I express my special thanks to Prof. Petitjean for his support and help on the empirical fund data used for this thesis.

Lastly, I would like to express my deepest gratitude to my parents, Yanbin Ren and Li Shi, for their selfless trust and devotion. I do appreciate that they have given me enough courage to pursue my career and supported my decision. Further, one of the best outcomes from these past six years is finding my best friend and boyfriend, Helu Xiao. There are no words to convey how much I love him. These past several years have not been an easy ride, both academically and personally. I truly thank him for sticking by my side, even when I was irritable and depressed. I feel that we have learned a lot from each other and have been getting better, which has strengthened our commitment and determination to each other and to live life to the fullest.

Tiantian Ren

Lille, 4 August 2021

General Abstract

The multidimensional evaluation criteria with multiple moments and multiple times explored in this thesis are of importance for finance to handle mixed risk-aversion preferences of investors which aim at time persistence. This thesis first establishes novel nonparametric frontier methods to rate funds that can simultaneously handle multiple moments and multiple times (Chapter 2). It further proposes a new discrete time indicator to trace the performance of funds with respect to the ever changing nonparametric frontiers (Chapter 3). Finally, it develops the nonparametric matafrontier rating methods to assess the performance of funds across classifications accounting for heterogeneity (Chapter 4). In the empirical study, this thesis designs a buy-and-hold backtesting strategy to compare the potential benefits of the existing and newly proposed rating methods to select the best performing funds.

The performance measure for fund rating that can handle both multiple moments and multiple times has been unavailable in the existing literature. This motivates our work in Chapter 2 where a new performance measure in the multi-time and multi-moment rating framework is first defined. This performance measure is capable of not only assessing to which extent a fund performs well in the several moments following mixed risk-aversion preferences, but it simultaneously measures to which extent a fund performs well in all these moments in different times as well. Then, a series of novel nonparametric frontier rating models are proposed, and are further empirically applied to hedge funds. Finally, we define a simple buy-and-hold backtesting strategy to test for the impact of multiple moments and

multiple times separately and jointly.

Chapter 3 develops a Luenberger portfolio productivity indicator (LPPI) and its decomposition to measure the evolution of fund performance in the multi-time and multi-moment framework. This LPPI captures the overall features of changes in fund performance over time, whose decomposition allows for identifying whether these changes are due to the efficiency change or to the frontier shift. The efficiency change component provides a pragmatic measurement to identify the contribution of fund managers to the performance evolution, while the frontier change component measures the local changes in the frontier movements induced by market volatility. Using the backtesting analysis, we empirically test the usefulness of the new LPPI (and its components) for rating and selecting funds.

Chapter 4 mainly responds to a practical need for rating funds across groups and develops a general procedure to assess the performance of heterogeneous portfolios based on the metatechnology. We define a metatechnology efficiency (MTE) by combining the shortage function with the nonparametric metafrontier, which allows to compare the performance of heterogeneous funds from distinct groups directly. Then, we extend this metafrontier rating procedure to the multi-time and multi-moment framework, which is compatible with the general mixed risk-aversion preferences of investors. Finally, the empirical part employs a large database of funds not only to offer extensive tests of the specification issues surrounding the application of these metafrontier models, but also to illustrate the performance of metafrontier models in fund selection.

Keywords: Shortage Function; Nonparametric Frontier; Fund Rating; Multiple Times; Multiple Moments

Résumé Général

Modèles de notation et de portefeuille multi-temporels et multi-moments avec des méthodes de frontière non paramétriques : Aperçu, nouvelles propositions et études empiriques

Les critères d'évaluation multidimensionnels avec des moments multiples et des temps multiples explorés dans cette thèse sont importants pour la finance afin de traiter les préférences mixtes d'aversion au risque des investisseurs qui visent la persistance temporelle. Cette thèse établit d'abord de nouvelles méthodes de frontière non paramétriques pour évaluer les fonds qui peuvent simultanément gérer des moments multiples et des temps multiples (Chapitre 2). Elle propose ensuite un nouvel indicateur en temps discret pour suivre la performance des fonds par rapport aux frontières non paramétriques en constante évolution (Chapitre 3). Enfin, elle développe les méthodes de notation matafrontières non paramétriques pour évaluer la performance des fonds à travers les classifications en tenant compte de l'hétérogénéité (Chapitre 4). Dans l'étude empirique, cette thèse conçoit une stratégie de backtesting buy-and-hold pour comparer les avantages potentiels des méthodes de notation existantes et nouvellement proposées pour sélectionner les fonds les plus performants.

La mesure de performance pour la notation des fonds qui peut gérer à la fois des

moments multiples et des temps multiples n'est pas disponible dans la littérature existante. Cela motive notre travail dans le chapitre 2, où une nouvelle mesure de performance dans le cadre de la notation multitemporelle et multi-moment est d'abord définie. Cette mesure de performance est capable non seulement d'évaluer dans quelle mesure un fonds est performant dans les différents moments suivant des préférences mixtes d'aversion au risque, mais aussi de mesurer simultanément dans quelle mesure un fonds est performant dans tous ces moments à différents moments.Ensuite, une série de nouveaux modèles non paramétriques d'évaluation de la frontière sont proposés, et sont ensuite appliqués empiriquement aux hedge funds. Enfin, nous définissons une simple stratégie de backtesting buy-and-hold pour tester l'impact des moments multiples et des périodes multiples séparément et conjointement.

Le chapitre 3 développe un indicateur de productivité de portefeuille de Luenberger (LPPI) et sa décomposition pour mesurer l'évolution de la performance des fonds dans le cadre multitemporel et multi-moment. Ce LPPI capture les caractéristiques globales des changements de performance des fonds dans le temps, dont la décomposition permet d'identifier si ces changements sont dus au changement d'efficience ou au déplacement de la frontière. La composante de changement d'efficacité fournit une mesure pragmatique pour identifier la contribution des gestionnaires de fonds à l'évolution de la performance, tandis que la composante de changement de frontière mesure les changements locaux dans les mouvements de la frontière induits par la volatilité du marché. En utilisant l'analyse backtesting, nous testons empiriquement l'utilité du nouveau LPPI (et de ses composantes) pour la notation et la sélection des fonds.

Le chapitre 4 répond principalement à un besoin pratique de notation des fonds entre groupes et développe une procédure générale d'évaluation de la performance

de portefeuilles hétérogènes basée sur la métatechnologie. Nous définissons une efficacité métatechnologique (MTE) en combinant la fonction de pénurie avec le métafrontier non paramétrique, qui permet de comparer directement la performance de fonds hétérogènes issus de groupes distincts. Ensuite, nous étendons cette procédure de notation métafrontière au cadre multitemporel et multi-moment, ce qui est compatible avec les préférences mixtes générales d'aversion au risque des investisseurs. Enfin, la partie empirique utilise une large base de données de fonds non seulement pour offrir des tests approfondis des questions de spécification entourant l'application de ces modèles métafrontières, mais aussi pour illustrer la performance des modèles métafrontières dans la sélection des fonds.

Mots clés : Fonction de Pénurie ; Frontière Nonparamétrique ; Notation des Fonds ; Temps multiples ; Moments Multiples

Contents

List of Figures

List of Tables

CHAPTER 1

General Introduction

A fund is a professionally managed and pooled investment vehicle, which allows individuals and institutions to combine smaller amounts of capital into a larger sum for investment. With the development of global financial markets, the number and diversity of funds (e.g., hedge funds, equity funds, bond funds, etc.) available to investors are steadily increasing. Typical investors rely heavily on the risk-adjusted return (i.e., "performance") measures to identify the funds worthwhile investing in among these numerous available funds, and in this regard, they are increasingly concerned with the ratings and/or rankings of funds determined by explicit performance measures. Clearly, an effective fund evaluation/rating can provide not only strategic support for investors' fund screening, but also investment benchmarking for fund managers to improve the performance of their managed portfolios.

Since the seminal work of Markowitz (1952), it has been recognized that portfolio efficiency should be measured with a trade-off between portfolio return and risk, mainly based on the mean-variance (MV) portfolio optimization problem of maximizing returns and minimizing risk. In a static context, he defines an efficient frontier of portfolios whose expected return cannot improve unless one is willing to assume more risk, i.e., a Pareto-optimal subset of portfolios. However, maintaining

strong assumptions on probability distributions and Von Neumann-Morgenstern utility functions make his work suffer many theoretical difficulties. As stated by numerous authors, the MV evaluation framework is not consistent with the Von Neumann–Morgenstern axioms of expected utility theory unless either (i) asset returns follow normal probability distributions, or (ii) utility functions representing investor preferences are quadratic. Starting from at least Mandelbrot (1963), empirical studies have repeatedly shown that the distributions of the asset return are widely non-normal (see Campbell, Lo, and McKinlay (1997)). Meanwhile, the broad class of mixed risk-aversion utility functions with a general preference for odd moments and an aversion to even moments is nowadays rather widely acknowledged (see Trautmann and van de Kuilen (2018) for an empirical review).

Probably in view of widespread criticisms of the MV bi-criteria approach, several authors have introduced alternative evaluation criteria based on the multidimensional performance of portfolios. One of the crucial concerns is that multiple moments are included in the portfolio evaluation, for example, the classic moments such as return and variance, and higher-order moments such as skewness and kurtosis. From a theoretical point of view, the evaluation criteria with multiple moments somewhat relaxes the strong assumptions of the traditional MV portfolio optimization regarding the return distribution of assets or the investor preferences. From a practical point of view, the use of multiple moments allows for gauging the performance of portfolios compatible with general investor preferences (see Briec and Kerstens (2010)). Therefore, enlarging the portfolio performance analysis from the MV framework to the multi-moment framework is of interest to financial analysts and individual investors.

Another important concern is to account for the time dimension when establishing a portfolio evaluation criteria. In practice, a typical investor attempting to

select a fund for investment, has available various funds' performances over several time periods: e.g., the past year to date; 1 year; 3 years; 5 years and 10 years. The question now is how one can integrate all the performance information over different time periods and develop an overall multi-time performance mesasure to recognize funds with persistently positive performance.

Clearly, some traditional financial measures (e.g., Sharpe ratio, Sortino ratio, etc.) partly or mostly derived from the Markowitz framework share the same drawbacks as the MV bi-criteria for funds rating, and they are of little use of beyond two-dimensional problems (see Caporin, Jannin, Lisi, and Maillet (2014), for a review). In recent years, the frontier-based methods used for assessing the performance of funds have gained some popularity, especially for addressing multi-dimensional evaluation problems. With the help of efficiency measures from production theory, these frontier methodologies measure the efficiency of a fund by estimating the distance between this observed portfolio and its reference portfolio on a portfolio frontier (or extremum estimators of this frontier). The use of frontier or extremum estimators allows for rating funds along a multitude of dimensions instead of using just some combination of two dimensions as in most traditional financial performance measures. To our knowledge, Sengupta (1989) is the first to introduce an efficiency measure borrowed from production theory in relation to the MV portfolio frontier. This efficiency measure relates to the distance function that for a long time has been employed in consumer theory and especially in production theory (e.g., Cornes (1992)). The current application of frontier-based methods to handle multiple moments or multiple times in portfolio evaluation could be summarized as follows.

First, for handling the multiple moments, Joro and Na (2006) propose a radial measure under input (i.e., variance) orientation in a mean–variance-skewness

(MVS) framework, where the distance between the variance of the projection on the efficient frontier and the portfolio variance is used as the performance measure to gauge this portfolio. This oriented-radial efficiency measure ignores the possibility that the investor looks for simultaneously increasing the output while reducing the input level. Following Briec, Kerstens, and Lesourd (2004) who establish the duality between a distance function and MV utility functions, Briec, Kerstens, and Jokung (2007) use a general distance function (named shortage function) to look for improvements in efficiency in MVS space by looking for simultaneous expansions in mean return and positive skewness and reductions in risk. Even more general, for the class of mixed risk-aversion utility functions, Briec and Kerstens (2010) assess portfolio performance for the general moments case by simultaneously looking for improvements in odd moments and reductions in even moments. Examples of related studies include, e.g., Bacmann and Benedetti (2009), Branda (2013), Branda and Kopa (2014), Khemchandani and Chandra (2014), Massol and Banal-Estañol (2014), Branda (2015), Jurczenko, Maillet, and Merlin (2006), Jurczenko and Yanou (2010), Boudt, Cornilly, and Verdonck (2020), Adam and Branda (2020), among others.

The above portfolio evaluation models are based on the diversified portfolio frontier, and thus are also referred to as diversified portfolio models in the literature. These diversified models require nonlinear programming in most cases, and their potential complexity and computational burden make these models suffer from the dilemma of being unsuitable for large-scale evaluations. This dilemma may be especially evident when higher-order moments are included. Probably in view of solution complexity of diversified portfolio models, nonparametric production frontiers are transposed into the financial literature in an effort to offer alternative fund ratings, which are called nonparametric (production) frontier rating methods. In this context, the diversified portfolio frontier is approximately

estimated by nonparametric production frontier specifications, and the performance of each fund is measured relative to these nonparametric estimators of the portfolio frontier (see Basso and Funari (2016) for a review). Kerstens, Mounir, and Van de Woestyne (2011) argue that funds can be interpreted as fee-based financial products whose price distribution is characterized by some combination of moments by analogy to the characteristics approach of consumer theory. This argument is of great relevance in providing theoretical supports for the application of nonparametric frontier methods in funds rating. Moreover, the authors systematically discuss a variety of specification issues surrounding the application of nonparametric frontier-based methods when assessing the financial performance of funds. Compared to the diversified portfolio models, these nonparametric frontier rating methods are commonly solved by the linear (or binary mixed integer) programming. A great development on the application of nonparametric frontier-based methodology for handling multiple moments in fund ratings, e.g., Matallín-Sáez, Soler-Domínguez, and Tortosa-Ausina (2014), Brandouy, Kerstens, and Van de Woestyne (2015), Nalpas, Simar, and Vanhems (2017), Mehlawat, Kumar, Yadav, and Chen (2018), Krüger (2021), Lin and Li (2020), Gong, Yu, Min, and Ge (2021).

Second, for handling multiple times, Morey and Morey (1999) is the first to establish a multiple time period assessment within a standard MV framework. Their method establishes respectively two efficiency measures under risk and return orientations from a multi-period perspective for MV portfolio performance appraisals. The first efficiency measure aims at contracting all risk dimensions without affecting any return dimensions, and the other efficiency measure aims at augmenting all return dimensions with no increase in any risk dimensions. Briec and Kerstens (2009) develop a multi-period MV portfolio analysis by means of a temporal shortage function (see Briec and Kerstens (2009)). Instead of either

proportionally contracting risk dimensions or proportionally expanding return dimensions, this temporal shortage function simultaneously attempts to reduce the risk and to expand the return over all times. These multi-period evaluation methods allow for the difference in risk-return preferences of investors across time and can reveal more useful information on fund selection compared to gauging the fund's performance over a single time (see Ren, Zhou, and Xiao (2021)). To the best of our knowledge, these multi-time/period assessments are based on the diversified MV portfolio models, while none of the current research has conducted a multi-time portfolio performance analysis using the nonparametric production frontier methods.

Having the importance of multiple moments and multiple times for portfolio evaluation in mind, the fundamental objective of our work in **Chapter 2** is to develop a series of nonparametric convex and nonconvex frontier rating methods that can simultaneously handle both multiple moments and multiple times by extending the shortage function. These newly proposed methods aim not only to evaluate to which extent a fund performs well in the several moments following mixed risk-aversion preferences, but they simultaneously are assessing to which extent a fund performs well in all these moments over different times. This is important given the concern in the financial literature that traditional performance measures may exhibit limited stability over time (e.g., Bodson, Coen, and Hubner (2008), Menardi and Lisi (2012) and Grau-Carles, Doncel, and Sainz (2019), among others). In the empirical analysis of Chapter 2, these rating methods are applied to actual hedge fund data (HF), since this category of funds is known to be subject to non-normal return distributions. We define a simple buy-and-hold backtesting strategy to test for the impact of multiple moments and multiple times separately and jointly on the funds rating and selection.

The performance measures proposed in Chapter 2 fall into the category of static performance evaluation, which focuses on measuring a fund's efficiency by a cross-sectional comparison among all observed funds using the return information over a finite horizon. In recent years, differing from this static analysis, some studies are devoted to measure the evolution of fund performance over time by a cross-period comparison to investigate the attribution problem for the efficiency changes (see Babalos, Caporale, and Philippas (2012) Brandouy, Briec, Kerstens, and Van de Woestyne (2010), and Gregoriou and Moschella (2017), among others). This can be regarded as a dynamic evaluation for a fund performance in discrete time. Therefore, the primary interest of **Chapter 3** is to introduce a novel method for measuring the evolution of fund performance in the multi-time and multi-moment framework using some extensions of the shortage function combined with a Luenberger productivity indicator (Chambers, Färe, and Grosskopf (1996)). This new Luenberger portfolio productivity indicator and its components (i.e., the efficiency change component and the frontier change component) are constructed to measure fund's dynamic performance with the consideration of multiple time periods and multiple moments, and to distinguish whether performance changes over time are due to the efficiency change or due to the shift in the frontier.

As introduced in Brandouy, Briec, Kerstens, and Van de Woestyne (2010), the Luenberger portfolio productivity indicator captures essential features of the whole return distribution for funds, and positive values in this indicator reflect that the fund performance increases over time. With regard to its two components, the efficiency change component in principle detects the eventual ability of fund managers for stock picking and market timing, and the positive values in this component capture the positive contribution of fund managers to the performance evolution. The frontier change component measures the local changes in the frontier movements induced by market volatility, and the positive values in this

component indicate the improvements of financial market performance. Therefore, in the empirical part of Chapter 3, we conduct the buy-and-hold backtesting analysis for each of the three alternative selection scenarios: (i) HFs are selected depending on the ranking of Luenberger portfolio productivity indicator; (ii) HFs are selected depending on the ranking of Luenberger efficiency change component; (iii) HFs are selected depending on the ranking of Luenberger frontier change component. To the best of our knowledge, this work of Chapter 3 is the first to develop a proposal on the methodology and application of the Luenberger portfolio productivity indicator with multi-times and multi-moments.

In the existing frontier-based rating literature, homogeneity of the observed funds is a common assumption made in the application of nonparametric frontier methods to assess the performance of funds. Strictly speaking, this homogeneity assumption requires that the performance of a fund is evaluated relative to the peer group, and therefore the efficiency measured against the group-specific frontier is regarded as the group-specific efficiency. Clearly, this group-specific efficiency serves a limited usefulness for typical investors who aim to select best performing funds across groups. In addition, given a practical fact that the ratings provided by the rating agencies available to investors are also based on the comparison of funds within the same group, a cross-group fund rating methodology accounting for heterogeneity is of interest to both academics and industry.

This motivates the work of **Chapter 4**. The basic objective of Chapter 4 is to tackle the problem of assessing heterogeneous funds belonging to different groups by borrowing recent developments in the nonparametric metafrontier (see Afsharian and Podinovski (2018), Kerstens, O'Donnell, and Van de Woestyne (2019) and Jin, Kerstens, and Van de Woestyne (2020), among others). Without limiting ourselves to a standard MV portfolio setting, we further develop the nonparamet-

ric metafrontier rating methods with multiple times and multiple moments, which allows for making the fund evaluation in line with general investor preferences. The empirical part employs a large database of heterogeneous funds covering five different groups (types) not only to offer extensive tests of the specification factors considered in these models, but also to test the performance of different rating methods in a simple backtesting setup.

In summary, the overarching objective of this thesis is to establish a series of novel nonparametric frontier-based rating methods in the multi-time and multi-moment framework for handling mixed risk-aversion preferences of investors which aim at time persistence. From a theoretical point of view, the first aim is to define a new shortage function or performance measure that can simultaneously handle both multiple moments and multiple times. To the best of our knowledge, this basic idea is new and unavailable in the current literature. All these existing nonparametric frontier studies are single time, and this contribution is the first to develop a multi-time framework with multiple moments following general preferences of investors. The second idea is by incorporating this new performance measure with the production indicator to construct a new portfolio productivity indicator that can trace the evolution of fund performance in the multi-time and multi-moment framework. This proposed Luenberger portfolio production indicator and its decomposition provide a useful measurement tool to recognize the attribution of fund performance changes. The last objective is to introduce metafrontier technology into funds rating to design a general metafrontier-based evaluation procedure that allows for gauging the performances of funds across groups. From an empirical point of view, the actual data of funds is employed to offer extensive tests of these proposed methods, as well as other existing funds rating methods. Furthermore, the proposed models are all simply computed by the linear (or binary mixed integer) programming, which are much easier and time-saving compared to

diversified portfolio models when applied in the large-scale and multi-dimensional fund diversified ratings.

The structure of this thesis is as follows. **Chapter 2** proposes a series of new frontier models to rate funds that can simultaneously handle multiple moments and multiple times. Based on these new performance measures, **Chapter 3** offers a novel method for measuring the evolution of fund performance in the multi-time and multi-moment framework by combining the productivity index theory. **Chapter 4** is devoted to develop a general procedure to rate funds across classifications accounting for heterogeneity by borrowing recent developments on metatechnology. In the empirical analysis of the three chapters, a simple buy-and-hold backtesting strategy is defined to test for the impact of multiple moments and multiple times separately and jointly on the ratings and selection of funds. Finally, conclusions and issues for future work are summarized in **Chapter 5**.

References

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

Afsharian, M., and V. Podinovski (2018): "A Linear Programming Approach to Efficiency Evaluation in Nonconvex Metatechnologies," *European Journal of Operational Research*, 268(1), 268–280.

Babalos, V., G. M. Caporale, and N. Philippas (2012): "Efficiency Evaluation of Greek Equity Funds," *Research in International Business and Finance*, 26(2), 317–333.

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

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

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

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

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

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

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

Brandouy, O., W. Briec, K. Kerstens, and I. Van de Woestyne (2010): "Portfolio Performance Gauging in Discrete Time Using a Luenberger Productivity Indicator," *Journal of Banking & Finance*, 34(8), 1899–1910.

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

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

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

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

Briec, W., K. Kerstens, and J. Lesourd (2004): "Single Period Markowitz Portfolio Selection, Performance Gauging and Duality: A Variation on the Lu-

enberger Shortage Function," *Journal of Optimization Theory and Applications*, 120(1), 1–27.

Campbell, J., A. Lo, and A. McKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

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

Chambers, R. G., R. Färe, and S. Grosskopf (1996): "Productivity Growth in APEC countries," *Pacific Economic Review*, 1(3), 181–190.

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

Gong, X., C. Yu, L. Min, and Z. Ge (2021): "Regret Theory-based Fuzzy Multi-objective Portfolio Selection Model Involving DEA Cross-efficiency and Higher Moments," *Applied Soft Computing*, 100, 106958.

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

Gregoriou, G. N., and J. Moschella (2017): "The Efficiency and Productivity of Hedge Fund Families," *Journal of Wealth Management*, 20(3), 101–112.

Jin, Q., K. Kerstens, and I. Van de Woestyne (2020): "Metafrontier Productivity Indices: Questioning the Common Convexification Strategy," *European Journal of Operational Research*, 283(2), 737–747.

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

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

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

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

Kerstens, K., C. O'Donnell, and I. Van de Woestyne (2019): "Metatechnology Frontier and Convexity: A Restatement," *European Journal of Operational Research*, 275(2), 780–792.

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

Krüger, J. (2021): "Nonparametric Portfolio Efficiency Measurement with Higher Moments," *Empirical Economics*, 61(3), 1435–1459.

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

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36(4), 394–419.

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

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

Matallín-Sáez, J., A. Soler-Domínguez, and E. Tortosa-Ausina (2014): "On the Informativeness of Persistence for Evaluating Mutual Fund Performance using Partial Frontiers," *Omega*, 42(1), 47–64.

Mehlawat, M. K., A. Kumar, S. Yadav, and W. Chen (2018): "Data Envelopment Analysis based Fuzzy Multi-objective Portfolio Selection Model Involving Higher Moments," *Information sciences*, 460, 128–150.

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

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

Nalpas, N., L. Simar, and A. Vanhems (2017): "Portfolio Selection in a Multi-Moment Setting: A Simple Monte-Carlo-FDH Algorithm.," *European Journal of Operational Research*, 263(1), 308–320.

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

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

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

CHAPTER

2

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

Abstract

This contribution introduces new frontier models to rate mutual funds that can simultaneously handle multiple moments and multiple times. These new models are empirically applied to hedge fund data, since this category of funds is known to be subject to non-normal return distributions. We define a simple buy-and-hold backtesting strategy to test for the impact of multiple moments and multiple times separately and jointly. The empirical evidence points to a remarkable superior performance of the proposed frontier methods compared to traditional financial performance measures and single-time MV frontier methods.

KEYWORDS: Shortage function; Frontier; Fund rating.

2.1 Introduction

The foundational work of Markowitz (1952) in modern portfolio theory has learned every investor that to gauge the performance of portfolio management one must consider risk in addition to return. This mean-variance (MV) dual objective of maximizing returns and minimizing risks turns performance evaluation into a controversial task involving trade-offs related to the risk preferences of the investor. The two-dimensional nature of this nonlinear quadratic optimization problem allows to display the efficient frontier as a Pareto-optimal subset of portfolios whereby the expected return can only increase when also the variance increases.

A large part of modern portfolio theory continues developing variations on these bi-objective MV optimization problems. A wide offer of alternative risk measures is available in the portfolio literature: entropy, expected shortfall, mean absolute deviation, semi-variance and other partial moment measures, Value-at-Risk (VaR) in all its variations, etc. (see, e.g., Bacon (2008) and Feibel (2003) for surveys).[1]

This focus on the first two moments of a random variable's distribution is only consistent with the von Neumann-Morgenstern axioms of choice underlying expected utility (EU) theory when: (i) asset processes follow normal distributions, or (ii) investors have quadratic utility functions. A substantial empirical literature has documented that normality of asset returns can be rejected for a variety of financial asset classes in both developed and emerging financial markets (e.g., Jondeau and Rockinger (2003)). At least since Scott and Horvath (1980), investors have been attributed a positive preference for skewness as well as a negative preference for kurtosis to explain financial behavior. Meanwhile, decision-theoretic arguments

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

exist for what has become known as the broad class of mixed risk-aversion utility functions that are characterized by a preference for odd moments and an aversion for even moments (see Eeckhoudt and Schlesinger (2006)). Furthermore, via surveys and experiments traditional risk preferences like risk aversion, but also higher order risk preferences like prudence and temperance are nowadays better understood (see Trautmann and van de Kuilen (2018) for a review).

Over time, several alternative portfolio selection criteria based on preferences for higher-order moments have been developed. But, so far not a single widely accepted criterion seems to have emerged. It is possible to distinguish between primal and dual approaches to determine such higher-order moments portfolio frontiers. One example of the primal approach is found in Lai (1991) who determines mean-variance-skewness (MVS) optimal portfolios via a Polynomial Goal Programming procedure. The dual approach necessitates a specification of some indirect higher-moment utility function and yields optimal portfolios via its parameters reflecting higher-moment preferences (e.g., Harvey, Liechty, Liechty, and Müller (2010)).

To our knowledge, Sengupta (1989) is the first to introduce an efficiency measure -borrowed from production theory- into a diversified MV portfolio model. This efficiency measure relates to the distance function that for a long time has been employed in consumer theory and especially in production theory (e.g., Cornes (1992)). In consumer theory the distance function is dual to the expenditure function: it serves to characterize multiple commodity and single utility choice sets.[2] In production theory the input distance function is dual to the cost function: it basically serves to characterize multiple input multiple output production possibility sets (e.g., Hackman (2008)). This has opened up a booming research field

[2]This distance function has sometimes been employed to make welfare comparisons (e.g., Slesnick (1998)). More recently, Briec, Dumas, and Mekki (2021) stress that the directional distance function is dually linked to the weighted and indirect Rawlsian social welfare functions.

where parametric but particularly nonparametric specifications of production and dual (e.g., cost) frontiers are specified based on minimal maintained axioms (e.g., constant or variable returns to scale, convexity or not, etc.). Applied to a plethora of private and public sectors, these frontier methodologies analyse technical, scale or cost efficiency, economies of specialization, mergers, etc. (e.g., Färe, Grosskopf, and Lovell (1994)).[3]

The introduction of an efficiency measure into portfolio theory allows to gauge performance over multiple dimensions and it opens up new perspectives. On the one hand, following Briec, Kerstens, and Lesourd (2004) who establish duality between a distance function and MV utility functions, Briec, Kerstens, and Jokung (2007) use a general distance function (named shortage function) to look for improvements in efficiency in MVS space by looking for simultaneous expansions in mean return and positive skewness and reductions in risk. Furthermore, these authors provide a duality result with a MVS utility function.[4] Even more general, for the class of mixed risk-aversion utility functions, Briec and Kerstens (2010) assess portfolio performance for the general moments case by simultaneously looking for improvements in odd moments and reductions in even moments. In addition, these authors establish duality with general moment utility functions.

Empirical applications of this diversified multi-moment approach are found in Adam and Branda (2020), Branda (2013), Branda and Kopa (2014), Branda (2015), Joro and Na (2006), Jurczenko, Maillet, and Merlin (2006), Khemchandani and Chandra (2014), Krüger (2021), Massol and Banal-Estañol (2014), among others. Furthermore, Bacmann and Benedetti (2009), Boudt, Cornilly, and Verdonck

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

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

(2020), and Jurczenko and Yanou (2010), among others, are empirical diversified multi-moment contributions focusing on hedge funds (HF).

On the other hand, within a standard MV framework, Morey and Morey (1999) develop a multiple time horizon assessment: in particular, these authors use either a risk contraction or a return expansion efficiency measure to evaluate MV performance over three time horizons simultaneously (in particular, a 3, 5 and a 10-year time period). This contribution is slightly generalized in Briec and Kerstens (2009).[5] An empirical application is available in Ren, Zhou, and Xiao (2021).

To the best of our knowledge, Murthi, Choi, and Desai (1997) is the seminal article that has been rating mutual funds (MF) by simultaneously trying to maximize the return and minimizing standard deviation, expense ratio, load, and turnover using a nonparametric production frontier specification that maintains convexity and constant returns to scale. Following Farrell (1957) and Charnes, Cooper, and Rhodes (1978), nonparametric production frontiers are transposed into the financial literature in an effort to provide alternative fund ratings. Intuitively, nonparametric production frontiers can envelop the observations of any multi-dimensional choice set and position each of the observations relative to the boundary of the choice set using some efficiency measure. This has led to a growing literature that has been applied to a large variety of financial assets (e.g., exchange traded funds, hedge funds, pension funds, etc.). Furthermore, a wide variety of model specifications are available in terms of some combination of ordinary moments, lower and/and upper partial moments, as well as in terms of production frontier specifications (constant or variable returns to scale, etc.), and the choice

[5]Note that the use of multiple time horizons within a MV framework is not particularly computationally challenging, but moving from a quadratic convex MV problem to a cubic nonconvex MVS portfolio optimization problem is computationally harder. Evidently, the same remark applies when one moves from a cubic nonconvex MVS to a quartic nonconvex mean-variance-skewness-kurtosis portfolio optimization problem, or beyond by including even higher order moments.

of efficiency measure (e.g., reducing variables for which less is better (like inputs), or expanding variables for which more is better (like outputs), or some combination of both). This frontier-based MF rating literature has been rather recently surveyed in Basso and Funari (2016).

Following Heffernan (1990) and Blake (1996), among others, Kerstens, Mounir, and Van de Woestyne (2011) interpret this funds rating literature as a transposition of the characteristics approach in consumer theory into finance: MF are seen as fee-based financial products characterized by distributional characteristics of the asset price distribution as summarized by some combination of moments. Compared to the diversified portfolio models that require nonlinear programming, these nonparametric production frontier MF rating models can normally be solved using simple linear programming.

An open question is how the diversified portfolio models relate to the nonparametric production frontier specifications? Recently, Liu, Zhou, Liu, and Xiao (2015) state that a convex variable returns to scale nonparametric production frontier specification provides an inner approximation to the traditional MV diversified portfolio model. This is certainly correct. One basic idea implicit in their contribution is that nonparametric production frontier specifications should ideally underestimate the eventual performance of a diversified portfolio model. In the more general case where we want to explore a nonconvex diversified MV (e.g., with some integer constraints) or a nonconvex higher moment portfolio model, then one can argue that the nonconvex nonparametric production frontier specification with variable returns to scale already advocated by Kerstens, Mounir, and Van de Woestyne (2011) provides a conservative underestimation of the corresponding nonconvex diversified portfolio model within some common subspace of moments (see also Germain, Nalpas, and Vanhems (2011)). By contrast, the more widely

used convex nonparametric production frontier specification may overestimate the corresponding nonconvex diversified portfolio model within the common subspace of moments. The latter argument seems to have escaped attention so far: this explains why most nonparametric production frontier MF rating models with higher moments do impose convexity (for instance, Gregoriou, Sedzro, and Zhu (2005)).

The use of distance functions or efficiency measures in both the diversified portfolio models and the nonparametric production frontier specifications leads to the question how these gauges relate to traditional financial performance measures (see, e.g., the surveys in Bacon (2008), Feibel (2003) and Caporin, Jannin, Lisi, and Maillet (2014)). While relative performance measures that are variations on returns per unit of risk (e.g., Sharpe ratio) are useful to handle bi-objective (e.g., MV) optimization problems, they are of little use beyond two dimensional problems. E.g., adding a skewness constraint to a MV diversified model weakly decreases return and weakly increases variance inevitably yielding a weakly worse Sharpe ratio: hence, the Sharpe ratio cannot assess higher moment portfolios. If finance wants to handle mixed risk-aversion preferences of investors, then it must consider a multidimensional performance measure. Some performance measures try to assess the tail risk, like VaR or the Conditional Value-at-Risk (CVaR), but they most of the time focus on the risk component and do not include the first moment of the return distribution. One exception is the Omega ratio that we include in our analysis. Caporin, Jannin, Lisi, and Maillet (2014) classify the distance (shortage) function approach correctly among the absolute performance measures: these performance measures are based on rewards when compared to those of a reference portfolio on a portfolio frontier. The choice for distance (shortage) function brings finance and portfolio analysis in line with consumer and production analysis where these micro-economic tools have a proven track record in representing

multidimensional choice sets.[6]

The major objective of this contribution is to define new distance functions or efficiency measures that can simultaneously handle both multiple moments and multiple times (instead of either multiple moments or multiple times separately). This performance measure offers a simple and powerful tool for assessing MF performance based on the moment characteristics over all time periods. To the best of our knowledge, this basic idea is new and unavailable in the literature. This performance measure thus aims not only to evaluate to which extent a MF performs well in the several moments following mixed risk-aversion preferences, but it simultaneously is assessing to which extent a MF performs well in all these moments over different times. This is important given the concern in the financial literature that traditional performance measures may exhibit limited stability over time (e.g., Bodson, Coen, and Hubner (2008), Menardi and Lisi (2012) and Grau-Carles, Doncel, and Sainz (2019), among others).

This new performance measure is applied to HFs, a fund accessible only to institutional investors and high net worth individuals. Among MFs, HFs have a unique compensation structure. The most widespread fee structure is the so-called 2/20, i.e., 2% of assets under management for annual management fees and 20% of any profits made as a performance incentive fee. Consequently, HFs are marked by their heterogeneity and unusual (i.e., non-normal) statistical properties, as compared to more traditional MFs. Indeed, HFs tend to exhibit some more strongly asymmetric and fat tailed return characteristics compared to other MFs (see Gregoriou (2003), Darolles and Gourieroux (2010), Eling and Faust (2010), among others, and especially El Kalak, Azevedo, and Hudson (2016) for a survey).

[6]Tammer and Zălinescu (2010) show that the shortage function is linked to the scalarization function that is used in vector optimization problems, of which multi-objective optimization problem is a special case.

They are globally viewed as riskier but are also associated with higher rewards. This is why our empirical study specifically focuses on HFs since these are most likely to be affected by higher order moments.

The traditional financial performance measures (e.g., Sharpe ratio, Sortino ratio, etc.) used for HF rating have been subject to some criticism, because they basically follow the theoretical assumptions of the Capital Asset Pricing Model (CAPM) that the capital market is efficient and financial asset returns are normally, independently and identically distributed, among others. When asset returns do not obey the normal distribution, then the mean and variance no longer suffice to effectively summarize its return distribution. Given the complexities to assess the performance of HFs using traditional performance measures (e.g., see Smith (2017)), we think that our new performance measure may provide a suitable framework to evaluate both persistence across moments and across times.

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

The remainder of this contribution is organized as follows. The next Section 2.2 introduces the nonparametric production frontiers that serve to approximate the diversified portfolio models: we first discuss single-time multi-moment models, then we introduce the new multi-time multi-moment models. In Section 2.3, we develop the buy-and-hold backtesting strategy in detail. Section 2.4 describes

the hedge fund data in detail and comments upon the empirical results. Finally, Section 2.5 concludes.

2.2 Nonparametric Frontier Rating Models: Methodology

2.2.1 Single-Time and Multi-Moment Rating Framework

The nonparametric frontier rating methods gauge the financial performance of MF, and these evaluations are done mostly using frontier-based models which originate from production theory. In this section, we only introduce the basic definitions and properties needed for applications within finance. Assume that there are n MFs under evaluation over a given time horizon. At time t in this time horizon, the j-th MF ($j \in \{1, \ldots, n\}$) is characterized by m input-like values x_{ij}^t ($i \in \{1, \ldots, m\}$) and s output-like values y_{rj}^t ($r \in \{1, \ldots, s\}$). Input-like variables need to be minimized and output-like variables need to be maximized.

We introduce one widely used production frontier-based model with variable returns to scale (VRS). Following Briec, Kerstens, and Vanden Eeckaut (2004), a unified algebraic representation of convex and nonconvex production possibility sets (PPS) under the VRS assumption for a sample of n MFs at time t is:

$$P_\Lambda^t = \Big\{(x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s \mid \forall i \in \{1, \ldots, m\} : x_i^t \geq \sum_{j=1}^{n} \lambda_j x_{ij}^t, \\ \forall r \in \{1, \ldots, s\} : y_r^t \leq \sum_{j=1}^{n} \lambda_j y_{rj}^t, \lambda \in \Lambda\Big\}, \qquad (2.2.1)$$

where:

$\Lambda \equiv \Lambda^C = \{\lambda \in \mathbb{R}^n \mid \sum_{j=1}^n \lambda_j = 1 \text{ and } \forall j \in \{1,\ldots,n\} : \lambda_j \geq 0\}$ if convexity is assumed, and
$\Lambda \equiv \Lambda^{NC} = \{\lambda \in \mathbb{R}^n \mid \sum_{j=1}^n \lambda_j = 1 \text{ and } \forall j \in \{1,\ldots,n\} : \lambda_j \in \{0,1\}\}$ if nonconvexity is assumed.

At time t, if there exists an input-output combination $(\sum_{j=1}^n \lambda_j x_{ij}^t, \sum_{j=1}^n \lambda_j y_{ij}^t)$ in the convex or nonconvex PPS using less inputs and producing more outputs than the observed MF, then this MF is considered inefficient since it can improve its inputs and/or outputs. MFs are efficient if no improved input-output combinations can be found. The input-output combinations of these efficient MFs are all located at the boundary of P_Λ^t which is called the convex or nonconvex VRS nonparametric frontier.

Using the nonparametric PPS defined in (2.2.1), the shortage function of any observed MF at time t is now defined as follows:

Definition 2.2.1. *At time t, let $g^t = (-g_x^t, g_y^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ and $g^t \neq 0$. For any observation $z^t = (x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s$, the shortage function S_Λ^t at time t in the direction of vector g^t is defined as:*

$$S_\Lambda^t(z^t; g^t) = \sup\{\beta \in \mathbb{R} \mid z^t + \beta g^t \in P_\Lambda^t\}.$$

This shortage function simultaneously permits the enhancement of output-like variables and the reduction of input-like variables. If the shortage function value $S_\Lambda^t(z_o^t; g_o^t) > 0$ for the input-output combination $z_o^t = (x_o^t, y_o^t)$ of a specific MF at time t, then z_o^t is not located on the frontier of P_Λ^t. Hence, its inputs and/or outputs can be improved to catch up with the VRS nonparametric frontier. By contrast, if the shortage function value $S_\Lambda^t(z_o^t; g_o^t) = 0$, then z_o^t is located on the frontier.

Consider a MF with index $o \in \{1, \ldots, n\}$ in need of assessment at time t by means of the shortage function with direction vector $g_o^t = (-g_{xo}^t, g_{yo}^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$. Combining (2.2.1) and Definition 2.2.1, the efficiency status of this MF at time t can be determined by solving the following program:

$$
\begin{aligned}
\max \quad & \beta \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij}^t \leq x_{io}^t - \beta g_{io}^t, \quad i = 1, \ldots, m, \\
& \sum_{j=1}^{n} \lambda_j y_{rj}^t \geq y_{ro}^t + \beta g_{ro}^t, \quad r = 1, \ldots, s, \\
& \sum_{j=1}^{n} \lambda_j = 1, \quad \beta \geq 0, \\
& \forall j = 1, \ldots, n : \begin{cases} \lambda_j \geq 0, & \text{under convexity,} \\ \lambda_j \in \{0, 1\}, & \text{under nonconvexity.} \end{cases}
\end{aligned}
\tag{2.2.2}
$$

Note that Model (2.2.2) results in a linear programming (LP) problem under convexity and a mixed binary integer programming (MBIP) problem under nonconvexity.[7]

The setting defined in the previous section is general and flexible and can thus handle a large choice of inputs and outputs. We now particularize the above formulation to characterize the efficient frontier in the MVS and the mean-variance-skewness-kurtosis (MVSK) spaces. Suppose that there are n MFs under evaluation. At time t, let $R_1^t, \ldots, R_n^t$ denote the raw returns of the n funds, which are characterized by their expected return $E(R_j^t)$, variance $V(R_j^t)$, skewness $S(R_j^t)$ and kurtosis $K(R_j^t)$ for $j \in \{1, \ldots, n\}$.[8] Here, the calculations of vari-

[7]The direction vector of the shortage function used in the nonparametric frontier-based methods is set as $g^t = (-|x_{1o}^t|, \ldots, -|x_{mo}^t|, |y_{1o}^t|, \ldots, |y_{so}^t|)$, whereby all input-like values x_{io}^t, $(i = 1, \ldots, m)$ and output-like values y_{ro}^t, $(r = 1, \ldots, s)$ are simultaneously increased and decreased in proportion to their initial values respectively.

[8]In the application of single-time MVSK frontier models, we choose the historical returns of

ance, skewness and kurtosis are expressed as follows: $V(R_j^t) = E[(R_j^t - E(R_j^t))^2]$, $S(R_j^t) = E[(R_j^t - E(R_j^t))^3]$, and $K(R_j^t) = E[(R_j^t - E(R_j^t))^4]$.[9] To obtain a detailed specification of the PPS, as defined in (2.2.1), we need to classify the different goals of the investor in terms of either inputs (i.e., objectives to minimize), or outputs (i.e., objectives to maximize). As discussed in the previous section, the need for multiple moments is apparent to assess MFs (and most particularly HFs) whose return distributions may exhibit strong asymmetry and fat tails. Given mixed risk-aversion utility functions, investors express a preference for odd moments and a dislike for even moments of the distribution of asset returns. Therefore, when the MVSK framework is considered, we can define the first and second inputs of MFs as $x_{1j}^t = V(R_j^t)$ and $x_{2j}^t = K(R_j^t)$, and the first and second outputs as $y_{1j}^t = E(R_j^t)$ and $y_{2j}^t = S(R_j^t)$ for $j \in \{1, \ldots, n\}$. Obviously, for the MVS case only the first input is considered.

For a MF o under evaluation at time t, denote $E_o = E(R_o^t)$, $V_o = V(R_o^t)$, $S_o = S(R_o^t)$ and $K_o = K(R_o^t)$. Then both models, either with convexity or nonconvexity, allow to project the input-output combination (V_o, K_o, E_o, S_o) of this MF in such a way that inputs (i.e., variance and kurtosis) are decreased and outputs (i.e., expected return and skewness) are increased in the direction g_o^t. The optimal solution β^* of model (2.2.2) measures how many times the direction vector g_o^t fits in the line segment from the input-output combination of the MF o to the efficient frontier in the direction of g_o^t.

In model (2.2.2) under convexity, the left-hand sides of the constraints are all linear. All possible linear combinations of inputs and outputs of the observed

the last 5 years prior to time t to estimate the first four moments for the fund under evaluation.

[9]Note that the higher moments included in the proposed models are the 3rd and 4th moments of the returns distribution, rather than the normalized versions in finance(obtained by dividing by the standard deviation to the powers 3 and 4, respectively.). Krüger (2021) provide a thorough discussion regarding the association between the 3rd and 4th moments of the returns distribution with the expected utility for a risk-averse investor.

MFs are used to construct a convex VRS frontier for evaluation. For the MF o, if $\beta^* = 0$, the corresponding input-output combination is on the convex frontier and efficient at time t. If $\beta^* > 0$, there exist input-output combinations yielding a higher or equal return and skewness together with a lower or equal variance and kurtosis. When nonconvexity is assumed in model (2.2.2), evaluation is done with respect to a nonconvex VRS frontier determined by all efficient MFs (excluding the convex input-output combinations of these).

2.2.2 Multi-Time and Multi-Moment Rating Framework

Differing from MF ratings in a single-time framework, MF ratings in a multi-time framework consider performance over a time horizon consisting of multiple discrete time periods. To develop the nonparametric frontier rating models in this multi-time framework, some definitions and properties are presented.

Consider n MFs under evaluation. Let T denote the number of consecutive times in a time horizon of interest. In addition, define a multi-time path of inputs and outputs as $Z_j = (x_j^t, y_j^t)_{t=1}^T$ for MF j, $(j = 1, \ldots, n)$, where $x_j^t = (x_{1j}^t, \ldots, x_{mj}^t)$ and $y_j^t = (y_{1j}^t, \ldots, y_{sj}^t)$ represent m inputs and s outputs at time t, respectively. Assuming VRS for all times $t \in \{1, \ldots, T\}$ and strong free disposability of all inputs and outputs, the multi-time PPS with convexity and nonconvexity can be defined as:

$$\mathbf{P}_\Lambda^T = P_\Lambda^1 \times \cdots \times P_\Lambda^T \subset (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m \times T} \times \mathbb{R}^{s \times T}, \tag{2.2.3}$$

where P_Λ^t, $(t = 1, \ldots, T)$, is the PPS at time t mentioned previously in (2.2.1).

The idea is now for each MF to simultaneously expand its multiple outputs

and decrease its multiple inputs over all discrete times in a given time horizon by means of the multi-time shortage function. To allow a general definition, we first introduce some abbreviating notations.

The time dependent direction vector denoted by $G = (g^1, \ldots, g^T) \in (\mathbb{R}^m_- \times \mathbb{R}^s_+)^T \cong \mathbb{R}^{m\times T}_- \times \mathbb{R}^{s\times T}_+$ represents a given multi-time direction path, where $g^t = (-g^t_x, g^t_y) \in \mathbb{R}^m_- \times \mathbb{R}^s_+$ represents the direction vector at time $t \in \{1, \ldots, T\}$. In addition, we denote $\Theta = (\beta_1, \ldots, \beta_T) \in \mathbb{R}^T$ and $\Theta \cdot G = (\beta_1 g^1, \ldots, \beta_T g^T) \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$. Considering the time preference of an investor in a portfolio context, we introduce a time discounting factor denoted ξ $(0 < \xi < 1)$ to weight the efficiency measures over the time horizon.[10] Then, the time discounted multi-time shortage function assuming convexity or nonconvexity is defined as follows:

Definition 2.2.2. *With the notations introduced above, for any observation $Z \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$, the time discounted multi-time shortage function S^T_Λ in the direction of G is defined as:*

$$S^T_\Lambda(Z; G) = \sup\left\{\frac{1}{T}\sum_{t=1}^{T} \xi^{T-t}\beta_t \mid Z + \Theta \cdot G \in \boldsymbol{P}^T_\Lambda\right\}.$$

For a given time horizon T, this amounts to looking for the largest arithmetic mean of time discounted distances over all times in a given time horizon of the input-output combinations of an observed MF to boundary of $\mathbf{P}^T_\Lambda$. This definition adapts a weighted (discounted) temporal efficiency measure, whereby the weights decline as one moves away from the present into the past.[11] If the time discounted

[10]Given that efficiency gains in the far future should be weighted less than efficiency gains in the near future when planning ahead for an economic agent, the time discounting factor ξ is assumed to remain $0 < \xi < 1$.

[11]For retrospective benchmarking based on observed past behavior when assessing performance, the distant past is less valuable than the nearby present (as indicated by Briec and Kerstens

multi-time shortage function value $S_\Lambda^T(Z;G) > 0$ for the input-output path Z of the MF being evaluated, then it means that its inputs and outputs can be reduced and improved simultaneously in one or more time periods.

Based on Definition 2.2.2, we are now in the position to determine the non-parametric frontier rating models in a general formulation. Suppose there are n MFs under evaluation. Let T denote the number of consecutive times in a time horizon under consideration. In particular, the multi-time rating methods used in Section 2.3 focus on 3 distinct time periods: 1, 3 and 5 years. For a given multi-time direction path $G = (g^t)_{t=1}^T \in \mathbb{R}_-^{m\times T} \times \mathbb{R}_+^{s\times T}$, the efficiency of the MF o under evaluation can be determined by the time discounted multi-time shortage function value resulting from the following program:

$$
\begin{aligned}
\max \quad & \frac{1}{T}\sum_{t=1}^{T} \xi^{T-t}\beta_t \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j^t x_{ij}^t \le x_{io}^t - \beta_t g_{io}^t, \quad i = 1,\dots,m, \quad t = 1,\dots,T, \\
& \sum_{j=1}^{n} \lambda_j^t y_{rj}^t \ge y_{ro}^t + \beta_t g_{ro}^t, \quad r = 1,\dots,s, \quad t = 1,\dots,T, \\
& \sum_{j=1}^{n} \lambda_j^t = 1, \quad \beta_t \ge 0, \quad t = 1,\dots,T, \\
& \forall j = 1,\dots,n : \begin{cases} \lambda_j^t \ge 0, \quad t = 1,\dots,T, & \text{under convexity,} \\ \lambda_j^t \in \{0,1\}, \quad t = 1,\dots,T, & \text{under nonconvexity.} \end{cases}
\end{aligned}
\tag{2.2.4}
$$

In the multi-time framework, we select variance and kurtosis of each time t, $(t = 1,\dots,T)$, as inputs and expected return and skewness as outputs, whereas for the MVS case only variance for each t is considered as inputs. With the help of the time discounted multi-time shortage function, the observed MF with index o can improve its multiple return and skewness dimensions and reduce its multiple

(2009)). In that sense, the distant past contributes less weight to efficiency gains than the nearby past.

variance and kurtosis dimensions along a given direction path G over all time periods. The value of the objective function of model (2.2.4) indicates the amount of (in)efficiency of the MF o representing the multi-time shortage function. A value greater than zero indicates that the inputs and outputs of the evaluated MF can be improved in one or more time periods, and thus is inefficient from a multi-time perspective.

In the following Sections 2.4 and 2.5, we employ MF data to compare the proposed multi-time and multi-moment measures with traditional financial measures, as well as with single-time MV measures. These comparisons are aimed not only to illustrate the impact of multiple moments and multiple times on MF performance evaluation, but more importantly to further explore the potential benefits of the newly proposed performance measures for MF selection by means of backtesting. We now turn to explain the backtesting framework in Section 2.3.

2.3 Backtesting Framework

Our main objective in this contribution is to test that the multi-time and multi-moment performance measures can be expected to perform well for MF ratings and selection. To this end, a comparative approach based on a backtesting methodology is adopted. Backtesting refers to executing fictitious investment strategies using historical data to simulate how these strategies would have performed if they had actually been adopted by MF managers in the past. It is powerful for evaluating and comparing the performance of different investment strategies without using real capital. Some examples of a backtesting approach are found in DeMiguel, Garlappi, and Uppal (2009), Tu and Zhou (2011), Brandouy, Kerstens, and Van de Woestyne (2015), Zhou, Xiao, Jin, and Liu (2018) and Lin and Li (2020), among

others.

For comparison, there are 15 fund rating methods in total being collected in our work. On the one hand, we test some traditional financial indicators: Sharpe ratio, Sortino ratio and Omega ratio. The exact definition for the Sharpe, Sortino and Omega ratios can be found in Feibel (2003, p. 187 and p. 200) and Eling and Schuhmacher (2007, p. 2635), respectively. Based on these definitions, these three traditional financial ratios are presented as follows:

$$\text{Sharpe} = \frac{E(R) - r_f}{\sigma(R)}, \tag{2.3.1}$$

$$\text{Sortino} = \frac{E(R) - r_f}{\sigma_{-}(R)}, \tag{2.3.2}$$

$$\text{Omega} = \frac{E(R) - L}{E[\max(L - R, 0)]} + 1, \tag{2.3.3}$$

where $E(R)$ and r_f represent the mean value of a raw return R and the risk-free rate, respectively; $\sigma(R)$ and $\sigma_{-}(R)$ denote the standard and lower semi-standard deviations of a raw return R, respectively; L is the loss threshold, in particular, above this threshold returns are considered gains, while below this treshold these are regarded as losses. Using the above three ratios, we obtain the financial indexes for the above n MFs (i.e., Sharpe_j, Sortino_j and Omega_j, where $j = 1, \ldots, n$) which can be used to measure their performance at the given time horizon T, and the higher the value, the better the performance. The risk-free rate r_f and the loss threshold L are here specified as zero.[12]

Furthermore, in line with the properties of the shortage function used in the nonparametric frontier-based methods, we define the following traditional finance-

[12]As our dataset consists in the raw returns (not excess returns over a risk-free rate), we consider setting the risk-free benchmark as zero to calculate the financial ratios here, which ensures consistency with the nonparametric frontier models that assume jointness and null jointness between risk and return. In most finance literature, the risk-free rate could be set as the bank term deposit rate or Treasury bill rate.

based efficiency measures that bound the values between zero and unity and that make sure that the zero indicates full efficiency:

$$\text{Eff}(\text{Sharpe}_j) = \frac{\max\{\text{Sharpe}_j \mid j = 1, \ldots, n\} - \text{Sharpe}_j}{\max\{\text{Sharpe}_j \mid j = 1, \ldots, n\} - \min\{\text{Sharpe}_j \mid j = 1, \ldots, n\}}, \quad (2.3.4)$$

$$\text{Eff}(\text{Sortino}_j) = \frac{\max\{\text{Sortino}_j \mid j = 1, \ldots, n\} - \text{Sortino}_j}{\max\{\text{Sortino}_j \mid j = 1, \ldots, n\} - \min\{\text{Sortino}_j \mid j = 1, \ldots, n\}}, \quad (2.3.5)$$

$$\text{Eff}(\text{Omega}_j) = \frac{\max\{\text{Omega}_j \mid j = 1, \ldots, n\} - \text{Omega}_j}{\max\{\text{Omega}_j \mid j = 1, \ldots, n\} - \min\{\text{Omega}_j \mid j = 1, \ldots, n\}}. \quad (2.3.6)$$

On the other hand, we include convex and nonconvex nonparametric frontier-based ratings in different frameworks. All these 15 rating methods (3 traditional financial rating methods plus 12 frontier-based rating methods) are listed in Table 2.1.

Table 2.1: List of various rating models compared: 3 financial rating methods and 12 frontier rating methods

Classification	Methods
Traditional financial measures	Eff(Sharpe)
	Eff(Sortino)
	Eff(Omega)
Convex frontier rating methods	Single-time and MV framework
	Single-time and MVS framework
	Single-time and MVSK framework
	Multi-time and MV framework
	Multi-time and MVS framework
	Multi-time and MVSK framework
Nonconvex frontier rating methods	Single-time and MV framework
	Single-time and MVS framework
	Single-time and MVSK framework
	Multi-time and MV framework
	Multi-time and MVS framework
	Multi-time and MVSK framework

To simplify names of the frontier-based methods, some notation indicates which frontier rating method is used for ranking MFs. This can be done in both single-time (ST) and multiple-time (MT) frameworks, using a convex (subscript 'c') or a non-convex (subscript 'nc') frontier rating methods, and focusing on the first two (MV), three (MVS), or four moments (MVSK), respectively. For instance, MTMVSKc refers to the convex frontier model with the mean, variance, skewness and kurtosis over multiple times. Note that all the empirical results concerning these 15 rating methods are reported using these simplified notations.

We consider a simple buy-and-hold backtesting strategy consisting of buying in each time the 10, 20 and 30 best performing MFs ranked by rating method, respectively.[13] Our work now is to empirically test the out-of-sample performance of these 15 buy-and-hold strategies. Since the Sharpe ratio and other relative performance measures are only suitable for the MV world, we opt for the shortage function as an absolute performance measure that is capable to assess the performance of these strategies in multiple dimensions simultaneously (i.e., mean, variance, skewness and kurtosis). Hence, the 15 buy-and-hold backtesting strategies are compared based on the MVSK performance of their holding values evaluated by the single-time and multi-moment rating models (with convexity and nonconvexity).

Based on the fundamental logic of backtesting summarized so far, we design a backtesting analysis in detail for the buy-and-hold strategies constructed by the 15 rating methods. Our backtesting analysis is performed multiple times by rolling the time window. We first collect a sample of HFs with monthly return data starting from October 2006 till October 2020.[14] The detailed description of this

[13]Strictly speaking, the buy-and-hold backtesting framework designed in this contribution is a process of selecting and holding best-performing funds based on their rankings.

[14]Compared to rebalancing strategies, the buy-and-hold strategies designed in this contribution are likely to be less affected by periods of the extreme events because they has a relatively long

sample funds is presented in the following section (Section 2.4). Then, we split the period from the beginning of the sample period to the end of October 2015 in time windows of a given length, where the 5 years before the end of the sample period are kept apart to test the long-term holding performance of these strategies in the last backtesting period. Since the longest time period considered in our work is 5 years, it is appropriate to set the length of the rolling time window at 5 years. Therefore, the backtesting analysis is developed starting from November 2011, and is repeated 48 times (each time another month) with the rolling time window of 5 years till October 2015.

Using the first 5 year time window of data (from November 2006 to October 2011) to obtain the rankings for different rating methods, we determine the first buy-and-hold backtesting strategies in November 2011. These strategies are held for four holding scenarios: the end of October 2012 (for 1 year); the end of October 2014 (for 3 years); the end of October 2016 (for 5 years); and until the end of October 2020 (the end of the whole sample period). The process of the first backtesting is represented in Figure 2.1.[15]

holding period and remains unchanged after the funds are selected.

[15]The four moments of the fund's return distribution are estimated based on a 5-year estimation window covering 60 historical monthly returns. By contrast, when executing the backtesting exercise, one could also use the return data of 5 years ahead to estimate these statistics of funds instead of looking 5 years back which has the potential to capture future information under a perfect foresight assumption. It would be intriguing future work to repeat the current backtesting with this alternative estimation method based on a predictive view and compare with the current backtesting results based on a retrospective view.

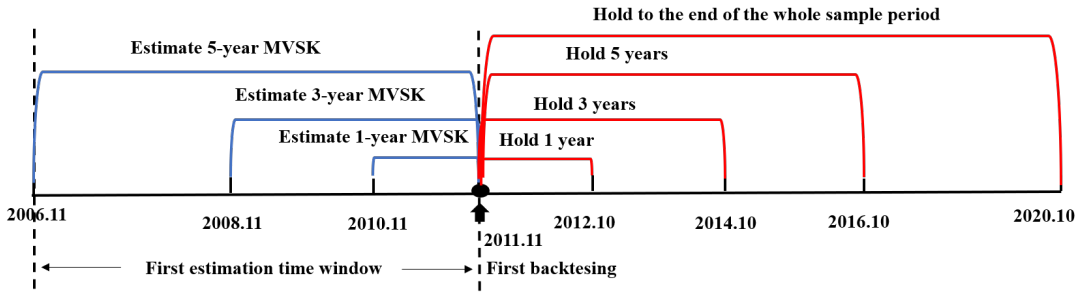

Figure 2.1: Process of the first backtesting window

Then, the time window is shifted with a step of a single month to develop the next backtesting analysis. For each time window or each backtesting event, the steps can be detailed as follows:

(1) Adopt the 5-year time window of data to compute the single-time frontier rankings, as well as the traditional financial rankings. In combination with the other two time periods (i.e., 1-year and 3-year) of data from this time window, the multi-time frontier ratings are computed.

(2) Depending on the ranking computed by this time window of data for each rating method, the 10, 20 or 30 best performing HFs are selected for the backtesting exercise, and then one holds these selected HFs for 1 year, for 3 years, for 5 years, and till the end of the whole sample period, respectively.

(3) In each of the above four holding period scenarios, we compute and store the complete historical track record of the holding value per buy-and-hold backtesting strategy, and then we calculate the mean, variance, skewness and kurtosis of these holding value series.

The above steps for backtesting are repeated over 48 time windows in total. For each of the four holding period scenarios, the performance of these MVSK

observations (15 times 48 observations) that are generated by the 15 strategies over 48 backtesting exercises are all evaluated by the shortage functions in the single-time and multi-moment frameworks (with convexity and nonconvexity). In particular, we first establish the convex and nonconvex VRS nonparametric frontiers in the single-time and multi-moment framework for these MVSK observations, and then measure their efficiency scores using the shortage functions. Clearly, each buy-and-hold strategy yields the efficiency scores of 48 MVSK observations. The average efficiency score and the number of efficient units, as well as the distribution of inefficiency scores across these 48 observations, are adopted to evaluate the 15 strategies. For the four holding scenarios, the same pattern is used to compare the 15 strategies based on the different rating methods.

2.4 Empirical Backtesting Results

As previously mentioned, the purpose of the empirical analysis is twofold. First, we examine whether the consideration of multiple moments and multiple times has an impact on both the efficiencies and the rankings of HFs. Second, we aim to further illustrate the eventual superiority of the proposed multi-time and multi-moment frontier rating methods by the backtesting analysis.

2.4.1 Sample Description

Considering the use of backtesting in the newly proposed multi-time and multi-moment ratings, the sample data collected requires the availability of continuous data for at least 14 years. Hence, we choose 187 HFs with monthly returns from October 2006 to October 2020 to test the 15 rating methods. The data is all down-

loaded from Lipper for Investment Management made available by Hedge Funds database. It needs to be stated that we initially specify these nonparametric frontier rating methods following the idea of Kerstens, Mounir, and Van de Woestyne (2011) that higher order moments and cost components are included. But, since HF cost data is unavailable in this database, our empirical analysis is limited to focus on the characteristics of the return distributions for these HFs without considering cost factors. In the following, we make a basic analysis of the monthly return characteristics of the 187 HF sample over the whole sample period. Table 2.2 reports descriptive statistics on the first four moments of the sample.

Table 2.2: Descriptive statistics for all 187 HFs over the whole sample period

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

From the descriptive statistics of the monthly returns reported in Table 2.2, we find that the series consisting of 187 HFs'skewness present positive mean and negative median, while the dispersion is quite large. Furthermore, all 187 HFs display positive kurtosis and also have a high dispersion. It is evident that some HFs do not perform well in terms of skewness and kurtosis. Therefore, for investors seeking non-negative skewness with small positive kurtosis, the multi-moment rating methods can be of great importance to select well-performing HFs from a large and heterogeneous HF universe. To assess the stability and persistence of these return characteristics over time, we further report the first four moments of the sample

over three time periods: a 1-year, a 3-year and a 5-year time periods, respectively, is presented in Table 2.6 in Appendix 2.A. Fundamentally, the same results regarding the return characteristics are available for these three time periods.

2.4.2 Evaluation Results

For the first aim of the empirical analysis, we compare both the efficiency distributions and the rankings of the 187 HFs calculated by the 15 rating methods.[16] In the single-time rating framework, we extract the monthly returns of these samples for the past 5 years to date to calculate the efficiency and ranking.While in the multi-time rating framework, the monthly returns for the past 1 year, 3 years and 5 years to date are integrated and applied to evaluate the performance of these funds.[17]

First, the efficiency distributions computed for the 15 rating methods are compared by means of nonparametric tests comparing two entire distributions initially developed by Li (1996) and refined by Fan and Ullah (1999) and most recently by Li, Maasoumi, and Racine (2009). It tests for the eventual statistical significance of differences between two kernel-based estimates of density functions f and g of a random variable x. The null hypothesis maintains the equality of both density functions almost everywhere: $H_0 : f(x) = g(x)$ for all x; while the alternative hypothesis negates this equality of both density functions: $H_1 : f(x) \neq g(x)$ for some x.[18] Table 2.3 provides Li-test statistics for all rating methods considered in

[16]Note: a fund o under evaluation, the direction vector g_t of each time t is $(-|V(R_o^t)|, |E(R_o^t)|)$ in the MV case, $(-|V(R_o^t)|, |E(R_o^t)|, |S(R_o^t)|)$ in the MVS case and $(-|V(R_o^t)|, -|K(R_o^t)|, |E(R_o^t)|, |S(R_o^t)|)$ in the MVSK case.

[17]The estimation time windows range of the 1-year, 3-year and 5-year MVSK respectively from November 2019 till October 2020, November 2017 till October 2020 and November 2017 till October 2020.

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

this contribution: In total, we report 105 relevant rating methods comparisons.[19]

Several observations can be made regarding the results in Table 2.3. First, it is clear that the efficiency distributions computed by traditional financial performance measures and those computed by frontier-based rating methods are significantly different at the 1 % significance level.

Second, in both convex and nonconvex frontier ratings, the single-time and multi-time rating methods yield significantly different efficiency distributions. This implies that the consideration of multiple times has a significant impact on the efficiency distributions.

Third, the effect of adding multiple moments on the efficiency distributions are somewhat different in single-time and multi-time ratings. For instance, in the case of convexity, adding skewness and kurtosis jointly has a significant effect on the efficiency distributions at the 1 % significance level in multi-time ratings. In single-time ratings, adding higher moments does not contribute in a significant way. Furthermore, the nonconvex frontier rating methods are more discriminatory in the impact of adding multiple moments. Compared to the above results in the case of convexity, in the case of nonconvexity, both adding skewness in itself and adding skewness and kurtosis jointly have significant effects on the efficiency distributions at 1 % significance level in multi-time ratings, and adding these jointly has a significant impact at 5 % significance level in single-time ratings.

Fourth, for multi-time ratings, imposing convexity always has a significant impact on the efficiency distributions. The efficiency distributions obtained by convex and nonconvex frontier ratings in MV, MVS and MVSK cases all yield differences at 1 % significance level, respectively. For the single-time ratings, the

[19]Note that Li-test statistics is applied to test whether the three specification factors considered in the proposed nonparametric frontier rating methods(i.e., multiple times, multiple moments and the imposition of convexity) have an impact on the efficiency distribution.

Table 2.3: Li-test statistics comparing the efficiency distributions computed by different rating methods (3 financial rating methods and 12 frontier rating methods)

	Eff(Sortino)	Eff(Omega)	STMVc	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
Eff(Sharpe)	13.105***	52.572***	32.974***	26.262***	28.112***	28.856***	19.95***	10.787***	37.267***	29.058***	33.588***	22.357***	12.428***	14.98***
Eff(Sortino)		36.715***	34.775***	27.804***	26.431***	15.464***	9.735***	4.772***	34.704***	28.537***	32.088***	11.644***	8.669***	11.052***
Eff(Omega)			30.826***	23.574***	25.33***	31.387***	28.876***	24.233***	31.146***	32.792***	36.324***	20.163***	25.316***	27.69***
STMVc				-5.818	-2.186	43.486***	40.519***	39.723***	-0.011	4.092***	6.917***	43.546***	37.76***	38.265***
STMVSc					-0.969	41.352***	36.06***	32.862***	0.162	2.763***	5.764***	38.787***	31.598***	32.38***
STMVSKc						6.845***	4.645***	28.251***	0.434	0.172	1.62*	36.439***	23.893***	23.503***
MTMVc							0.629	4.607***	47.637***	49.282***	52.157***	1.699**	12.015***	13.962***
MTMVSc								1.311*	39.891***	38.643***	41.47***	0.68	8.052***	10.026***
MTMVSKc									36.62***	28.562***	30.288***	-1.789	1.09	2.774***
STMVnc										0.206	2.272**	44.249***	33.723***	33.033***
STMVSnc											-1.476	39.818***	21.488***	19.946***
STMVSKnc												41.843***	21.709***	19.273***
MTMVnc													4.327***	6.209***
MTMVSnc														-2.158

Li test: critical values at 1% level= 2.33(***); 5% level= 1.64(**); 10%level= 1.28(*).

Table 2.4: Kendall rank correlations comparing the rankings computed by different rating methods (3 financial rating methods and 12 frontier rating methods)

	Eff(Sortino)	Eff(Omega)	STMVc	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
Eff(Sharpe)	0.956***	0.961***	0.833***	0.778***	0.627***	0.718***	0.639***	0.398***	0.824***	0.605***	0.592***	0.728***	0.406***	0.401***
Eff(Sortino)		0.933***	0.813***	0.776***	0.641***	0.737***	0.663***	0.427***	0.812***	0.620***	0.608***	0.745***	0.437***	0.431***
Eff(Omega)			0.823***	0.770***	0.620***	0.711***	0.632***	0.392***	0.822***	0.604***	0.589***	0.722***	0.400***	0.394***
STMVc				0.908***	0.749***	0.782***	0.668***	0.423***	0.948***	0.703***	0.690***	0.775***	0.435***	0.427***
STMVSc					0.836***	0.793***	0.741***	0.506***	0.875***	0.785***	0.772***	0.789***	0.520***	0.513***
STMVSKc						0.700***	0.706***	0.624***	0.725***	0.886***	0.887***	0.692***	0.630***	0.627***
MTMVc							0.857***	0.579***	0.770***	0.676***	0.665***	0.938***	0.580***	0.573***
MTMVSc								0.713***	0.661***	0.690***	0.680***	0.832***	0.713***	0.705***
MTMVSKc									0.418***	0.602***	0.606***	0.562***	0.905***	0.901***
STMVnc										0.732***	0.712***	0.775***	0.438***	0.429***
STMVSnc											0.967***	0.677***	0.651***	0.643***
STMVSKnc												0.668***	0.651***	0.657***
MTMVnc													0.594***	0.589***
MTMVSnc														0.977***

efficiency distributions of the convex and the nonconvex models are different at the 1 % and 10 % significance level in MVS and MVSK cases, respectively.

We further determine the Kendall rank correlations to test the degree of concordance in rankings determined by these performance measures. Table 2.4 shows the rank correlation between different HF ratings. In this table, *** indicates that the correlation coefficient between the rankings is significantly different from zero at 1 % significance level. The following key findings are revealed from Table 2.4. First, it is clear that the traditional financial ratings present a consistently low correlation (around 0.39-0.43) with the multi-time and multi-moment (MVS & MVSK) frontier ratings, but a high correlation (more than 0.8) with the single-time MV ratings. Second, turning to the comparisons between frontier ratings in single-time and multi-time frameworks, the single-time frontier rating and multi-time frontier rating show a low correlation overall. Third, the MV frontier rating exhibits a lower correlation with the multi-moment (MVS & MVSK) frontier ratings in multi-time framework compared in single-time framework. Moreover, the MV frontier rating has a lower correlation with the MVSK frontier rating compared with the MVS frontier rating. Finally, regarding comparisons between the rating models with convexity and nonconvexity, both the second and third findings tend to be more pronounced in the nonconvex case compared to the convex case.

From these analyses, we can conclude that the multiple moments and multiple times both separately and jointly have an impact on the HF efficiency and ranking for our data, and this impact is more significant when the two factors are considered jointly. Furthermore, nonconvexity may prove to be a more modest hypothesis in the proposed multi-time and multi-moment ratings since it exhibits a stronger discriminatory power with respect to the effect of adding multiple moments. This confirms earlier comparative results between the convex and nonconvex models

with higher order moments in the contribution of Kerstens, Mounir, and Van de Woestyne (2011).

2.4.3 Backtesting Results

We analyze the backtesting scenarios with a selection of the 10, 20 or 30 best performing HFs, respectively.[20] As stated previously, the 15 buy-and-hold strategies are compared in terms of the MVSK performances of their holding value series that are evaluated by the shortage functions based on the convex and nonconvex VRS frontiers in single-time and multi-moment frameworks (with convexity and nonconvexity). Table 2.5 presents an overall analysis with respect to the performances of the MVSK observations generated per strategy held until the end of the whole sample period. This table is structured as follows: The first series of four columns list the results with regard to the 10 best HFs selected for the backtesting exercise, and the second and third series of four columns present the results for selecting 20 and 30 best HFs, respectively. Within each selecting (buying) scenario, the first two columns report the average inefficiency scores and the number of efficient units for each strategy when evaluated using the convex VRS frontier in single-time and multi-moment framework (VRSc), while the last two columns report these results in the nonconvex case (VRSnc).

We first analyze the main findings in the context of buying and holding until the end of the whole sample period, as presented in Table 2.5. From these results, there are four main conclusions.

The first key finding is that all the frontier-based strategies outperform the

[20] All funds with an efficiency of 0 are ranked as 1 in our calculations. As a consequence, when we have to take a certain amount of funds among these ties then we take these randomly among the tied units (as in Brandouy, Kerstens, and Van de Woestyne (2015).

Table 2.5: Performance results for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 frontier rating methods): Descriptive statistics of the values of shortage function

Methods	HF(10)				HF(20)				HF(30)			
	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
Eff(Sharpe)	0.064	0	0.040	9	0.081	2	0.047	10	0.078	0	0.034	9
Eff(Sortino)	0.063	1	0.034	10	0.084	2	0.055	7	0.077	1	0.037	9
Eff(Omega)	0.064	0	0.031	10	0.084	1	0.059	4	0.077	0	0.040	7
STMVc	0.077	0	0.045	17	0.101	1	0.064	5	0.096	0	0.047	11
STMVSc	0.059	7	0.027	28	0.090	2	0.055	14	0.076	4	0.033	16
STMVSKc	0.044	6	0.014	31	0.070	4	0.039	17	0.059	1	0.031	15
MTMVc	0.061	1	0.020	22	0.075	1	0.038	14	0.078	2	0.032	11
MTMVSc	0.063	4	0.025	22	0.078	2	0.044	14	0.065	2	0.028	16
MTMVSKc	0.041	9	0.008	30	0.065	1	0.033	17	0.053	1	0.020	17
STMVnc	0.068	2	0.031	20	0.100	0	0.062	8	0.090	0	0.038	11
STMVSnc	0.042	5	0.023	16	0.054	4	0.029	19	0.039	5	0.014	25
STMVSKnc	0.042	4	0.026	13	0.040	6	0.022	27	0.035	7	0.012	26
MTMVnc	0.047	3	0.013	26	0.075	0	0.035	18	0.074	0	0.030	15
MTMVSnc	0.034	9	0.010	27	0.049	9	0.024	19	0.039	6	0.013	28
MTMVSKnc	0.039	5	0.012	31	0.047	7	0.021	21	0.032	7	0.009	28

strategies based on traditional financial indicators, except the strategies constructed by the single-time MV frontier rating methods. From the average inefficiency scores reported in Table 2.5, it is easy to see that the average inefficiency scores of all strategies based on the multi-moment and/or the multi-time frontier ratings are lower than those of Sharpe-, Sortino- and Omega-driven strategies. This result is valid when buying the 10, 20 and 30 best HFs. Combining the numbers of efficient units given in Table 2.5, the frontier-based strategies clearly yield more efficient units compared to those based on traditional indicators.

The second key result is that the buy-and-hold strategies according to the multi-moment ratings present superior results compared to those based on the MV ratings. Again, this result is confirmed when buying the 10, 20 and 30 best HFs. Both in the single-time and multi-time rating frameworks, we find that the strategies driven by the multi-moment ratings yield lower average inefficiency scores and a higher number of efficient units over strategies driven by the MV ratings.

Third, combining the two evaluation indicators of average inefficiency scores

and the number of efficient units, it is found that in the majority of cases the buy-and-hold strategies consisting of the HFs selected by the multi-time rating methods perform better than strategies consisting of the HFs selected by the single-time rating methods. This result remains valid when buying the 10, 20 and 30 best HFs.

A last key finding is that strategies determined by the nonconvex frontier-based ratings always outperform those determined by the convex frontier-based ratings. Moreover, by comparing the average inefficiency scores and the number of efficient units between the two in MVS and MVSK frameworks, it can be seen that when multiple moments are considered, the strategies based on the nonconvex frontier-based ratings usually display a more significant advantage. The reason for this finding is that skewness and kurtosis imply nonconvexities in diversified portfolio optimisation. As stated above, nonconvex production frontier models used for fund rating underestimate the nonconvex diversified portfolio models, while the convex production frontier models may tend to overestimate these same nonconvex diversified portfolio models.

Thus, this backtesting analysis shows that the buy-and-hold strategies constructed by our proposed multi-moment and multi-time rating methods exhibit superior performance in most scenarios. We therefore believe that the joint consideration of multi-moments and multi-times provides additional useful information for HF selection in practice.

As a sensitivity analysis, we test the performance of the 15 buy-and-hold backtesting strategies held for 1 year, 3 years and 5 years, which can be regarded as their short-, medium- and long-term holding performance. Table 2.7 in Appendix 2.B summarizes the performance results of the 15 strategies held for these three alternative holding periods. The above four findings are also evidenced in

most cases for these three holding period scenarios. Moreover, the buy-and-hold backtesting strategies consisting of the best HFs rated by the multi-moment and multi-time performance measure tend to show a consistent performance over the different holding periods. We basically conclude that the buy-and-hold strategies driven by the multi-moment and multi-time ratings exhibit favorable and consistent short-, medium- and long-term holding performance, somewhat implying that the performance of the best-performing HFs rated by the proposed multi-moment and multi-time performance measure would be sustained over time. A more detailed discussion on the sensitivity analysis is provide in Appendix 2.B.

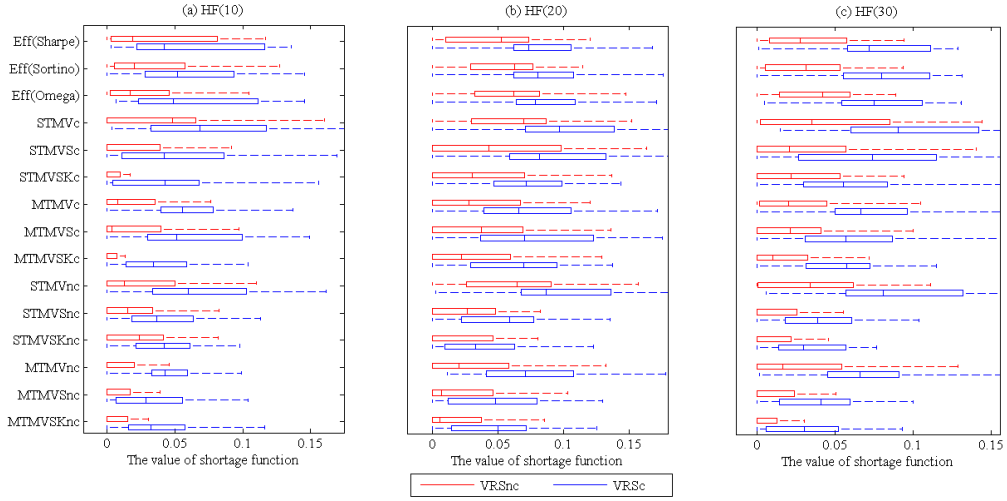

Figure 2.2: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 frontier rating methods)

Besides evaluating strategies based on the two summarized indicators reported in Table 2.5, we further provide the entire distribution of the inefficiency scores per strategy to compare these intuitively. Figure 2.2 presents a graphical overview of the performance of all strategies by integrating the box-plot per strategy held to end in the buying scenarios with 10, 20 and 30 HFs selected. In this figure, the sub-figures (a) to (c) correspond to the performance results of these three buying

scenarios. The box-plots for the performance of strategies based on the convex VRS frontier are in blue, and those based on the nonconvex VRS frontier are in red. In these box-plots, the box indicates the interquartile range where the small vertical lines reporting the location of the median. Their locations closer to the left suggests that the entire distribution of inefficiency scores for the strategy is at a lower level, which implies that the strategy has a better performance in backtesting analysis. As we can observe from Figure 2.2, comparing the performance of these strategies in each buying (backtesting) scenario, the buy-and-hold strategies constructed by the multi-moment and multi-time frontier rating methods are superior to strategies constructed by the existing rating methods in most cases.

Equally so, the entire distributions of the inefficiency scores for the 15 strategies held for 1, 3 and 5 years are presented in Figures 2.3, 2.4 and 2.5 in Appendix 2.B, respectively. From Figures 2.3, 2.4 and 2.5, one can observe that the dominance of the strategies driven by the multi-moment and multi-time ratings over other strategies remains valid and that this relation is strengthened as the holding period increases. It is therefore clear that the good performance of the strategies driven by the proposed frontier-based performance measures including multiple moments and multiple times exhibits good stability (see Appendix 2.B for details).

2.5 Conclusion

Inspired by recent nonparametric frontier rating methods contributing to assessing MF performance, this contribution has aimed to define a new shortage function or performance measure for rating MFs that can simultaneously handle both multiple moments and multiple times. Furthermore, we have explored the potential benefits of this new performance measure for selecting the best performing MF. We are now

in a position to summarize the main contributions.

First, we establish a series of nonparametric convex and nonconvex frontier rating methods with multi-moments and multi-times. The proposed rating methods are capable of not only assessing to which extent a MF performs well in the several moments following mixed risk-aversion preferences, but it simultaneously measures to which extent a MF performs well in all these moments in different times as well. These new multi-time and multi-moment performance measures are suitable for handling mixed risk-aversion preferences of investors which aim at time persistence.

Second, the proposed rating methods are empirically applied to HFs, given that HFs tend to exhibit strong asymmetric and long-tail return characteristics compared to other MFs. Using Li-test and Kendall rank correlation, the multi-time and multi-moment ratings are compared with traditional financial indicators and basic single-time MV rating methods to examine the impact of multiple moments and multiple times. From the comparison among 15 various rating methods, we find that in both convex and nonconvex cases, the multiple moments and multiple times both separately and jointly have an impact on the HF efficiency and ranking, and this impact is more significant when the two factors are considered jointly. Furthermore, the nonconvex rating models have stronger discriminatory power with respect to the effect of adding multiple moments over the convex rating models. This confirms earlier comparative results between convex and nonconvex models with higher order moments in Kerstens, Mounir, and Van de Woestyne (2011).

Third, having the impact of the multi-moments and multi-times in mind, we develop a simple buy-and-hold backtesting strategy to test whether the new ratings perform any better than more traditional financial ratings and single-time MV

ratings in HF selection. In most backtesting exercises, the buy-and-hold strategies based on the multi-time and multi-moment ratings exhibit a superiority over those based on traditional financial ratings and single-time MV ratings. This superiority is clearly confirmed by comparing the MVSK performance of holding values with respect to various buy-and-hold backtesting strategies. The multi-time and multi-moment strategies tend to exhibit more stable and favorable short-, medium- and long-term holding performance than the other strategies. Equally so, we focus on the comparison of these multi-time and multi-moment strategies in the convex and nonconvex cases. The strategies based on the nonconvex frontier ratings usually display a more significant advantage over the convex frontier ratings probably for reasons of a closer fit with the nonconvex skewness and kurtosis in diversified portfolio optimisation.

Overall, the proposed multi-time and multi-moment performance measures provide a novel idea into the important topic of rating and selecting MF. From the basic backtesting setup in our empirical analysis, further extensive backtesting studies can be developed to exploit the potential of the new performance measures in constructing fund of funds. This is one of the main avenues for future research. Another desirable extension is to transfer the current methodological framework and perform a backtesting analysis in diversified models. It is worthwhile to compare the performance in MF selection between the backtesting strategies driven by diversified frontier rating methods and those driven by nondiversified frontier rating methods (i.e., convex and nonconvex frontier rating methods).

References

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

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

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

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

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

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

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

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

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

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

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

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

Briec, W., A. Dumas, and A. Mekki (2021): "Directional Distance Functions and Social Welfare: Some Axiomatic and Dual Properties," *Mathematical Social Sciences*, 113, 181–190.

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

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

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

Briec, W., K. Kerstens, and J. Lesourd (2004): "Single Period Markowitz Portfolio Selection, Performance Gauging and Duality: A Variation on the Lu-

enberger Shortage Function," *Journal of Optimization Theory and Applications*, 120(1), 1–27.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Jurczenko, E., and G. Yanou (2010): "Fund of Hedge Funds Portfolio Selection: A Robust Non-parametric Multi-moment Approach," in *The Recent Trend*

of Hedge Fund Strategies, ed. by Y. Watanabe, pp. 21–56. Nova Science, New York.

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

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

Krüger, J. (2021): "Nonparametric Portfolio Efficiency Measurement with Higher Moments," *Empirical Economics*, 61(3), 1435–1459.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2.A Sample Description: Further Details

The descriptive statistics on the first four moments of the 187 HF sample over 1-year time period (sample period: Nov 2019 to Oct 2020), 3-year time period (sample period: Nov 2017 to Oct 2020) and 5-year time period (sample period: Nov 2015 to Oct 2020) are provided in Table 2.6.

Table 2.6: Descriptive statistics for all 187 HFs over 1-, 3- and 5-year time periods

1 year monthly return: From Nov 2019 to Oct 2020				
	Mean	Variance	Skewness	Kurtosis
Min.	−3.913	0.241	−4765.126	0.305
Q1	−0.467	8.580	−135.113	259.251
Median	0.356	18.270	−12.409	1277.641
Mean	0.423	45.237	414.608	71518.803
Q3	0.930	45.155	14.606	6685.385
Max.	10.857	937.351	60038.381	5381641.258
3 year monthly return: From Nov 2017 to Oct 2020				
	Mean	Variance	Skewness	Kurtosis
Min.	−2.836	0.258	−1820.695	0.217
Q1	−0.134	7.925	−65.187	344.122
Median	0.175	14.259	−17.723	945.365
Mean	0.248	27.364	240.996	36348.550
Q3	0.463	28.983	2.409	3974.036
Max.	4.623	533.743	30000.036	2865464.753
5 year monthly return: From Nov 2015 to Oct 2020				
	Mean	Variance	Skewness	Kurtosis
Min.	−1.618	0.228	−1142.152	0.196
Q1	−0.005	7.290	−45.376	275.106
Median	0.254	12.959	−12.149	715.167
Mean	0.345	26.036	243.423	39749.791
Q3	0.609	25.073	0.299	3002.459
Max.	3.943	705.232	27466.851	3289535.317

As observed from column 4 in Table 2.6, we find that for each time, the series

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

2.B Backtesting Results: Sensitivity Analysis

To develop a sensitivity analysis with respect of the holding period, this Appendix focuses on testing the short-, medium- and long-term holding performance of the buy-and-hold backtesting strategies based on the proposed multi-moment and multi-time rating methods. The performance of strategies held for only 1 year is regarded as a short-term holding performance, for 3 years as a medium-term holding performance, and for 5 years as a long-term holding performance. For each of the three holding scenarios, the 15 strategies are compared in terms of the MVSK performances of their holding values that are always evaluated by the shortage function based on the convex and nonconvex VRS frontiers in single-time and multi-moment framework. Table 2.7 reports the summarized results with respect to the performance per buy-and-hold backtesting strategy held for 1, 3 and 5 years.

Table 2.7 is organized as follows: the three series consisting of four columns

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

We now analyze the results on the three holding scenarios presented in Table 2.7, following the same basic logic of analysing the 15 strategies in the main text. Thus, the performance of strategies generated by two family of ratings (frontier vs. finance) is compared first, and then the comparison between the frontier families of ratings is developed separately (i.e., multi-moments vs. MV; multi-times vs. single time; convexity vs. nonconvexity).

We first discuss the short-term holding performance of the 15 buy-and-hold strategies, as shown in columns 2-6 of Table 2.7. First, it can be observed that minor difference on the short-term holding performance is observed between the strategies depending on the multi-time frontier ratings and those depending on traditional financial ratings, and both their performances are superior over other frontier-based strategies. Second, in most cases, the strategies based on the multi-moment ratings do not show superiority compared to those based on the MV ratings when these strategies are held for only 1 year. This result is somewhat at odds with the one reported in the main text. Third, combining average inefficiency scores and the number of efficient units, the strategies constructed in the multi-time rating framework perform better over those in the single-time rating framework under the 1-year holding scenario. Finally, in terms of short-term holding performance, the strategies determined by the nonconvex frontier-based

Table 2.7: Performance results for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 frontier rating methods) held for 1, 3 and 5 years: Descriptive statistics of the values of shortage function

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

ratings outperform those determined by the convex frontier-based ratings in the majority of cases. The latter finding is in line with the one shown in the main text. It needs to be mentioned that some of the findings may be somewhat unstable with respect to the 1-year holding period due to the limited data for testing the short-term holding performance of the 15 buy-and-hold strategies.

Looking at columns 7-10 of Table 2.7 for the medium-term holding performance of the 15 strategies, one can draw the following observations. The frontier-based strategies with consideration of multi-moments and multi-times (separately or jointly) largely outperform the finance-based strategies. It is easy to observe that the strategies driven by the multi-moment and multi-time frontier ratings generally yield lower average inefficiency scores and more efficient units compared to Sharpe-, Sortino- and Omega-driven strategies. Turning to the comparisons between various frontier-based rating methods, the buy-and-hold strategies based on the multi-moment ratings (MVS & MVSK) perform better than those based on the basic MV ratings. This is confirmed in both single-time and multi-time rating frameworks. Moreover, consistent with the finding on considering multiple times in the 1-year holding scenario (see the third finding), the multi-time frontier-based strategies outperform the single-time frontier-based strategies in most cases in the 3-year holding scenario. Again, when comparing convex and nonconvex frontier-based strategies in the medium-term holding scenario, the same coherent finding emerges as in the short-term holding scenario (see the final finding analyzed in the 1-year holding context).

Following up the results regarding the 5-year holding scenario as reported in columns 11-14 of Table 2.7, the above four findings emerging in the 3-year holding period are also evidenced in this holding scenario. These results in the medium- and long-term holding scenarios are rather in line with the ones reported in the

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

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

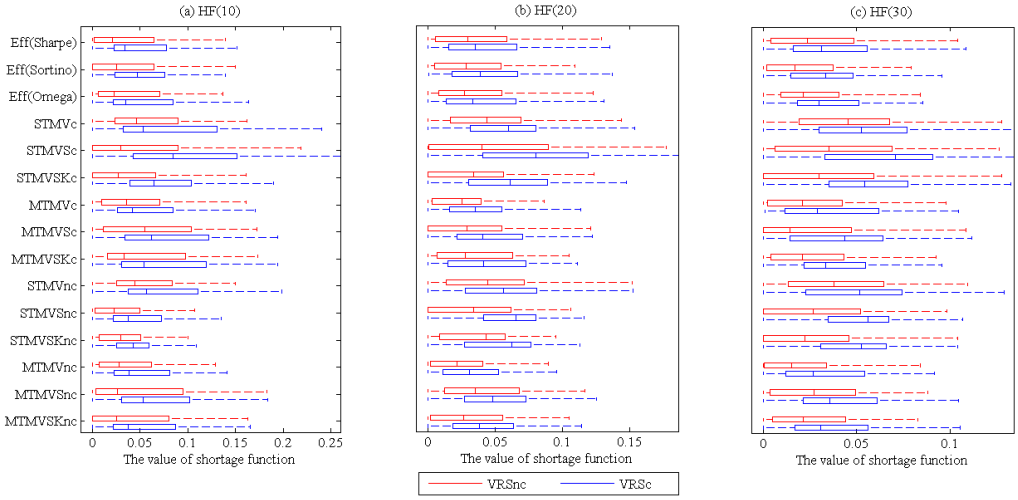

Figure 2.3: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 frontier rating methods): held for 1 year

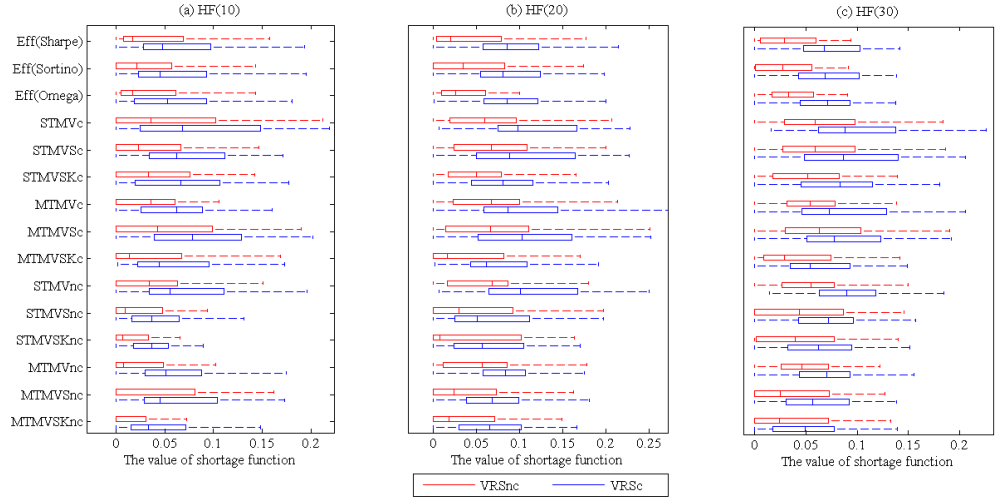

Figure 2.4: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 frontier rating methods): held for 3 years

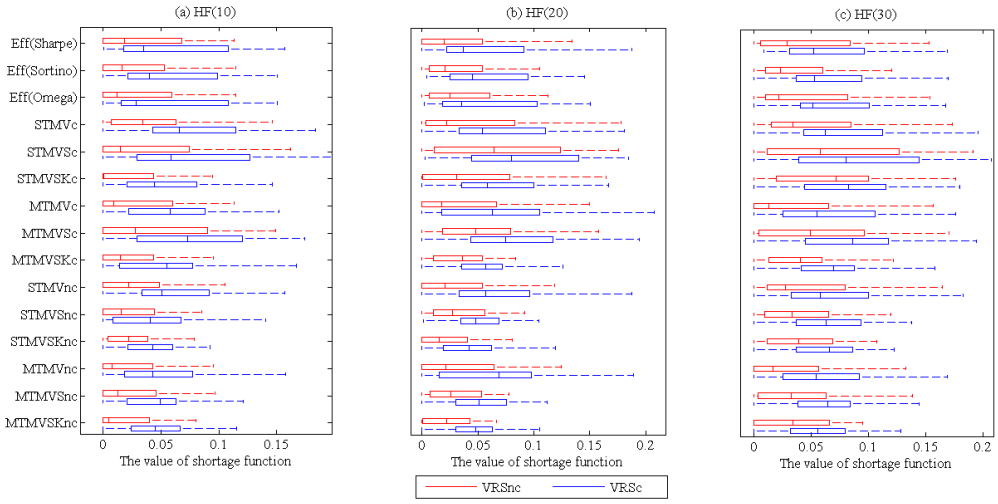

Figure 2.5: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 frontier rating methods): held for 5 years

To compare the 15 buy-and-hold strategies intuitively, Figures 2.3, 2.4 and 2.5 offer box-plots to describe the entire distributions of the inefficiency scores per strategy held for 1, 3 and 5 years, respectively. In each figure, the sub-figures (a) to (c) correspond to the performance results of the buying scenarios with 10, 20 and 30 best HFs selected, whereby the performance of strategies based on the convex VRS frontier are depicted in blue, and those based on the nonconvex VRS frontier are displayed in red. As introduced in the main text, the box of these box-plots indicates the interquartile range where the small vertical line reports the location of the median. Straightforwardly, the location of the median closer to the left indicates that the entire distribution of inefficiency scores for one strategy is somewhat skewed to the left, which signals that the strategy performs better in the backtesting analysis because the probability mass of the inefficiency is closer to zero. Two major observations can be made with regard to these results in Figures 2.3, 2.4 and 2.5. First, although the buy-and-hold strategies constructed by the

multi-moment and multi-time frontier rating methods do not exhibit a significant superiority in the 1-year holding scenario, they establish a clear dominance over the other strategies in both the 3- and 5-year holding scenarios (see sub-figures (a) and (b) of Figures 2.3, 2.4 and 2.5). Second, concentrating on Figures 2.3, 2.4 and 2.5 individually, the good performance of the buy-and-hold strategies depending on the multi-moment and multi-time ratings tends to be consistent and stable over time.

CHAPTER

3

Multi-Time and Multi-Moment Nonparametric Frontier-Based Fund Rating: Luenberger Portfolio Productivity Indicator Approach

Abstract

This contribution introduces a novel method for measuring the evolution of fund performance in a multi-time and multi-moment framework using some extensions of the shortage function combined with a Luenberger productivity indicator. This new Luenberger discrete time portfolio productivity indicator and its decomposition are constructed to measure fund's dynamic performance with the consideration of multiple time periods and multiple moments, and to distinguish whether performance changes over time are due to the multi-time efficiency change or due to the shift in the frontier. The empirical part employs actual hedge fund (HF) data to illustrate the validity and practicality of the proposed methodology. The evaluation results show that multiple time periods and higher-order moments

have an clear impact on the Luenberger portfolio productivity indicator, as well as the indicator-based ratings. Moreover, the backtesting results point to a remarkable superior performance of the multi-time and multi-moment Luenberger-based ratings compared to traditional finance-based ratings and the single-time MV Luenberger-based ratings for the actual HF selection.

KEYWORDS: Shortage function; Nonparametric frontier; Fund rating; Luenberger productivity indicator.

3.1 Introduction

Financial analysts and individual investors rely heavily on the risk-adjusted return (i.e., "performance") measures to select investment funds. Since the foundational work of Markowitz (1952) and Tobin (1958) on modern portfolio mean-variance (MV) theory, the MV efficiency is defined as the ability of a set of assets to yield the maximum return for a given level of risk or, alternatively, to produce the minimum level of risk for a given expected return. The modern portfolio MV theory pioneers a perspective for measuring fund efficiency, but this dual objective of MV makes performance evaluation a controversial task. Indeed, none of the current methods in the literature seems to be universally approved.

Some traditional financial efficiency indicators (such as the Sharpe ratio) basically developing under the MV evaluation framework have received widespread criticism. As stated by numerous authors (see Hanoch and Levy (1969) and Leland (1999)), the MV evaluation is not consistent with the Von Neumann–Morgenstern axioms of expected utility theory unless either (i) the asset returns obey a normal probability distribution, which is one of theoretical assumptions of the Capital Asset Pricing Model (CAPM); or (ii) the utility functions of investor preferences are

quadratic, and depend on the first and second moments (i.e., mean and variance) of the asset's distribution. One the one hand, there is a lot of evidence showing that asset returns are non-normally distributed (see Arditti (1975); Konno, Shirakawa, and Yamazaki (1993); Harvey, Liechty, Liechty, and Müller (2010), and Aksarayl ı and Pala (2018)). On the other hand, the broad class of mixed risk-aversion utility functions with a general preference for odd moments and an aversion to even moments is nowadays widely accepted (see Trautmann and van de Kuilen (2018) for a review). Furthermore, some investigations indicate that portfolio selection based on the MV criterion in the presence of non-normally asset returns and mixed risk-aversion investor preferences can lead to severe welfare losses (e.g., Hwang and Satchell (1998); Dittmar (2002); Jondeau and Rockinger (2006)).

Scott and Horvath (1980) confirm that investors are willing to accept lower expected return and higher volatility compared to the MV benchmark in exchange for higher skewness and lower kurtosis. Thereafter, several alternative methods have developed in the financial literature to incorporate the individual preferences for higher-order moments into portfolio evaluation and selection. For instance, Lai (1991) develops a polynomial goal programming (PGP) approach to estimate the set of the mean-variance-skewness-kurtosis (MVSK) efficient funds. Guidolin and Timmermann (2002) describe a particular specification of the indirect MVSK utility by using Taylor series expansion to determine the optimal portfolios. Although none of these procedures has managed to obtain widespread acceptance, it is nowadays rather generally recognized that the skewness and kurtosis are critical when considering alternative assets or funds (e.g., hedge funds, managed futures, private equity, and leveraged buy-out) in constructing a portfolio. To this end, in this contribution, we develop a general portfolio evaluation procedure with multiple moments in the methodology part, while in the empirical part we specifically place ourselves in the MVSK framework.

Taking the mixed risk-aversion preference structures for granted, a particular distance function, the shortage function, has been proven to be an excellent tool to derive efficiency measures in combination with investors' preferences for higher-order moments. Briec, Kerstens, and Lesourd (2004) are the first to develop the shortage function (Luenberger (1995)) to measure portfolio efficiency, where investors can both contract variance and expand expected return simultaneously. While their work is still developed within the MV evaluation framework, Briec, Kerstens, and Lesourd (2004) establish a link between portfolio theory and developments in production theory. Following their contribution, one can summarize the basic idea for assessing portfolio performance: By the shortage function, the efficiency of a given portfolio is measured based on the distance between it and the MV portfolio frontier (theoretical frontier). Immediately thereafter, this evaluation method based on the shortage function is extended to more generalized portfolio frameworks to be compatible with general mixed risk-aversion investor preferences. Briec, Kerstens, and Jokung (2007) demonstrate that the shortage function can project any inefficient portfolio exactly on the three dimensional MVS portfolio frontier and that the function is connected via duality to an indirect MVS utility function if one is able to articulate preferences. Briec and Kerstens (2010) discuss the shortage function in a general higher-order moment setting, in which the MVSK case is covered.

Differing from gauging funds based on the portfolio frontier that considers diversification effect, Kerstens, Mounir, and Van de Woestyne (2011b) launch a new proposal in favor of the use of shortage function in terms of convex/nonconvex nonparametric frontiers, and systematically test for the need of higher-order moments in defining the efficiency measures based on these non-diversified methods. There has been a great development on the convex or nonconvex frontier-based methodology for assessing funds, e.g., Matallín-Sáez, Soler-Domínguez, and Tortosa-

Ausina (2014), Brandouy, Kerstens, and Van de Woestyne (2015), Nalpas, Simar, and Vanhems (2017), Mehlawat, Kumar, Yadav, and Chen (2018), Krüger (2021), Lin and Li (2020), Gong, Yu, Min, and Ge (2021). In particular, one can find that the shortage function based on the non-diversified frontier techniques has become very popular in performance evaluation combining simultaneously several characteristics of fund return distributions into a single performance measure (i.e., mean, variance, skewness and kurtosis).

In the mean time, several studies have focused on the role of time dimensions in performance evaluation of funds. Morey and Morey (1999) are the first to propose two efficiency measures under risk and return orientations respectively from a multi-period perspective for MV portfolio performance appraisals. The first efficiency measure aims at contracting all risk dimensions without altering any return dimensions, and the other efficiency measure aims at augmenting all return dimensions with no increase in any risk dimensions. Briec and Kerstens (2009b) develop a multi-period MV portfolio analysis by means of the shortage function. Instead of either proportionally contracting risk dimensions or proportionally expanding return dimensions, a temporal shortage function simultaneously attempts to reduce the risk and to expand the return over all times. These multi-period evaluation methods allow to arise the difference in preferences across time and can release more useful information on fund selection to investors compared to gauging the fund's performance from a single time (see Ren, Zhou, and Xiao (2021)).

The foregoing discussion on time dimensions and higher-order moments inspire our fundamental idea to offer a multi-time and multi-moment performance measure by the extension of the existing shortage function. Considering the computational superiority of non-diversified models, the contents of this contribution are developed based on the convex and nonconvex nonparametric non-diversified

frontier technology. Indeed, it seems more appropriate that the proposed performance measure offers a general framework for portfolio evaluation, allowing for multiple factors associated with the investment decision process to be considered. However, although this temporal performance measure takes into account multiple times in a time horizon, admittedly, it still falls into the category of static performance evaluation that focuses on measuring a fund's efficiency by a cross-sectional comparison among all observed funds using the return information over a finite horizon.[1]

In recent years, differing from this static analysis, some studies are devoted to measure the evolution of fund performance over time by a cross-period comparison to investigate the attribution problem for the efficiency changes. This can be regarded as a dynamic evaluation for a fund performance in discrete time. In a dynamic context, both the window analysis and the productivity indicator analysis are commonly applied to measure the performance evolution of a decision-making unit (see, e.g., Gardijan and Škrinjarić (2015), Zeitun, Abdulqader, and Alshare (2013), Asmild, Paradi, Aggarwall, and Schaffnit (2004), Babalos, Caporale, and Philippas (2012), Gregoriou and Moschella (2017)). The window analysis suffers from some limitations, e.g., the absence of interpretation for the performance change and low test frequency at the beginning and ending windows (Lin, Chen, Hu, and Li (2017)). Compared to the window analysis, the productivity indicator analysis based on frontier-based evaluation models can identify the progress or regress in efficiency along with the progress or regress of the frontier over time (Cooper, Seiford, and Tone (2007)). Borrowing from the theory of productivity indicators, Brandouy, Briec, Kerstens, and Van de Woestyne (2010) integrate the shortage function based on the MV portfolio frontier into the Luenberger

[1]Note Briec, Comes, and Kerstens (2006) introduce the notions of an temporal technical and profit efficiency measurement which do not allow for linkages between optimal decisions between time periods and are thus only dynamic in a limited sense.

productivity indicator (Chambers, Färe, and Grosskopf (1996)) to measure the portfolio performance change over time. This portfolio productivity indicator is decomposed into an efficiency change component and a frontier change component. The former estimates the efficiency changes over time attributable to the ability of the portfolio manager, and the latter component captures the efficient frontier shift induced by market volatility. Their contribution is the first to establish a link between the indicator theory and the portfolio performance evaluation framework, to some extent revealing the potential of the Luenberger portfolio productivity indicator and its decomposition as a measurement tool for fund performance appraisal in discrete time. However, their work is limited to discuss the portfolio performance in a single-time framework. In addition, since all diversified models for computing this Luenberger portfolio productivity indicator are nonlinear, the computation process is time-consuming, which is not suitable for the large-scale and multidimensional portfolio evaluation problems. Therefore, we opt for non-diversified production frontier models.

Following Brandouy, Briec, Kerstens, and Van de Woestyne (2010), the aim of this contribution is threefold. First, we propose a series of performance measures in the shortage function framework, which are compatible with general investor preferences regarding multiple time periods and multiple moments. Using the new performance measures, we further develop a multi-time and multi-moment Luenberger portfolio productivity indicator and its decomposition for measuring dynamic fund performance. Second, applying these models to actual HF data, we test the impact of multiple time periods and multiple moments separately and jointly on the Luenberger portfolio productivity indicators and their decomposition, as well as the indicator-based ratings for the HF sample. Third, we explore the potential benefit of the proposed Luenberger portfolio productivity indicators on fund selection by a backtesting analysis approach. To the best of our know-

ledge, this contribution is the first to develop a proposal on the methodology and application for the Luenberger portfolio productivity indicators in a multi-time and multi-moment framework.

The structure of this contribution is as follows. Section 3.2 presents the discussion on the nonparametric frontier methodology dealing with estimation of the single-time and multi-time Luenberger portfolio productivity indicators. Section 3.3 describes the details of the backtesting setup in dynamic analysis. The empirical results are presented in Section 3.4. Conclusions and issues for future work are summarized in the final section.

3.2 Nonparametric Frontier Rating in Dynamic Analysis Framework: Methodology

3.2.1 Single-Time and Multi-Moment Luenberger Portfolio Productivity Indicator

To introduce some basic notations and definitions, assume a sample of n funds is observed over a given time horizon, for each time t within this time horizon, each fund is characterized by m input-like values x_{ij}^t and s output-like values y_{sj}^t, where $i = 1, \ldots, m,\ r = 1, \ldots, s$. We introduce one widely used production frontier-based model with variable returns to scale (VRS). Following Briec, Kerstens, and Vanden Eeckaut (2004), a unified algebraic representation of convex and nonconvex production possibility sets (PPS) under the VRS assumption for a sample of n

funds at time t is:

$$P_{\Lambda}^{t} = \Big\{(x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s \mid \forall i \in \{1, \ldots, m\} : x_i^t \geq \sum_{j=1}^{n} \lambda_j x_{ij}^t, \\ \forall r \in \{1, \ldots, s\} : y_r^t \leq \sum_{j=1}^{n} \lambda_j y_{rj}^t, \lambda \in \Lambda\Big\}, \qquad (3.2.1)$$

where
$\Lambda \equiv \Lambda^C = \{\sum_{j=1}^{n} \lambda_j = 1 \text{ and } \lambda_j \geq 0\}$ if convexity is assumed, and $\Lambda \equiv \Lambda^{NC} = \{\sum_{j=1}^{n} \lambda_j = 1 \text{ and } \lambda_j \in \{0, 1\}\}$ if nonconvexity is assumed.

For any fund being evaluated when there exists a combination $(\sum_{j=1}^{n} \lambda_j x_{ij}^t, \sum_{j=1}^{n} \lambda_j y_{ij}^t)$ in the convex PPS or the nonconvex PPS using less inputs and producing more outputs than it, this fund can always improve its inputs or outputs.

Using the above nonparametric PPS (see (3.2.1)), the shortage function of any fund for a single time t is now defined as follows:

Definition 3.2.1. *For time t in a given time horizon, let $g^t = (-g_x^t, g_y^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ and $g^t \neq 0$. For any fund $v^t = (x^t, y^t) \in P_{\Lambda}^t$ in the direction of vector g^t, the shortage functions with and without convexity can be uniformly defined as:*

$$S_{\Lambda}^{t}(v^t; g^t) = sup\{\beta \in \mathbb{R} | v^t + \beta g^t \in P_{\Lambda}^{t}\}. \qquad (3.2.2)$$

This shortage function is termed as the single-time shortage function in this contribution, and simultaneously permits the enhancement of output-like variables and the reduction of input-like variables for time period t. If the shortage function value $S_{\Lambda}^{t}(v_o^t; g_o^t) > 0$ for a specific fund $v_o^t = (x_o^t, y_o^t)$ with index $o \in \{1, \ldots, n\}$, it means that v_o^t must not be located on the frontier derived from P_{Λ}^{t}, and its inputs and outputs can be improved to catch up with the nonparametric VRS frontier.

Following Brandouy, Briec, Kerstens, and Van de Woestyne (2010), the combination of some shortage function extensions with the Luenberger productivity indicators provides an effective way to evaluate the dynamic performance of funds. To define the Luenberger portfolio productivity indicator in a general context, we assume τ ($\tau \geq 0$) denotes the time interval between the two different times of interest.[2] To assess a fund's performance evolution across times t and $t+\tau$, the Luenberger portfolio productivity indicator for measuring the change of the shortage function is defined between times t and $t+\tau$:

$$\begin{aligned} L_{\Lambda}(v^t, v^{t+\tau}; g^t, g^{t+\tau}) =& \frac{1}{2}[(S^t_{\Lambda}(v^t; g^t) - S^t_{\Lambda}(v^{t+\tau}; g^{t+\tau})) \\ &+ (S^{t+\tau}_{\Lambda}(v^t; g^t) - S^{t+\tau}_{\Lambda}(v^{t+\tau}; g^{t+\tau}))] \end{aligned} \tag{3.2.3}$$

The Luenberger portfolio productivity indicator based on the single-time shortage function is referred to as single-time Luenberger indicator in our work. Accordingly, an equivalent decomposition of this Luenberger indicator is provided.

$$\begin{aligned} L_{\Lambda}(v^t, v^{t+\tau}; g^t, g^{t+\tau}) =& E_{\Lambda}(v^t, v^{t+\tau}; g^t, g^{t+\tau}) \\ &+ F_{\Lambda}(v^t, v^{t+\tau}; g^t, g^{t+\tau}) \end{aligned} \tag{3.2.4}$$

with

$$E_{\Lambda}(v^t, v^{t+\tau}; g^t, g^{t+\tau}) = S^t_{\Lambda}(v^t; g^t) - S^{t+\tau}_{\Lambda}(v^{t+\tau}; g^{t+\tau}) \tag{3.2.5}$$

and

$$\begin{aligned} F_{\Lambda}(v^t, v^{t+\tau}; g^t, g^{t+\tau}) =& \frac{1}{2}[(S^{t+\tau}_{\Lambda}(v^{t+\tau}; g^{t+\tau}) - S^t_{\Lambda}(v^{t+\tau}; g^{t+\tau})) \\ &+ (S^{t+\tau}_{\Lambda}(v^t; g^t) - S^t_{\Lambda}(v^t; g^t))] \end{aligned} \tag{3.2.6}$$

In the above decomposition, $E_{\Lambda}(\cdot)$ measures the efficiency change of fund between

[2]The time interval between different times in this contribution is general and flexible. In the following empirical computations, this value is fixed as $\tau = 3$ representing the three-month interval between two different times, which mainly accounts for the adjustment cycle of the fund manager's strategy in practice.

times t and $t+\tau$, while $F_\Lambda(\,\cdot\,)$ captures the average change in fund performance between the two times evaluated in t and $t+\tau$. Hence, (3.2.4) decomposes the change of fund's performance into two components: one representing efficiency change relative to a moving VRS frontier (i.e., $E_\Lambda(\,\cdot\,)$), another indicating the average change in the frontier itself (i.e., $F_\Lambda(\,\cdot\,)$). This decomposition offers a measurement framework for fund market performance gauging: On the one hand, $E_\Lambda(\,\cdot\,)$ captures the performance of the fund managers over time relative to a shifting frontier, and on the other hand, $F_\Lambda(\,\cdot\,)$ indicates how the fund market itself has locally changed over time. When the Luenberger indicator of fund performance change $L_\Lambda(\,\cdot\,)$ or any of its components ($E_\Lambda(\,\cdot\,)$ or $F_\Lambda(\,\cdot\,)$) is positive (negative), then fund's performance increases (decreases) between the two times considered.

Now, we discuss the computational matters of the single-time Luenberger indicator for any fund under evaluation. Based on the above definitions and notations, the PPS (see (3.2.1)) are used to directly compute the various shortage functions and thus the Luenberger productivity indicator by recourse to mathematical programming models. Consider any fund $v_o^t = (x_o^t, y_o^t)$ with index $o \in \{1, \ldots, n\}$ to assess, for a given direction vector denoted by $g_o^t = (-g_{io}^t, g_{ro}^t)$, where $i = 1, \ldots, m$ and $r = 1, \ldots, s$, let us first solve for the shortage functions based on Definition

(3.2.1) resulting from the following model:

$$
\begin{aligned}
\max \quad & \beta \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j y_{rj}^t \geq y_{ro}^t + \beta g_{ro}^t, \quad r = 1, \ldots, s, \\
& \sum_{j=1}^{n} \lambda_j x_{ij}^t \leq x_{io}^t - \beta g_{io}^t, \quad i = 1, \ldots, m, \\
& \sum_{j=1}^{n} \lambda_j = 1, \\
& \forall j = 1, \ldots, n : \begin{cases} \lambda_j \geq 0, & \text{under convextiy} \\ \lambda_j \in \{0, 1\}, & \text{under nonconvextiy} \end{cases}
\end{aligned}
\tag{3.2.7}
$$

where β and $\lambda_j (j = 1, \ldots, n)$ are decision variables, and the direction vector is opted here to employ the observation under evaluation itself, that is, $g^t = (-|x_{1o}^t|, \ldots, -|x_{mo}^t|, |y_{1o}^t|, \ldots, |y_{so}^t|)$. To calculate the Luenberger indicator $L_\Lambda(\,\cdot\,)$ of the evaluated fund o, the four different shortage functions composing it must be computed: $S_\Lambda^t(v_o^t; g_o^t)$, $S_\Lambda^t(v_o^{t+\tau}; g_o^{t+\tau})$, $S_\Lambda^{t+\tau}(v_o^{t+\tau}; g_o^{t+\tau})$, and $S_\Lambda^{t+\tau}(v_o^t; g_o^t)$. To solve for $S_\Lambda^t(v_o^t; g_o^t)$ and $S_\Lambda^{t+\tau}(v_o^{t+\tau}; g_o^{t+\tau})$, one can compute (3.2.7) directly for times t and $t+\tau$, respectively. For the second (last) computation, it is found by using the VRS-based PPS at time t (or time $t+\tau$), occurring on the right-hand side of the s and m constraints of model (3.2.7) are set as those observations at time $t+\tau$ (or time t).[3]

It is noted that model (3.2.7) results in a linear programming (LP) problem under convexity and a mixed integer linear programming (MILP) problem under nonconvexity.

In addition, when nonconvexity is taken into account, we can obtain a noncon-

[3]The non-negativity constraint on the output projections in production (see Briec and Kerstens (2009a)) are not needed in finance, because the negative output projection for a fund is typically possible.

vex VRS frontier composed by all efficient funds excluding the convex combinations of these. In application, convexity is not always useful to guide investors in terms of choosing among funds, especially when higher moments are considered as input-like variables and output-like variables to describe the return characteristics of funds. The convex production frontier models may tend to overestimate the corresponding nonconvex diversified portfolio models when including higher-order moments, while the nonconvex production frontier models provide a conservative underestimation of these same nonconvex diversified portfolio models (see Kerstens, Mounir, and Van de Woestyne (2011b) and Nalpas, Simar, and Vanhems (2017)).[4]

The above setting for single-time shortage function and Luenberger indicator is very general and flexible and can thus handle a large choice of inputs and outputs. We now particularize the formulation and the characterization of the efficient frontier in the multi-moment (MVS and MVSK) spaces. Suppose that there are n funds under evaluation. For time period t, let $R_1^t, \ldots, R_n^t$ denote the random returns of the n funds, which are characterized by their mean $E(R_j^t)$, variance $V(R_j^t)$, skewness $S(R_j^t)$ and kurtosis $K(R_j^t)$ for $j \in \{1, \ldots, n\}$. Here, the calculations of variance, skewness and kurtosis are expressed as follows: $V(R_j^t) = E[(R_j^t - E(R_j^t))^2]$, $S(R_j^t) = E[(R_j^t - E(R_j^t))^3]$, and $K(R_j^t) = E[(R_j^t - E(R_j^t))^4]$. To obtain the inputs/outputs representation of the PPS, we need to classify the different goals of the investor in terms of inputs, (i.e. objectives to minimize) and outputs (i.e. objectives to be maximized). As discussed in the previous section, the need for multiple moments is apparent to assess funds whose return distributions exhibit strong asymmetry and fat tail. In general, investors express preference

[4]Note that the nonconvex diversified portfolio models introduced here could be used to geometrically construct and visualize in a systematic way portfolios frontiers. Kerstens, Mounir, and Van de Woestyne (2011a) systematically develop geometric representations of the MVS diversified portfolio frontier, and the non-convex nature of the efficient MVS frontier is illustrated visually.

for odd moments and reluctance for even moments of the distribution of asset returns. Therefore, when the MVSK framework is considered, the inputs of funds are defined as $x_{1j}^t = V(R_j^t)$ and $x_{2j}^t = K(R_j^t)$, as well as the outputs are defined as $y_{1j}^t = E(R_j^t)$ and $y_{2j}^t = S(R_j^t)$ for $j \in \{1, \ldots, n\}$, whereas for the MVS case, only the first input is considered.

3.2.2 Multi-Time and Multi-Moment Luenberger Portfolio Productivity Indicator

Similar to the procedure for defining the single-time Luenberger portfolio production indicator, this subsection is concerned with a multi-time Luenberger portfolio productivity indicator to track the evolution of fund multi-time performance over time. Equally so, to develop the nonparametric frontier rating models in multi-time framework, some definitions and properties are presented. Consider that there are n funds under evaluation. Let T denote the number of consecutive times in a given time horizon of interest. In addition, define a multi-time path of inputs and outputs as $\mathbf{V}_j^{t,T} = (x_j^k, y_j^k)_{k=t}^{T+t-1}$ for MF j, $(j = 1, \ldots, n)$, where $x_j^k = (x_{1j}^k, \ldots, x_{mj}^k)$ and $y_j^k = (y_{1j}^k, \ldots, y_{sj}^k)$ represent m inputs and s outputs at time k, respectively. Assuming VRS for all times k and strong free disposability of all inputs and outputs, the multi-time PPS with convexity and nonconvexity can be defined as:

$$\mathbf{P}_\Lambda^{t,T} = (P_\Lambda^k)_{k=t}^{T+t-1} \subset (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m \times T} \times \mathbb{R}^{s \times T}, \tag{3.2.8}$$

From the definition of the PPS in the multi-time framework, it is possible for each fund to reduce inputs and expand outputs over multiple time periods in this overall time horizon simultaneously.

Before defining the multi-time shortage function in a general case, we further simplify some symbolic expressions. Since the direction vector is assumed to be time-dependent, we define that $\mathbf{G}^{t,T} = (g^k)_{k=t}^{T+t-1} \in \mathbb{R}_-^{m\times T} \times \mathbb{R}_+^{s\times T}$ represents a given multi-time direction path, where $g^k = (-g_x^k, g_y^k) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ represents the direction vector for time k. In addition, we denote $\boldsymbol{\beta} = (\beta_k)_{k=t}^{T+t-1}$. Considering the time preference of an investor in a portfolio context, we introduce a time discounting factor denoted ξ $(0 < \xi < 1)$ to weight the efficiencies over the time horizon. Then, for any multi-time path of a fund denoted as $\mathbf{V}^{t,T} = (x^k, y^k)_{k=t}^{T+t-1}$, the discounted multi-time shortage functions with convexity and nonconvexity can be defined as follows:

$$S_\Lambda^{t,T}(\mathbf{V}^{t,T}; \mathbf{G}^{t,T}) = sup\Big\{\frac{1}{T}\sum_{k=t}^{T+t-1} \xi^{T+t-1-k}\beta_k : \mathbf{V}^{t,T} + \boldsymbol{\beta}\mathbf{G}^{t,T} \in P_\Lambda^{t,T}\Big\}, \qquad (3.2.9)$$

The definition of a multi-time shortage function does the same as a single-time shortage function, but over multiple time periods in a given time horizon: for a given overall time horizon, this amounts to looking for an arithmetic mean of simultaneous reductions in inputs and expansions in outputs into a path of direction $\mathbf{G}^{t,T}$ over all possible times. If the multi-time shortage function value $S_\Lambda^{t,T}(\mathbf{V}^{t,T}; \mathbf{G}^{t,T}) > 0$ for the input-output path $\mathbf{V}^{t,T}$ of the fund being evaluated, then it means that its inputs and outputs can be simultaneously improved in single or multiple times in this given overall time horizon.

Using the above multi-time shortage function, the multi-time Luenberger portfolio production indicators are further defined. Similarly, an adaptation of Definition (3.2.9) of the multi-time shortage function is required. This corresponding definition can be extended as follows:

Definition 3.2.2. *Given two initial times a and b, for any fund $\boldsymbol{V}^{a,T}$ in the direction of vector $\boldsymbol{G}^{a,T}$, the cross-period shortage functions with and without convexity*

are uniformly defined as:

$$S_{\Lambda}^{b,T}(\boldsymbol{V}^{a,T};\boldsymbol{G}^{a,T}) = sup\Big\{\frac{1}{T}\sum_{k=1}^{T}\beta_k : \boldsymbol{V}^{a,T} + \boldsymbol{\beta}\,\boldsymbol{G}^{a,T} \in \boldsymbol{P}_{\Lambda}^{b,T}\Big\}, \tag{3.2.10}$$

Similarly, these two times with a fixed time interval τ ($\tau \geq 0$) are denoted as: $(a,b) \in \{t, t+\tau\} \times \{t, t+\tau\}$.

To measure the evolution of a fund's multi-time performance, we also assume τ ($\tau \geq 0$) denotes the time interval between the different times of interest. Therefore, the multi-time Luenberger portfolio productivity indicator for measuring the change of the multi-time shortage function over time can be defined as:

$$\begin{aligned} L_{\Lambda}(\mathbf{V}^{t,T},\mathbf{V}^{t+\tau,T};\mathbf{G}^{t,T},\mathbf{G}^{t+\tau,T}) =& \frac{1}{2}[(S_{\Lambda}^{t,T}(\mathbf{V}^{t,T};\mathbf{G}^{t,T}) \\ &- S_{\Lambda}^{t,T}(\mathbf{V}^{t+\tau,T};\mathbf{G}^{t+\tau,T})) \\ &+ (S_{\Lambda}^{t+\tau,T}(\mathbf{V}^{t,T};\mathbf{G}^{t,T}) \\ &- S_{\Lambda}^{t+\tau,T}(\mathbf{V}^{t+\tau,T};\mathbf{G}^{t+\tau,T}))] \end{aligned} \tag{3.2.11}$$

Equally so, the multi-time Luenberger portfolio productivity indicator can decomposed as follows:

$$\begin{aligned} L_{\Lambda}(\mathbf{V}^{t,T},\mathbf{V}^{t+\tau,T};\mathbf{G}^{t,T},\mathbf{G}^{t+\tau,T}) =& E_{\Lambda}(\mathbf{V}^{t,T},\mathbf{V}^{t+\tau,T};\mathbf{G}^{t,T},\mathbf{G}^{t+\tau,T}) \\ &+ F_{\Lambda}(\mathbf{V}^{t,T},\mathbf{V}^{t+\tau,T};\mathbf{G}^{t,T},\mathbf{G}^{t+\tau,T}) \end{aligned} \tag{3.2.12}$$

with

$$\begin{aligned} E_{\Lambda}(\mathbf{V}^{t,T},\mathbf{V}^{t+\tau,T};\mathbf{G}^{t,T},\mathbf{G}^{t+\tau,T}) =& S_{\Lambda}^{t,T}(\mathbf{V}^{t,T};\mathbf{G}^{t,T}) \\ &- S_{\Lambda}^{t+\tau,T}(\mathbf{V}^{t+\tau,T};\mathbf{G}^{t+\tau,T}) \end{aligned} \tag{3.2.13}$$

and

$$\begin{aligned} F_\Lambda(\mathbf{V}^{t,T}, \mathbf{V}^{t+\tau,T}; \mathbf{G}^{t,T}, \mathbf{G}^{t+\tau,T}) = & \frac{1}{2}[(S_\Lambda^{t+\tau,T}(\mathbf{V}^{t+\tau,T}; \mathbf{G}^{t+\tau,T}) \\ & - S_\Lambda^{t,T}(\mathbf{V}^{t+\tau,T}; \mathbf{G}^{t+\tau,T})) \\ & + (S_\Lambda^{t+\tau,T}(\mathbf{V}^{t,T}; \mathbf{G}^{t,T}) \\ & - S_\Lambda^{t,T}(\mathbf{V}^{t,T}; \mathbf{G}^{t,T}))] \end{aligned} \tag{3.2.14}$$

In the above decomposition, $E_\Lambda(\,\cdot\,)$ measures the change on the multi-time efficiency that integrates several considered times over time, while $F_\Lambda(\,\cdot\,)$ captures the shift with respect to the frontiers of these sub times. More precisely, the efficiency change component measures the evolution of the weighted (discounted) multi-time efficiency of the fund relative to a changing multi-time frontiers from one time to the next. The frontier change component provides a local measure of the shift in the weighted value of multiple frontier movements induced by market volatility. The Luenberger indicator and its decomposition provide a unique tool for evaluating the relative success of different portfolio strategies implemented by fund managers over time. In fact, the efficiency change component $E_\Lambda(\,\cdot\,)$ of this multi-time Luenberger has an obvious use to track the performance evolution of a fund manager who, on average, adheres to a specified time-dependent risk profile and management style. It has the potential to serve for detecting the eventual ability of fund managers to generate superior performance, as this measure is not affected by the change in the financial market. Overall, the multi-time Luenberger portfolio productivity indicator represents the evolution of funds' multi-time performance. If the value of the multi-time Luenberger indicator (or its two components) is less than zero, then it indicates that multi-time performance of funds is decreasing. A value greater than zero implies a progression in the multi-time performance.

Next, we develop the computations of the multi-time Luenberger portfolio

productivity indicator for an observed fund $o(o \in 1,\ldots,n)$. As in the computation for the single-time Luenberger productivity indicator, the four different multi-time shortage functions need to be computed, that is, $S_{\Lambda}^{t,T}(\mathbf{V}_o^{t,T};\mathbf{G}_o^{t,T})$, $S_{\Lambda}^{t,T}(\mathbf{V}_o^{t+\tau,T};\mathbf{G}_o^{t+\tau,T})$, $S_{\Lambda}^{t+\tau,T}(\mathbf{V}_o^{t+\tau,T};\mathbf{G}_o^{t+\tau,T})$, and $S_{\Lambda}^{t+\tau,T}(\mathbf{V}_o^{t,T};\mathbf{G}_o^{t,T})$. For a given multi-time direction path $\mathbf{G}_o^{t,T}=(g_o^k)_{k=t}^{T+t-1}\in\mathbb{R}_-^{m\times T}\times\mathbb{R}_+^{s\times T}$, one must solve the following programs to obtain the own-period shortage function in the multi-time framework from a initial time t:

$$
\begin{aligned}
\max\quad & \frac{1}{T}\sum_{k=t}^{T+t-1}\xi^{T+t-1-k}\beta_k \\
\text{s.t.}\quad & \sum_{j=1}^{n}\lambda_j^k x_{ij}^k \le x_{io}^k-\beta_k g_{io}^k, \\
& i=1,\ldots,m,\quad k=t,\ldots,T+t-1, \\
& \sum_{j=1}^{n}\lambda_j^k y_{rj}^k \ge y_{ro}^k+\beta_k g_{ro}^k, \\
& r=1,\ldots,s,\quad k=t,\ldots,T+t-1, \\
& \sum_{j=1}^{n}\lambda_j^k=1, \\
& \forall j=1,\ldots,n:\begin{cases}\lambda_j^k\ge 0,\text{under convextiy}\\ \lambda_j^k\in\{0,1\},\text{under nonconvextiy}\\ k=t,\ldots,T+t-1,\end{cases}
\end{aligned}
\tag{3.2.15}
$$

Based on model (3.2.15), for the computation of the cross-period shortage function $S_{\Lambda}^{t,T}(\mathbf{V}_o^{t+\tau,T};\mathbf{G}_o^{t+\tau,T})$, one simply replaces the right-hand side of the first two constraints by the multi-time path of inputs and outputs $\mathbf{V}_o^{t+\tau,T}=(x_o^k,y_o^k)_{k=t+\tau}^{T+t+\tau-1}$ of the evaluated fund o and also the corresponding multi-path of direction vector $\mathbf{V}_o^{t+\tau,T}=(g_o^k)_{k=t+\tau}^{T+t+\tau-1}$.

Over a fixed time interval τ, for a given multi-time direction path $\mathbf{G}_o^{t+\tau,T}=$

$(g_o^k)_{k=t+\tau}^{T+t+\tau-1} \in \mathbb{R}_-^{m\times T} \times \mathbb{R}_+^{s\times T}$, the corresponding own-period shortage function from a initial time $t+\tau$ can be computed by the following program:

$$
\begin{aligned}
\max \quad & \frac{1}{T} \sum_{k=t+\tau}^{T+t+\tau-1} \xi^{T+t+\tau-1-k} \beta_k \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j^k x_{ij}^k \leq x_{io}^k - \beta_k g_{io}^k, \\
& i = 1, \ldots, m, \quad k = t+\tau, \ldots, T+t+\tau-1, \\
& \sum_{j=1}^{n} \lambda_j^k y_{rj}^k \geq y_{ro}^k + \beta_k g_{ro}^k, \\
& r = 1, \ldots, s, \quad k = t+\tau, \ldots, T+t+\tau-1, \\
& \sum_{j=1}^{n} \lambda_j^k = 1, \\
& \forall j = 1, \ldots, n : \begin{cases} \lambda_j^k \geq 0, \text{under convextiy} \\ \lambda_j^k \in \{0,1\}, \text{under nonconvextiy} \\ k = t+\tau, \ldots, T+t+\tau-1, \end{cases}
\end{aligned}
\tag{3.2.16}
$$

Based on model (3.2.16), the cross-period shortage function $S_\Lambda^{t+\tau,T}(\mathbf{V}_o^{t,T}; \mathbf{G}_o^{t,T})$ can be computed by replacing the right-hand side of the first two constraints by the multi-time path of inputs and outputs $\mathbf{V}_o^{t,T} = (x_o^k, y_o^k)_{k=t}^{T+t-1}$ of the evaluated fund o and also the corresponding multi-path of direction vectors $\mathbf{G}_o^{t,T} = (g^k)_{k=t}^{T+t-1}$.

When the multi-time Luenberger indicator is placed in the multi-moment (MVS or MVSK) framework, then the variance and the kurtosis of each sub time k are selected as inputs, and the expected return and the skewness are set as outputs, whereas for the MVS case only variance of each k is considered as an input. The specific shortage functions with these given inputs and outputs over multiple time periods in a given time horizon cross-period are applied to compute the multi-time and multi-moment Luenberger portfolio productivity indicators.

Next, we employ the actual fund data to compare the proposed multi-time and multi-moment Luenberger indicators with the basic single-time MV Luenberger indicators to illustrate the impact of multiple moments and multiple time periods on the values of indicators, as well as the rating determined by them. More importantly, we further explore the potential benefits of the proposed Luenberger indicators for fund selection by using a backtesting approach. Before these key issues are tested in the empirical analysis, we specify the backtesting framework in discrete time designed in this contribution.

3.3 Backtesting Framework in Dynamic Analysis

Our main objective is to explore the potential benefits of the multi-time and multi-moment Luenberger portfolio productivity indicators for fund ratings and selection in discrete time. Towards this end, we adopt a backtesting approach to compare the buy-and-hold strategies constructed by various fund rating methods. Backtesting refers to executing fictitious investment strategies using historical data to simulate how these strategies would have performed if they had actually been adopted by MF managers in the past. It is powerful for evaluating and comparing the performance of different investment strategies without using real capital. Some examples of a backtesting approach are found in DeMiguel, Garlappi, and Uppal (2009), Tu and Zhou (2011), Brandouy, Kerstens, and Van de Woestyne (2015), Zhou, Xiao, Jin, and Liu (2018) and Lin and Li (2020), among others.

In our backtesting analysis, we not only compare 12 different Luenberger portfolio productivity indicators computed in different rating frameworks, but also contrast those to 3 traditional performance appraisal tools: Sharpe ratio, Sortino ratio and Omega ratio. The exact definition for the Sharpe, Sortino and Omega

ratios can be found in Feibel (2003, p. 187 and p. 200) and Eling and Schuhmacher (2007, p. 2635), respectively. Table 3.1 lists the rating methods considered in our work and their abbreviations.

Table 3.1: List of various rating models considered for computing Luenberger portfolio productivity indicators

Models	Abbreviation
Traditional financial measures	
Sharpe ratio	Sharpe
Sortino ratio	Sortino
Omega ratio	Omega
Convex Frontier Rating	
Single-time and MV framework	STMVc
Single-time and MVS framework	STMVSc
Single-time and MVSK framework	STMVSKc
Multi-time and MV framework	MTMVc
Multi-time and MVS framework	MTMVSc
Multi-time and MVSK framework	MTMVSKc
Nonconvex Frontier Rating	
Single-time and MV framework	STMVnc
Single-time and MVS framework	STMVSnc
Single-time and MVSK framework	STMVSKnc
Multi-time and MV framework	MTMVnc
Multi-time and MVS framework	MTMVSnc
Multi-time and MVSK framework	MTMVSKnc

The first column of this table gives the 3 traditional financial measures and the 12 Luenberger portfolio productivity indicators under different frameworks, and the last column gives the abbreviation used to represent these rating methods.

For example, MTMVSKnc refers to the nonconvex (nc) frontier model with the means, variances, skewness and kurtosis (MVSK) of multiple time (MT) periods. It is noted that all the empirical results on the Luenberger portfolio productivity indicators and their corresponding decompositions are reported using the above abbreviations.

Here, we consider a simple buy-and-hold backtesting strategy consisting of buying and holding 10, 20 and 30 best performing HFs ranked by the rating methods. In particular, we design three selecting (buying) scenarios: (i) HFs are selected depending on the ranking of Luenberger portfolio productivity indicator; (ii) HFs are selected depending on the ranking of Luenberger efficiency change component; (iii) HFs are selected depending on the ranking of Luenberger frontier change component. Following Brandouy, Briec, Kerstens, and Van de Woestyne (2010), the Luenberger portfolio productivity indicator captures essential features of the whole return distribution for funds, and positive values in this indicator reflect that the fund performance increases over time. With regard to its components, the efficiency change component in principle detects the eventual ability of fund managers for stock picking and market timing, and the positive values in this component capture the positive contribution of fund managers to the performance evolution. The frontier change component measures the local changes in the frontier movements induced by market volatility, and the positive values in this component indicate the improvements of financial market performance.

For each of the above selecting scenarios, we set up four holding scenarios: hold the selected funds till the end of the whole sample period, for 1 year, for 3 years and for 5 years. The performance of a strategy held for 1, 3 and 5 years can be considered as its short, medium and long term holding performance, which are tested for the purpose to develop a sensitivity analysis regarding the out-of-sample

performance of this strategy.

Now, our work is to empirically test the out-of-sample performance of these indicator-driven buy-and-hold strategies. Since the Sharpe ratio and other relative performance measures are only suitable for the MV world, we opt for the shortage function as an absolute performance measures that is capable to assess the performance of these strategies in multiple dimensions simultaneously (i.e., mean, variance, skewness and kurtosis). Hence, the 15 buy-and-hold backtesting strategies are compared based on the MVSK performance of their holding values evaluated by combining shortage functions with the single-time and multi-moment frontiers (with convexity and nonconvexity), denoted as VRSc and VRSnc, respectively.

Based on the fundamental logic of backtesting, we design a backtesting analysis in detail for the various buy-and-hold strategies depending on the ranking of Luenberger portfolio productivity indicators. To empirically examine the out-of-sample performance of the backtesting strategies, we first collect 187 active HFs with monthly returns from August 2006 to October 2020. The detailed description for these sample HFs is presented in the following empirical section. Then, our backtesting analysis is performed multiple times replying on rolling the time windows. Specially, we split the period from the beginning of the sample period to the end of October 2015 in time windows of a given length, where the 5 years before the end of the sample period are maintained to test the long-term holding performance of these strategies in the last backtesting.

The length of this rolling time window is set to 63 months in our work. In each 63-month time window, we assume that there exist two different times with a fixed time interval $\tau = 3$ such that they are separated by 3 months and that decision-makers are able to observe monthly data for the last 60 months at the first and

second times, respectively. The observed historical returns at the first and second times are used to estimate the parameters of the two different times, respectively, so as to obtain the Luenberger portfolio productivity indicators and their decompositions of the observed HFs over 3 months. Such a rolling time window is designed mainly due to two concerns. First, the longest time period considered in the proposed multi-time performance measures is 5 years (60 months), so the basic time periods should at least equal the longest evaluation horizon. Second, considering the fact that fund managers usually adjust their portfolio strategies every three or six months, it is appropriate to set the interval between two different times as 3 months.

Our backtesting analysis is developed starting from November 2011, and is repeated 48 times (each time a month) with the 63-month rolling time windows till October 2015. Using the first time window of data (from August 2006 to October 2011, 63 months in total) to obtain the Luenberger portfolio productivity indicators and the corresponding decompositions for different frontier-based rating models, we determine the first buy-and-hold backtesting strategies in November 2011. These strategies are held until the end of October 2020 (the whole sample period), the end of October 2012 (for 1 year), the end of October 2014 (for 3 years), or the end of October 2016 (for 5 years). The process of the first backtesting is represented by Figure 3.1.

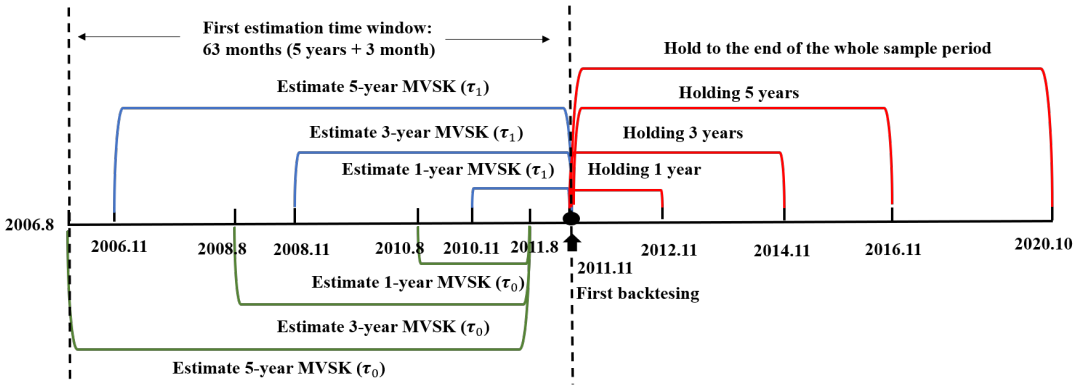

Figure 3.1: The process of the first backtesting in discrete framework

Then, the time window is shifted with a step of a single month to develop the next backtesting analysis. For each time window or each backtesting, the detailed steps are as follows:

(1) For a given 63-month time window, using the data of 60 months observed prior to two different times respectively, HF samples are evaluated with 4 different shortage functions generated by each frontier rating models. Based on these, the Luenberger portfolio productivity indicators and the corresponding decompositions are computed.

(2) Depending on the values and the decompositions of Luenberger indicator computed by this time window of data for each rating method, the 10, 20 or 30 best performing HFs are selected for the backtesting exercise.

(3) For each buy-and-hold scenario, we compute and store the complete historical track record of the holding value for per buy-and-hold backtesting strategy, and calculate the mean, variance, skewness and kurtosis of their holding value series.

The above steps for backtesting are repeated over 48 time windows in total. For the 3 traditional financial methods, we use monthly returns of the sample HFs over each 63-month rolling time window to compute the Sharpe ratio, Sortino ratio and Omega ratio respectively, whereby the corresponding buy-and-hold strategies based on these traditional financial ratings are constructed.

For each of the buy-and-hold period scenarios, the performance of these MVSK observations (15 times 48 observations) that are generated by the 15 strategies over 48 backtesting exercises are all evaluated by the shortage functions in the single-time and multi-moment frameworks (with convexity and nonconvexity). In particular, we first establish the convex and nonconvex VRS nonparametric frontiers in the single-time and multi-moment framework for these MVSK observations, and then measure their efficiency scores using the shortage functions. Clearly, each buy-and-hold strategy yields the efficiency scores of 48 MVSK observations. The average efficiency score and the number of efficient units, as well as the distribution of inefficiency scores across these 48 observations, are adopted to evaluate the 15 strategies.

3.4 Empirical Testing

In this section, our empirical analysis aims to examine two issues. First, we investigate the impact of three crucial factors on the Luenberger portfolio productivity indicators and their corresponding decompositions, as well as on the rankings determined by these indicators: (i) consideration of multiple time periods; (ii) inclusion of multiple moments;[5] (iii)imposition of convexity. Second, we further

[5]In our empirical testing, the 5-year MVSK of the return distribution for evaluated funds are estimated based on a 5-year estimation window covering 60 historical monthly returns. For the multi-time case, the historical monthly returns for the past 1 year and 3 years to date within this

examine the performance of the proposed rating methods based on the multi-time and multi-moment Luenberger indicators and their decompositions with respect to fund selection.

We apply 187 active HF samples with monthly returns from August 2006 to October 2020 as the test dataset to develop a backtesting exercise. The sample data is downloaded from the Hedge Funds database provided by Lipper for Investment Management. Since HF cost data is unavailable in this database, our frontier rating models used to establish Luenberger indicators are limited to focusing on the characteristics of the return distribution for these HFs, while ignoring the cost factors. As the backtesting framework designed in Section 3, the period from the beginning of the sample period to the end of October 2015 is split with a 63-month time window. As we introduced previously, the rolling tick for such time window is one month. Therefore, since we dispose of 111 months in the data set, we end up with 48 time windows. In each of 48 time windows, all 187 HF samples are evaluated with 4 different shortage functions computed by the 12 nonparametric frontier models (see Table 3.1). It needs to be mentioned that none of the computations for the cross-period evaluations suffers from the infeasibility problem mentioned here above.

Figure 3.2 plots the mean values of the Luenberger indicator (denoted L) and its components (denoted E and F) per frontier models for the 187 samples over the 48 time windows by means of the box-plots. The box indicates the interquartile range where a small horizontal line reports the location of the median. From these results, two things stand out. First, the sample mean values of Luenberger indicators and their corresponding components calculated per frontier models generally differ. Overall, these values obtained by the multi-time frontier models are relatively

estimation window are applied to estimate their 1-year MVSK and 3-year MVSK, respectively.

larger than those obtained from the single-time frontier models over the 48 time windows. Additionally, the inclusion of multiple moments seems to contribute to these larger value. Second, the overall distributions on sample mean values of these indicators are somewhat different, especially when the higher order moments are considered. These basic observations are further tested and discussed in the following subsection by means of the Li-test and Kendall rank correlations.

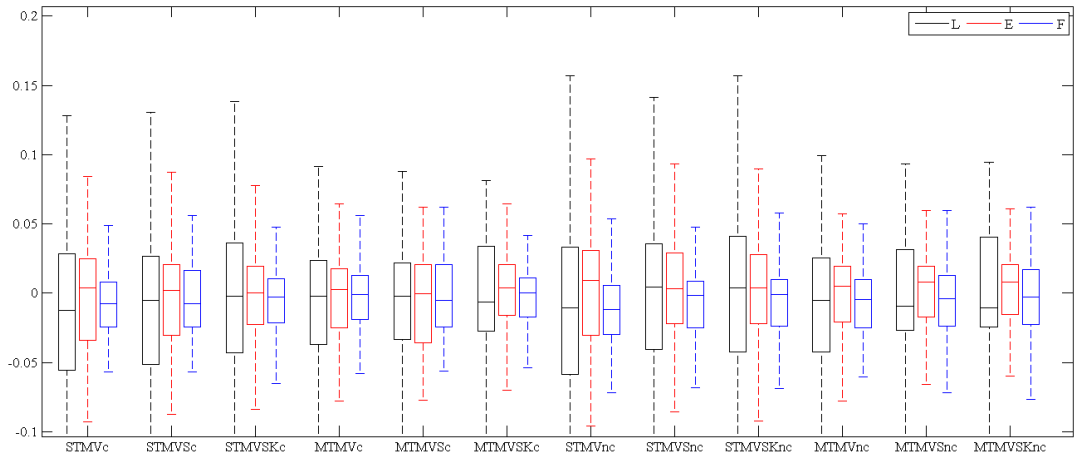

Figure 3.2: Luenberger decomposition distributions for 12 methods: sample mean values

In addition, note that in our backtesting exercises, the selection of the 30 best rated funds is the case in which we select the largest number of best performing funds. To simply show that most of these selected funds have a positive indicator, we report now the third quartile (Q3) of the Luenberger indicator and its component computed by each frontier model for the 187 samples over 48 time windows. Looking at the results regarding these sample Q3 values in Figure 3.3, one can find that the sample Q3 values of these computed Luenberger indicators (or their components) over the 48 time windows are mostly positive and only exceptionally negative. Since the total number of values greater than or equal to Q3 is roughly

25% of the 187 samples (around 47, clearly more than 30), this indicates that funds selected in our backtesting exercises by the ranking of Luenberger indicators (or their components) have a positive Luenberger indicator (or efficiency change and frontier change components).

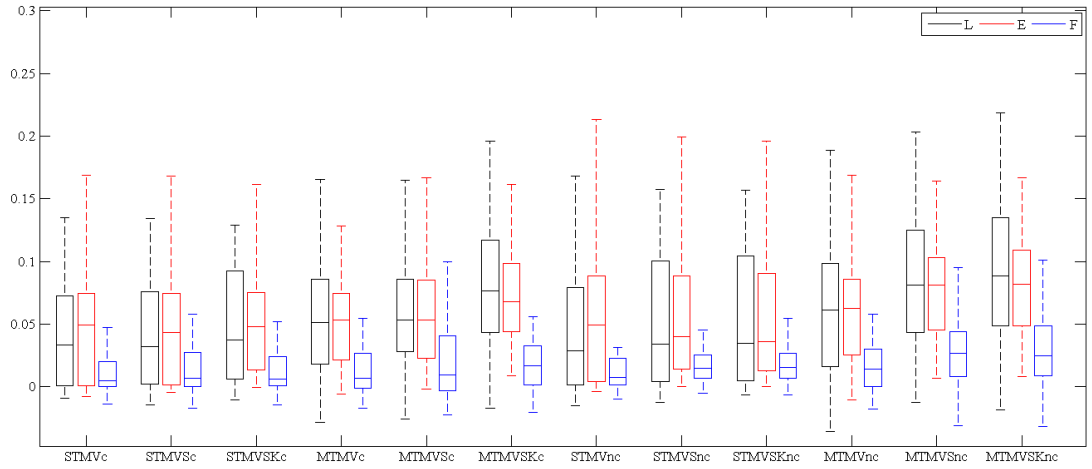

Figure 3.3: Luenberger indicators distributions for 12 methods: sample Q3 values

3.4.1 Evaluation Results Comparison

According to the results of the Luenberger portfolio productivity indicators and their components computed by the 12 nonparametric frontier models, we compare the distributions of these various Luenberger indicators and their components in each time window using a Li-test approach. The Li-test is a nonparametric test approach comparing two entire distributions initially developed by Li (1996) and refined by Fan and Ullah (1999) and most recently by Li, Maasoumi, and Racine (2009). It tests for the eventual statistical significance of differences between two kernel-based estimates of density functions f and g of a random variable x. The null hypothesis maintains the equality of both density functions almost everywhere:

$H_0 : f(x) = g(x)$ for all x; while the alternative hypothesis negates this equality of both density functions: $H_1 : f(x) \neq g(x)$ for some x.[6]

In each of 48 time windows, the Li-test are computed between different Luenberger indicators, as well as between their corresponding components, which contains in total 66 relevant model comparisons. For reasons of readability, we aggregate the results of the Li-test over 48 time windows, and simply report the times that the distributions between the Luenberger portfolio productivity indicators or between their components are significantly different at a 5 % significance level in Table 3.2 to 3.4

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

Table 3.2: Li-test comparing $L_\Lambda(\cdot)$ distributions computed by different rating models

	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
STMVc	0	9	40	47	48	0	13	13	42	48	48
STMVSc		0	39	45	48	15	2	3	38	48	48
STMVSKc			34	41	40	18	1	2	39	43	45
MTMVc				2	33	41	39	39	1	45	42
MTMVSc					12	46	41	43	5	29	34
MTMVSKc						48	39	39	25	1	2
STMVnc							4	9	41	48	48
STMVSnc								0	38	46	42
STMVSKnc									37	45	43
MTMVnc										27	31
MTMVSnc											0

Table 3.3: Li-test comparing $E_\Lambda(\cdot)$ distributions computed by different rating models

	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
STMVc	0	4	42	44	46	7	31	37	41	47	46
STMVSc		0	39	40	43	22	37	45	35	39	39
STMVSKc			41	46	45	13	17	38	41	38	35
MTMVc				3	22	44	45	47	4	31	33
MTMVSc					15	46	47	48	14	24	26
MTMVSKc						47	48	48	22	2	3
STMVnc							6	14	45	45	46
STMVSnc								0	46	48	48
STMVSKnc									48	48	48
MTMVnc										9	12
MTMVSnc											0

Table 3.4: Li-test comparing $F_\Lambda(\cdot)$ distributions computed by different rating models

	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
STMVc	0	10	47	48	48	26	47	44	48	48	48
STMVSc		9	47	46	48	35	41	37	46	48	48
STMVSKc			45	45	45	33	32	33	42	47	47
MTMVc				17	34	46	48	46	22	48	48
MTMVSc					30	48	46	48	34	47	46
MTMVSKc						48	47	47	21	36	44
STMVnc							12	13	46	48	48
STMVSnc								0	44	45	46
STMVSKnc									44	47	45
MTMVnc										36	38
MTMVSnc											0

Several observations can be made from results in Table 3.2 regarding the Luenberger portfolio production indicators. First, it is clear that the distributions between the single-time and multi-time Luenberger portfolio productivity indicators yield significantly different overall. The basic statistics for these Li-test computations over 48 time windows show that the single-time and multi-time Luenberger indicators exhibit significantly different distributions at a 5 % significance level in almost all cases. This reveals that the consideration of multiple time periods has a significant impact on the Luenberger indicators for all cases in the MV, MVS and MVSK rating framework with both convexity and nonconvexity.

Second, the effects of adding higher order moments on the single-time and multi-time Luenberger indicators are rather different. For instance, the Luenberger indicators in the STMVnc vs. STMVSnc framework show significantly different distributions only 4 times within Li-test computations over 48 time windows, whereas the Luenberger indicators in the MTMVnc vs. MTMVSnc framework significantly different distributions 27 times. In addition, higher order moments tend to have a stronger influence on calculating these indicators by nonconvex frontier rating models than by convex frontier rating models. As we can observe in Table 3.2, the significant difference between the Luenberger indicators in the MTMVc vs. MTMVSc framework emerges 2 times over the 48 time windows, while the significant difference these indicators calculated in the MTMVnc vs. MTMVSnc framework arises 27 times.

Third, the Luenberger productivity indicators in the convex and nonconvex cases always exhibit a significant difference in multi-time MVS rating framework over the 48 time windows. This coincides with the above finding that the bias of applying the convexification Luenberger indicators is evident when considering multiple horizons and multiple moments for our data. To some extent, noncon-

vexity may prove to be a more modest and realistic hypothesis to be maintained in the proposed multi-time and multi-moment Luenberger portfolio productivity indicator.

Equally so, the distributions of the two components $E_\Lambda(\cdot)$ and $F_\Lambda(\cdot)$ of these 12 Luenberger indicators are compared using Li-test at a 5 % significance level, respectively. The clustering results computed in the 48 time windows are shown in Tables 3.3 and 3.4. Regarding the impacts of these three factors (ST vs. MT, MV vs. MVS and MVSK, and convexity vs. nonconvexity) on these two components, one can clearly first find that the impact of multiple times is significant on both the efficiency change component $E_\Lambda(\cdot)$ and the frontier change component $F_\Lambda(\cdot)$. Second, the effect of adding multiple moments on the two components is clearly different. This effect is somewhat more modest on the distribution of $E_\Lambda(\cdot)$ than on that of $F_\Lambda(\cdot)$. For instance, in the multi-time and nonconvex framework, adding skewness and kurtosis contributes a significant effect on the distribution of $F_\Lambda(\cdot)$ occurring 38 times in total, while this emerges only 7 times in the case of $E_\Lambda(\cdot)$. Third, imposing convexity leads to a stronger influence on the distributions of $F_\Lambda(\cdot)$ compared to the distributions of $E_\Lambda(\cdot)$. The reason for this finding is that the frontier change component $F_\Lambda(\cdot)$ captures the changes in frontiers. It is possible that the extent and direction of such changes are significantly different under both convex and nonconvex nonparametric frontiers.

Table 3.5: Kendall rank correlations comparing $L_\Lambda(\,\cdot\,)$ computed by different rating methods

	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
STMVc	0.886	0.670	0.434	0.380	0.261	0.801	0.520	0.510	0.407	0.236	0.225
STMVSc		0.774	0.412	0.390	0.288	0.726	0.582	0.568	0.387	0.261	0.251
STMVSKc			0.329	0.333	0.316	0.568	0.649	0.652	0.308	0.283	0.280
MTMVc				0.807	0.503	0.404	0.287	0.279	0.848	0.446	0.427
MTMVSc					0.671	0.354	0.289	0.283	0.725	0.588	0.567
MTMVSKc						0.245	0.270	0.275	0.469	0.763	0.771
STMVnc							0.596	0.571	0.411	0.236	0.225
STMVSnc								0.870	0.289	0.300	0.288
STMVSKnc									0.280	0.296	0.303
MTMVnc										0.462	0.440
MTMVSnc											0.896

Table 3.6: Kendall rank correlations comparing $E_\Lambda(\,\cdot\,)$ computed by different rating methods

	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
STMVc	0.879	0.643	0.420	0.362	0.252	0.713	0.456	0.442	0.384	0.229	0.220
STMVSc		0.729	0.385	0.367	0.272	0.640	0.502	0.487	0.350	0.247	0.239
STMVSKc			0.291	0.291	0.306	0.510	0.594	0.596	0.267	0.272	0.270
MTMVc				0.768	0.478	0.350	0.228	0.220	0.785	0.412	0.396
MTMVSc					0.629	0.301	0.226	0.221	0.650	0.533	0.519
MTMVSKc						0.211	0.232	0.236	0.433	0.718	0.722
STMVnc							0.594	0.563	0.374	0.221	0.211
STMVSnc								0.873	0.239	0.285	0.272
STMVSKnc									0.230	0.276	0.287
MTMVnc										0.446	0.427
MTMVSnc											0.899

Table 3.7: Kendall rank correlations comparing $F_\Lambda(\,\cdot\,)$ computed by different rating methods

	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
STMVc	0.850	0.610	0.296	0.217	0.194	0.575	0.331	0.315	0.231	0.124	0.118
STMVSc		0.648	0.271	0.249	0.214	0.502	0.348	0.334	0.207	0.143	0.138
STMVSKc			0.220	0.186	0.285	0.429	0.437	0.427	0.188	0.173	0.166
MTMVc				0.727	0.531	0.212	0.130	0.126	0.601	0.304	0.294
MTMVSc					0.532	0.153	0.111	0.104	0.487	0.300	0.288
MTMVSKc						0.163	0.179	0.182	0.438	0.456	0.443
STMVnc							0.510	0.471	0.293	0.171	0.159
STMVSnc								0.800	0.180	0.270	0.237
STMVSKnc									0.175	0.252	0.274
MTMVnc										0.410	0.393
MTMVSnc											0.762

We now move to the comparison for these various computed Luenberger ratings, as well as the ratings calculated by their components. For each of the 48 time windows, we compute a Kendall rank correlation among these 12 ratings, which delivers a matrix of 12 times 12 dimensions. Then, we have aggregated the Kendall rank correlations using a simple arithmetic mean of the computations in 48 time windows to report their overall degree of concordance in ranking. The latter results regarding the Luenberger productivity indicators and their components are reported in Tables 3.5 to 3.7. It is noted that the correlation coefficients between the different rankings are significantly different from zero at 1 % significance level for each time window calculation.

Looking at Table 3.5, one can draw three main conclusions. First, the single-time Luenberger rating and multi-time Luenberger rating show a low correlation on average (around 0.239-0.440 across all observations). Second, the Luenberger-driven ratings with MV moments exhibit a lower average correlation with the Luenberger-driven ratings with multi-moments (MVS & MVSK) in a multi-time framework compared to in a single-time framework. Furthermore, the MV vs. MVS & MVSK ratings exhibit a lower average correlation in the nonconvex frontier case compared to the convex frontier case. Lastly, the Luenberger ratings depending on the convex frontier models show a lower average correlation with those depending on the nonconvex frontier models in a multi-time and multi-moment framework compared to other frameworks. As can be observed from Table 3.6 and 3.7, the above conclusions are also revealed with respect to the ratings driven by the efficiency change component $E_{\Lambda}(\cdot)$ and the ratings driven by the frontier change component $F_{\Lambda}(\cdot)$.

To a large extent, these findings obtained from the Kendall ranking correlations are compatible with those obtained from the Li-test. Summarizing the above dis-

cussion so far, we can conclude that the Luenberger indicators and their components constructed by the multi-time and multi-moment performance measures show quite a difference from those constructed by the original single-time MV performance measures. This difference between these indicators is more pertinent under nonconvexity than under convexity. Finally, with regard to the components of the Luenberger indicators, adding multiple moments and imposing convexity seems to show a stronger effect on the frontier change than on the efficiency change.

3.4.2 Backtesting results

Following the backtesting framework in discrete time described in Section 3, we perform the backtesting analysis with the selection of 10, 20 or 30 best performing HFs ranked by three traditional financial measures and various frontier-based Luenberger indicators.[7] The same pattern of backtesting analysis is also applied to compare the strategies depending on the ranking of the two Luenberger components, respectively. Tables 3.8 to 3.10 present an overall analysis with respect to the performances of the MVSK observations generated by the buy-and-hold strategy held until the end of the entire sample period (October 2020). Specifically, Table 3.8 list the performance results per buy-and-hold strategy determined by the ranking of the Luenberger indicator ($L_\Lambda(\,\cdot\,)$), and Tables 3.9 and 3.10 report the results per strategy depending on the rankings of its efficiency change component ($E_\Lambda(\,\cdot\,)$) and frontier change component ($F_\Lambda(\,\cdot\,)$), respectively. Each table is structured as follows: The first series of four columns list the results with regard to the 10 best HFs selected for the backtesting exercise, and the second and third series of four columns present the results for selecting 20 and 30 best

[7] All funds with the same value of Luenberger indicator are ranked the same in our calculations. As a consequence, when we have to take a certain amount of funds among these ties then we take these randomly among the tied units.

HFs, respectively. Within each selecting (buying) scenario, the first two columns report the average inefficiency scores and the number of efficient units for each strategy when evaluated using the convex VRS frontier in single-time and multi-moment framework (VRSc), while the last two columns report these results in the nonconvex case (VRSnc).

Table 3.8: Performance results for 15 buy-and-hold backtesting strategies depending on $L_\Lambda(\,\cdot\,)$ (3 financial rating methods and 12 Luenberger rating methods): Descriptive statistics of the shortage function values

Methods	HF(10)				HF(20)				HF(30)			
	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
Sharpe	0.074	0	0.042	8	0.065	0	0.037	22	0.076	1	0.041	12
Sortino	0.068	1	0.043	14	0.069	1	0.041	20	0.073	0	0.033	14
Omega	0.072	0	0.044	14	0.075	0	0.044	21	0.081	0	0.043	13
STMVc	0.072	6	0.049	23	0.062	7	0.039	24	0.068	5	0.042	23
STMVSc	0.080	8	0.055	19	0.065	7	0.041	20	0.067	4	0.040	25
STMVSKc	0.076	4	0.040	22	0.064	4	0.033	21	0.062	2	0.035	21
MTMVc	0.070	6	0.036	17	0.064	6	0.033	20	0.069	7	0.032	13
MTMVSc	0.068	2	0.033	24	0.059	6	0.027	20	0.059	7	0.027	19
MTMVSKc	0.058	3	0.031	19	0.051	4	0.027	24	0.057	6	0.034	19
STMVnc	0.068	5	0.045	25	0.065	2	0.037	17	0.071	5	0.048	16
STMVSnc	0.085	5	0.057	15	0.069	3	0.035	16	0.074	3	0.044	18
STMVSKnc	0.078	4	0.043	19	0.066	3	0.033	19	0.072	4	0.042	14
MTMVnc	0.075	4	0.034	18	0.061	3	0.030	19	0.067	2	0.035	22
MTMVSnc	0.052	6	0.028	22	0.044	7	0.026	27	0.054	3	0.031	21
MTMVSKnc	0.050	8	0.025	25	0.045	8	0.027	21	0.053	4	0.031	20

Table 3.9: Performance results for 15 buy-and-hold backtesting strategies depending on $E_\Lambda(\cdot)$ (3 financial rating methods and 12 Luenberger rating methods): Descriptive statistics of the shortage function values

Methods	HF(10)				HF(20)				HF(30)			
	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
Sharpe	0.069	1	0.044	8	0.066	0	0.039	18	0.065	1	0.038	7
Sortino	0.065	0	0.044	12	0.070	1	0.041	15	0.062	0	0.037	9
Omega	0.067	0	0.046	13	0.072	0	0.042	18	0.066	1	0.041	9
STMVc	0.062	7	0.040	23	0.057	9	0.038	27	0.051	8	0.029	29
STMVSc	0.061	6	0.042	21	0.058	6	0.037	23	0.052	6	0.032	23
STMVSKc	0.073	3	0.053	17	0.058	6	0.040	21	0.052	3	0.037	21
MTMVc	0.065	8	0.034	23	0.055	5	0.029	18	0.050	6	0.023	24
MTMVSc	0.059	10	0.037	20	0.052	6	0.026	22	0.047	5	0.024	21
MTMVSKc	0.060	5	0.034	22	0.051	4	0.028	20	0.047	5	0.027	24
STMVnc	0.060	9	0.036	20	0.057	6	0.036	20	0.056	4	0.034	20
STMVSnc	0.061	5	0.044	21	0.056	3	0.033	20	0.056	4	0.036	16
STMVSKnc	0.062	6	0.040	22	0.060	3	0.035	17	0.057	1	0.039	14
MTMVnc	0.071	4	0.041	16	0.051	5	0.025	24	0.048	5	0.025	24
MTMVSnc	0.061	3	0.039	19	0.048	5	0.032	20	0.047	6	0.028	24
MTMVSKnc	0.055	7	0.033	19	0.047	7	0.032	21	0.048	5	0.030	21

Table 3.10: Performance results for 15 buy-and-hold backtesting strategies depending on $F_\Lambda(\cdot)$ (3 financial rating methods and 12 Luenberger rating methods): Descriptive statistics of the shortage function values

Methods	HF(10)				HF(20)				HF(30)			
	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
Sharpe	0.070	1	0.032	12	0.086	0	0.043	18	0.086	1	0.031	10
Sortino	0.069	1	0.035	15	0.089	1	0.042	18	0.082	0	0.032	14
Omega	0.071	0	0.033	14	0.096	0	0.052	19	0.087	0	0.039	13
STMVc	0.056	4	0.032	21	0.069	3	0.033	20	0.059	3	0.029	21
STMVSc	0.054	6	0.017	29	0.072	2	0.033	19	0.062	3	0.032	14
STMVSKc	0.071	5	0.036	18	0.072	1	0.029	18	0.051	5	0.022	20
MTMVc	0.064	4	0.033	21	0.069	5	0.044	19	0.057	4	0.037	21
MTMVSc	0.073	2	0.033	17	0.073	5	0.038	17	0.067	5	0.042	19
MTMVSKc	0.066	1	0.030	23	0.071	5	0.033	18	0.064	1	0.031	21
STMVnc	0.057	5	0.029	17	0.063	1	0.032	20	0.057	5	0.030	20
STMVSnc	0.060	3	0.028	27	0.062	6	0.033	17	0.052	3	0.024	21
STMVSKnc	0.050	7	0.027	25	0.060	3	0.032	14	0.053	3	0.027	18
MTMVnc	0.058	5	0.031	22	0.070	2	0.043	19	0.066	1	0.032	16
MTMVSnc	0.060	4	0.029	18	0.072	4	0.040	15	0.058	2	0.030	17
MTMVSKnc	0.067	2	0.034	17	0.066	1	0.033	17	0.058	2	0.029	16

Looking at Table 3.8, one can draw three main conclusions. First, the performance of the buy-and-hold strategies based on the ranking of the multi-time Luenberger indicators are superior to those based on the ranking of the traditional financial measures. From Table 3.8, it can be observed that the average inefficiency scores of these strategies driven by the multi-time Luenberger indicators are lower than those of Sharpe-, Sortino- and Omega-driven strategies, and the former yield more efficient units compared to the strategies driven by traditional financial ratings. This result is valid when buying the 10, 20 and 30 best HFs. From Tables 3.9 and 3.10, one can find that a similar conclusion emerges when comparing the strategies depending on the rankings of their efficiency change components or frontier change components with those depending on traditional measures. This somewhat reveals that the Luenberger portfolio productivity indicator is of potential in fund rating and selection in a discrete time framework.

Second, we turn to the comparison among the 12 various Luenberger indicator buy-and-hold strategies. Combining the average inefficiency scores and the number of efficient units, one can note that the strategies developed by the multi-time Luenberger ratings outperform those developed by the single-time Luenberger ratings. In most of the backtesting exercises, the multi-time Luenberger-based strategies yield lower average inefficiency scores and a higher number of efficient units compared to the single-time Luenberger-based strategies. When HF ratings are driven by the efficiency change component related to the individual skill of fund managers, the buy-and-hold strategies in the multi-time framework also exhibit superior performance compared to those in the single-time framework. When applying the frontier change component that reflects the market performance change to select best performing HFs, the advantage of the constructed buy-and-hold strategies in the multi-time framework is not significant.

Third, it is found that in the majority of cases the buy-and-hold strategies consisting of the HFs selected by the multi-moment (MVS & MVSK) Luenberger indicators show superior results compared to those consisting of the HFs selected by the MV Luenberger indicators. Moreover, this superiority tends to be more evident in the multi-time framework than in the single-time framework.

Equally so, this comparison is separately developed in the context of the efficiency change ratings and the frontier change ratings. In the later context, the buy-and-hold strategies in the multi-moment framework perform better than those in the original MV framework. However, this result is not validated for HF rating solely based on the efficiency change component.

Figures 3.4, 3.5, and 3.6 respectively provide the entire distribution of the inefficiency scores per strategy driven by Luenberger indicators and their components to compare these intuitively. In each figure, the sub-figures (a) to (c) correspond to the performance results of the backtesting scenarios with 10, 20 and 30 funds selected. Those based on the nonconvex VRS frontier are in red, while those based on the convex VRS frontier are in blue. In these box-plots, the box indicates the interquartile range where the small vertical lines reporting the location of the median. Their locations closer to the left suggests that the entire distribution of inefficiency scores for the strategy is at a lower level, which implies that the strategy has a better performance in backtesting analysis.

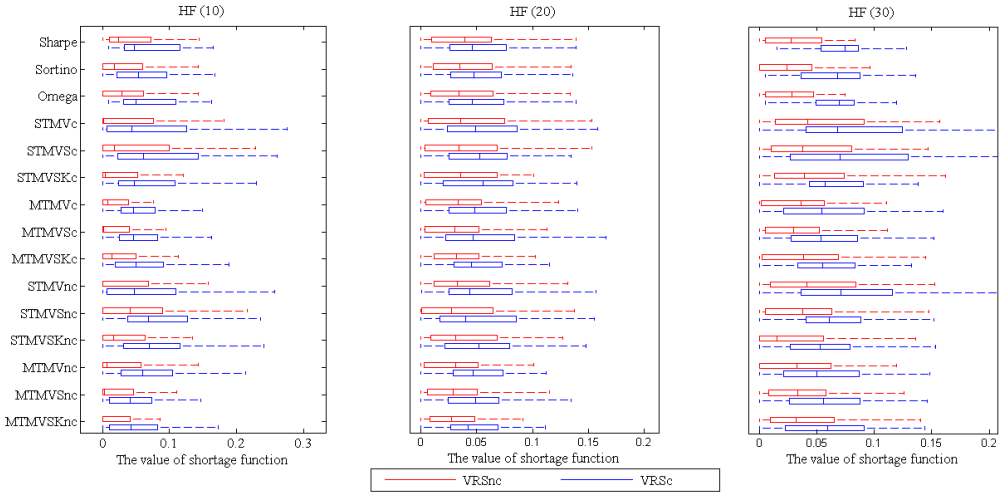

Figure 3.4: Inefficiency distributions for 15 buy-and-hold backtesting strategies driven by Luenberger indicators

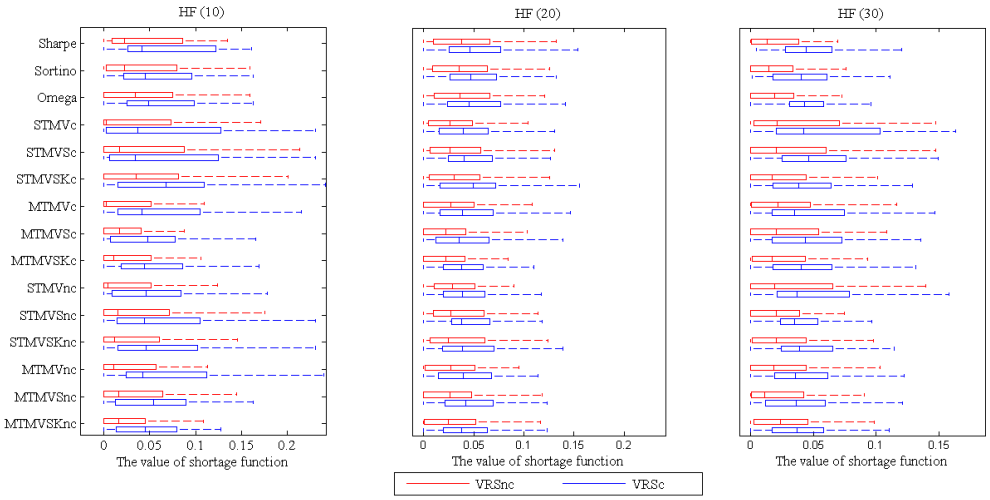

Figure 3.5: Inefficiency distributions for 15 buy-and-hold backtesting strategies driven by efficiency change components

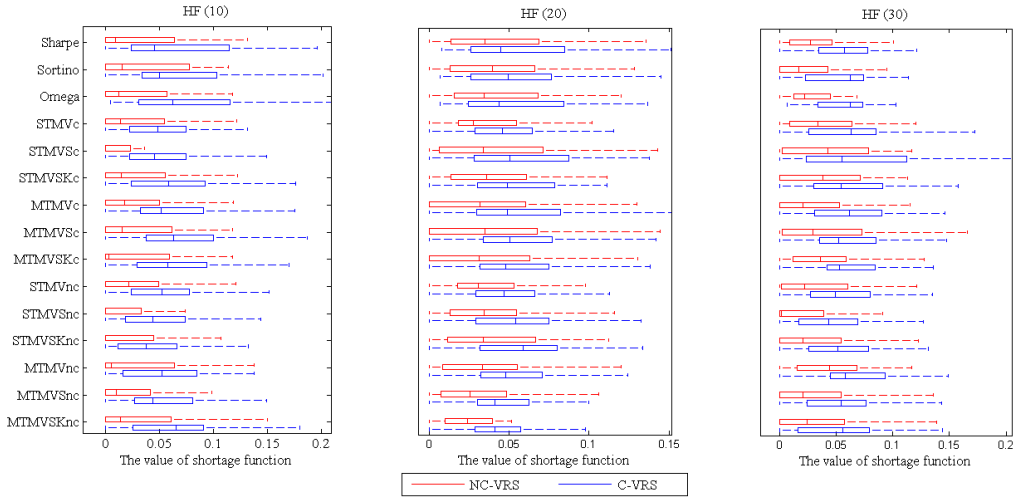

Figure 3.6: Inefficiency distributions for 15 buy-and-hold backtesting strategies driven by frontier change components

It is clear from Figure 3.4 that the buy-and-hold strategies constructed by the ranking of the multi-time and multi-moment Luenberger indicators are superior to strategies constructed by the ranking of financial measures and other existing Luenberger indicators in most cases. In terms of the contexts of efficiency change components and frontier change components, one can observe that the strategies constructed by the ranking of the two components also outperform those based on the financial ratings, respectively. When comparing the results in these two contexts, the finding that the multi-time Luenberger-based strategies are superior to the single-time Luenberger-based strategies is somewhat more robust in the efficiency change ratings, whereas the results that the multi-moment Luenberger-based strategies establish a dominance over the MV Luenberger-based strategies seems to be stronger in the frontier change ratings.

We therefore believe that the multi-time and multi-moment Luenberger portfolio productivity indicator gathers additional information regarding fund per-

formance changes to support the decision-making process of fund selection. Furthermore, the efficiency component of this proposed Luenberger indicator seems to better show the strengths of considering multiple times in fund selection, while the frontier change component tends to better capture the information from higher order moments.

For a more detailed analysis, we calculate the performance of these 15 buy-and-hold strategies held for 1 year, 3 years and 5 years, respectively. This can be regarded as testing the short-term holding performance, the medium-term holding performance and the long-term holding performance of an investment strategy. Tables 3.11 and 3.12 report the summarized results with respect to the performance per buy-and-hold backtesting strategy held for 1, 3 and 5 years , respectively.

Having the potential of the proposed multi-time and multi-moment Luenberger indicators in mind, we further analyse the short-, medium- and long-term performance of the buy-and-hold strategies determined by ranking this Luenberger indicators, as well as its two components. Looking at Tables 3.11 and 3.12, the above three findings drawn from Tables 3.8 to 3.10 are also evidence in most cases for the three holding period scenarios. Moreover, the buy-and-hold backtesting strategies consisting of the best HFs rated by the multi-moment and multi-time performance measure tend to show consistent performance over different holding periods. We basically conclude that the buy-and-hold strategies driven by the multi-moment and multi-time Luenberger ratings establish a dominance over other strategies and that this relation is strengthened as the holding period increases. It is somewhat revealed that the good performance of the strategies driven by the proposed Luenberger indicators with the inclusion of multiple moments and multiple times exhibit good stability.[8]

[8]In detail, the buy-and-hold backtesting strategies driven by the proposed multi-time and multi-moment Luenberger indicators have a stable superior performance over the other strategies.

Table 3.11: Performance results for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 Luenberger rating methods) held for 1, 3 and 5 years: Descriptive statistics of the shortage function values computed by convex VRS frontier

	Methods	Hold 1 year						Hold 3 years						Hold 5 years					
		L (·)		E (·)		F (·)		L (·)		E (·)		F (·)		L (·)		E (·)		F (·)	
		Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
	Sharpe	0.060	3	0.055	3	0.060	0	0.065	2	0.069	2	0.076	2	0.067	0	0.066	0	0.060	2
	Sortino	0.058	1	0.053	1	0.057	1	0.049	4	0.050	6	0.080	2	0.064	2	0.064	3	0.055	4
	Omega	0.054	3	0.050	3	0.054	2	0.052	0	0.054	0	0.070	0	0.066	0	0.066	0	0.058	0
	STMVc	0.068	2	0.060	2	0.062	2	0.073	4	0.073	4	0.076	4	0.081	8	0.074	5	0.057	3
	STMVSc	0.065	7	0.060	3	0.074	6	0.081	5	0.073	3	0.074	7	0.083	6	0.074	4	0.057	4
	STMVSKc	0.064	5	0.059	3	0.083	2	0.081	2	0.075	2	0.092	5	0.084	3	0.080	3	0.079	4
	MTMVc	0.070	1	0.056	3	0.070	3	0.075	2	0.066	2	0.085	1	0.070	4	0.060	4	0.063	2
HF(10)	MTMVSc	0.064	3	0.056	6	0.080	3	0.073	1	0.063	6	0.079	2	0.070	4	0.065	8	0.070	3
	MTMVSKc	0.057	3	0.048	3	0.078	2	0.059	3	0.060	1	0.093	0	0.060	4	0.062	7	0.071	4
	STMVnc	0.062	0	0.058	1	0.055	3	0.074	2	0.066	2	0.059	4	0.076	4	0.069	4	0.052	3
	STMVSnc	0.071	2	0.056	1	0.076	3	0.061	6	0.067	3	0.066	2	0.076	2	0.063	5	0.057	6
	STMVSKnc	0.060	3	0.051	3	0.068	4	0.052	4	0.061	5	0.058	7	0.072	3	0.062	3	0.049	5
	MTMVnc	0.056	1	0.051	3	0.057	2	0.064	3	0.060	4	0.079	3	0.070	4	0.065	3	0.061	4
	MTMVSnc	0.058	6	0.060	3	0.072	2	0.058	5	0.062	1	0.054	2	0.052	5	0.060	5	0.055	4
	MTMVSKnc	0.056	3	0.059	2	0.059	1	0.058	2	0.058	2	0.056	3	0.054	8	0.059	6	0.060	3
	Sharpe	0.053	1	0.053	1	0.056	0	0.078	3	0.049	2	0.077	0	0.070	0	0.069	0	0.072	0
	Sortino	0.053	2	0.052	2	0.055	0	0.074	2	0.047	0	0.070	0	0.072	0	0.073	0	0.074	0
	Omega	0.052	1	0.051	1	0.054	0	0.068	1	0.046	1	0.070	1	0.070	0	0.071	0	0.072	2
	STMVc	0.060	1	0.052	2	0.053	1	0.088	4	0.064	5	0.071	4	0.064	7	0.067	6	0.061	4
	STMVSc	0.056	7	0.052	2	0.060	3	0.088	3	0.062	4	0.078	3	0.070	6	0.067	4	0.074	2
	STMVSKc	0.055	2	0.051	4	0.054	3	0.080	0	0.055	5	0.075	2	0.066	4	0.064	4	0.079	1
	MTMVc	0.054	1	0.048	3	0.060	1	0.071	2	0.061	1	0.079	4	0.066	2	0.063	1	0.068	3
HF(20)	MTMVSc	0.056	2	0.046	6	0.062	3	0.071	1	0.058	6	0.075	3	0.062	3	0.060	4	0.068	2
	MTMVSKc	0.059	3	0.048	1	0.055	3	0.073	1	0.054	1	0.078	1	0.052	3	0.055	9	0.074	4
	STMVnc	0.058	0	0.052	2	0.050	2	0.090	1	0.063	0	0.064	2	0.064	5	0.064	3	0.054	4
	STMVSnc	0.052	5	0.051	3	0.056	1	0.066	5	0.054	4	0.054	4	0.067	1	0.056	0	0.061	1
	STMVSKnc	0.052	3	0.050	2	0.057	2	0.057	5	0.055	3	0.061	3	0.065	2	0.060	2	0.060	3
	MTMVnc	0.052	1	0.048	3	0.053	1	0.069	1	0.051	3	0.077	2	0.062	1	0.056	2	0.065	1
	MTMVSnc	0.056	3	0.050	3	0.051	1	0.068	4	0.047	4	0.068	2	0.046	9	0.049	6	0.063	4
	MTMVSKnc	0.054	2	0.047	1	0.047	2	0.073	2	0.048	2	0.060	1	0.044	8	0.050	5	0.058	3
	Sharpe	0.040	3	0.034	4	0.039	1	0.072	0	0.049	0	0.057	0	0.075	0	0.065	0	0.069	0
	Sortino	0.039	1	0.034	2	0.038	1	0.065	0	0.043	0	0.054	0	0.071	1	0.061	1	0.064	2
	Omega	0.042	2	0.036	2	0.038	2	0.066	0	0.044	1	0.056	0	0.072	0	0.063	0	0.066	0
	STMVc	0.055	3	0.041	4	0.044	3	0.081	3	0.058	4	0.063	4	0.070	5	0.058	3	0.058	3
	STMVSc	0.053	6	0.042	5	0.050	3	0.080	2	0.057	2	0.067	3	0.072	3	0.059	2	0.063	4
	STMVSKc	0.048	4	0.041	2	0.045	3	0.073	1	0.048	4	0.060	3	0.065	3	0.057	3	0.061	3
	MTMVc	0.051	2	0.039	3	0.053	3	0.061	4	0.049	6	0.061	2	0.065	3	0.055	5	0.059	3
HF(30)	MTMVSc	0.052	1	0.036	7	0.055	2	0.059	2	0.046	4	0.066	3	0.059	2	0.053	3	0.065	4
	MTMVSKc	0.052	1	0.044	3	0.046	3	0.061	3	0.046	2	0.060	1	0.053	5	0.049	4	0.064	4
	STMVnc	0.051	2	0.045	2	0.042	1	0.081	1	0.055	4	0.055	1	0.072	3	0.061	2	0.058	6
	STMVSnc	0.048	3	0.043	2	0.045	1	0.069	3	0.045	3	0.046	2	0.067	2	0.056	2	0.057	1
	STMVSKnc	0.046	3	0.042	2	0.046	3	0.058	3	0.046	3	0.054	2	0.063	2	0.056	2	0.057	0
	MTMVnc	0.048	3	0.039	1	0.050	1	0.059	3	0.047	5	0.067	3	0.061	3	0.051	4	0.062	3
	MTMVSnc	0.053	0	0.047	3	0.043	2	0.059	1	0.044	5	0.054	2	0.050	2	0.048	5	0.059	1
	MTMVSKnc	0.051	1	0.047	3	0.044	2	0.060	3	0.044	5	0.055	3	0.050	3	0.049	5	0.058	2

Table 3.12: Performance results for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 Luenberger rating methods) held for 1, 3 and 5 years: Descriptive statistics of the shortage function values computed by nonconvex VRS frontier

	Methods	Hold 1 year						Hold 3 years						Hold 5 years					
		L (·)		E (·)		F (·)		L (·)		E (·)		F (·)		L (·)		E (·)		F (·)	
		Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
	Sharpe	0.044	10	0.041	10	0.042	9	0.042	11	0.040	11	0.046	12	0.045	12	0.045	8	0.042	13
	Sortino	0.043	12	0.039	13	0.043	12	0.027	17	0.026	18	0.045	10	0.042	7	0.048	7	0.036	12
	Omega	0.039	11	0.036	12	0.041	12	0.026	16	0.025	14	0.044	7	0.043	11	0.046	6	0.038	13
	STMVc	0.054	6	0.049	7	0.051	5	0.048	17	0.052	15	0.046	11	0.055	17	0.051	13	0.039	12
	STMVSc	0.051	11	0.049	9	0.048	12	0.055	17	0.050	14	0.050	21	0.052	16	0.053	16	0.025	25
	STMVSKc	0.049	13	0.044	10	0.059	11	0.056	10	0.050	15	0.063	12	0.055	14	0.058	11	0.041	17
	MTMVc	0.055	10	0.041	14	0.052	7	0.046	12	0.042	12	0.051	14	0.038	17	0.033	19	0.037	16
HF(10)	MTMVSc	0.045	10	0.038	16	0.055	9	0.038	13	0.033	15	0.045	18	0.031	17	0.039	17	0.040	14
	MTMVSKc	0.043	12	0.029	16	0.056	9	0.038	11	0.035	16	0.060	13	0.034	14	0.041	13	0.042	18
	STMVnc	0.045	6	0.039	6	0.042	7	0.049	11	0.039	16	0.036	13	0.048	16	0.048	13	0.033	11
	STMVSnc	0.053	14	0.040	8	0.055	7	0.035	21	0.044	14	0.033	17	0.049	17	0.047	17	0.032	19
	STMVSKnc	0.047	15	0.038	9	0.049	8	0.026	26	0.042	14	0.028	22	0.045	17	0.044	17	0.027	17
	MTMVnc	0.041	8	0.036	13	0.036	14	0.038	14	0.037	19	0.050	16	0.045	18	0.041	18	0.039	15
	MTMVSnc	0.040	14	0.043	9	0.048	10	0.032	16	0.039	14	0.029	18	0.025	22	0.039	14	0.033	17
	MTMVSKnc	0.041	15	0.041	9	0.040	11	0.032	17	0.034	11	0.028	18	0.028	21	0.040	17	0.036	16
	Sharpe	0.043	8	0.042	8	0.043	4	0.031	16	0.020	14	0.033	8	0.049	2	0.049	1	0.039	9
	Sortino	0.042	8	0.041	9	0.044	3	0.027	18	0.019	20	0.029	14	0.053	2	0.052	2	0.043	9
	Omega	0.042	7	0.040	9	0.041	6	0.027	14	0.016	13	0.032	6	0.049	5	0.049	6	0.042	10
	STMVc	0.045	7	0.039	10	0.040	8	0.058	14	0.040	15	0.047	7	0.041	16	0.045	17	0.043	11
	STMVSc	0.041	9	0.039	10	0.042	10	0.055	10	0.035	15	0.054	16	0.044	16	0.044	15	0.044	15
	STMVSKc	0.039	10	0.038	10	0.039	6	0.051	7	0.034	11	0.054	9	0.043	16	0.045	15	0.039	10
	MTMVc	0.036	10	0.034	14	0.040	13	0.044	13	0.041	8	0.053	12	0.040	12	0.037	17	0.042	13
HF(20)	MTMVSc	0.039	11	0.030	17	0.042	14	0.045	13	0.034	9	0.048	14	0.041	19	0.037	11	0.037	13
	MTMVSKc	0.043	9	0.030	15	0.037	13	0.051	9	0.036	14	0.055	9	0.034	15	0.040	15	0.036	16
	STMVnc	0.041	8	0.041	6	0.037	4	0.061	8	0.043	13	0.035	14	0.041	17	0.040	13	0.032	13
	STMVSnc	0.039	12	0.038	7	0.037	5	0.034	18	0.032	12	0.029	19	0.041	14	0.039	14	0.039	13
	STMVSKnc	0.037	10	0.037	8	0.039	6	0.030	21	0.033	12	0.036	21	0.039	12	0.040	16	0.039	10
	MTMVnc	0.036	10	0.033	10	0.038	7	0.040	10	0.032	10	0.050	8	0.041	13	0.031	17	0.042	12
	MTMVSnc	0.040	9	0.034	13	0.033	10	0.045	15	0.029	15	0.043	15	0.027	22	0.033	17	0.039	17
	MTMVSKnc	0.039	11	0.034	12	0.029	8	0.051	15	0.031	11	0.032	16	0.029	19	0.032	16	0.035	14
	Sharpe	0.031	8	0.028	9	0.030	8	0.032	10	0.021	11	0.031	5	0.052	4	0.046	4	0.045	10
	Sortino	0.029	6	0.028	5	0.030	3	0.028	14	0.020	17	0.025	15	0.052	6	0.043	6	0.042	7
	Omega	0.031	9	0.029	10	0.029	10	0.029	10	0.020	13	0.027	5	0.053	3	0.043	2	0.041	10
	STMVc	0.041	9	0.034	12	0.030	10	0.055	9	0.041	11	0.041	8	0.050	13	0.040	15	0.037	13
	STMVSc	0.040	10	0.036	7	0.037	8	0.050	10	0.038	13	0.047	9	0.052	15	0.040	16	0.040	12
	STMVSKc	0.033	14	0.034	11	0.032	12	0.048	8	0.031	18	0.043	13	0.038	16	0.040	13	0.035	14
	MTMVc	0.030	10	0.029	12	0.039	10	0.039	12	0.033	12	0.033	14	0.043	11	0.037	16	0.039	15
HF(30)	MTMVSc	0.038	10	0.027	15	0.039	11	0.035	8	0.029	16	0.044	11	0.037	13	0.032	17	0.041	14
	MTMVSKc	0.036	7	0.031	12	0.035	10	0.040	11	0.028	11	0.038	8	0.039	17	0.035	18	0.038	13
	STMVnc	0.036	11	0.038	9	0.031	7	0.054	4	0.040	15	0.032	12	0.052	13	0.042	10	0.032	14
	STMVSnc	0.032	14	0.035	6	0.034	3	0.040	10	0.030	13	0.023	22	0.047	11	0.037	12	0.030	13
	STMVSKnc	0.030	11	0.032	5	0.033	8	0.033	13	0.030	12	0.031	15	0.043	13	0.039	10	0.032	10
	MTMVnc	0.028	16	0.027	19	0.033	11	0.038	14	0.029	15	0.043	11	0.039	11	0.033	18	0.036	15
	MTMVSnc	0.032	11	0.036	12	0.029	12	0.038	6	0.025	10	0.032	13	0.032	13	0.035	19	0.033	17
	MTMVSKnc	0.033	11	0.034	13	0.030	8	0.041	11	0.027	10	0.033	15	0.037	15	0.035	18	0.033	16

3.5 Conclusion

The main objective of this contribution has been to introduce a general method for measuring the evolution of fund performance in the multi-time and multi-moment framework using some extensions of the shortage function combined with the recent developments in productivity indicator theory. Before developing this dynamic portfolio evaluation problem, we first define a new performance measure with the help of the shortage function to evaluate fund performance by simultaneously looking for contraction in variance and kurtosis over multiple times, and augmentation in mean return and skewness over multiple times. Using the proposed performance measures based on the nonparametric VRS convex and nonconvex frontiers, we propose a new Luenberger portfolio productivity indicator and its decomposition to trace the performance change of funds.

The virtues of the proposed methodology can be summarized as follows: First, instead of the static analysis on fund performance, these indicators provide a feasible way to trace the evolution of fund performance with the consideration of multi-times and multi-moments. Even though the Luenberger portfolio productivity indicator is not strictly based on the utility theory of investment, it is important to note that the performance changes over time somewhat indicate gains and losses in utility (see Briec, Kerstens, and Lesourd (2004), Brandouy, Briec, Kerstens, and Van de Woestyne (2010)). Second, this Luenberger portfolio productivity indicator is developed based on the convex and nonconvex efficient VRS frontier methods instead of the diversified efficient frontier methods. The proposed non-diversified frontier-based efficiency measures and productivity indicators are all simply computed by the linear (or binary mixed integer) programming, which are much easier and time-saving when applied in the large-scale and multi-dimensional fund rat-

ing. Third, the decomposition of the Luenberger portfolio productivity indicator distinguishes whether changes in fund performance over time are due to the efficiency change or the frontier shift. While the latter component measures the local changes in the frontier movements induced by market volatility, the former can in principle capture efficiency changes attributable to the fund manager. This allows testing in an alternative, but conceptually promising way the eventual ability of fund managers to generate superior performances, since this measurement is not contaminated by any changes in the financial market itself.

In the empirical testing, we illustrate how the multi-time and multi-moment Luenberger productivity indicator works in the evaluation and selection of hedge funds, whose distribution characteristics are not characterized by mean and variance solely (see, e.g., Gregoriou, Sedzro, and Zhu (2005), Kumar, Roy, Saranga, and Singal (2010), Germain, Nalpas, and Vanhems (2011)). In particular, our empirical investigation is devoted to making the following two issues clear. First, multiple time periods and moments considered separately or jointly have an impact on the Luenberger portfolio productivity indicator and its decomposition, as well as the Luenberger-based ratings for our data. The second is to test the potential of the proposed indicator and its decomposition on fund selection based on our sample data. For the former issue, empirical results indicate that regardless of the Luenberger portfolio indicator and its decomposition, or the indicator-based rankings, they are significantly impacted by multiple time periods and multiple moments. Moreover, the joint inclusion of them tends to exhibit a more pronounced impact compared to the separate inclusion of them. For the second issue, we develop a simple buy-and-hold backtesting strategy to test performance of traditional finance-based ratings and various Luenberger-driven ratings in HF selection. In our backtesting analysis, the buy-and-hold strategies consisting of the best funds as rated by the proposed Luenberger indicator perform any better

than more finance-based ratings and single-time MV Luenberger-based ratings. Accordingly, the efficiency component of the proposed indicators tends to convey more the strength of the consideration of multiple time periods in fund selection, while the frontier component better captures the information from higher order moments.

Obviously, the proposed methodology and the resulting empirical results have some limitations one should be aware of. First, it would be good if the rating based on the proposed Luenberger indicators can be tested in a more realistic backtesting analysis setting. For instance, the diversified portfolio strategies consisting of funds selected by Luenberger-based rating could be developed for testing the performance of different rating methods. This calls for a generalization of the current diversified models in multi-time and multi-moment framework. Another restriction is that it does not account for other commonly used risk measure instead of the variance, such as Value-at-Risk (VaR) or the Conditional Value-at-Risk (CVaR) (see Alexander and Baptista (2002), Lwin, Qu, and MacCarthy (2017), Guo, Chan, Wong, and Zhu (2019)). Future research could focus on extending our fundamental methodology to the mean-VaR (or CVaR)-skewness-kurtosis framework.

References

Aksarayli, M., and O. Pala (2018): "A Polynomial Goal Programming Model for Portfolio Optimization Based on Entropy and Higher Moments," *Expert Systems with Applications*, 94, 185–192.

Alexander, G. J., and A. M. Baptista (2002): "Economic Implications of Using a Mean-VaR Model for Portfolio Selection: A Comparison with Mean-Variance Analysis," *Journal of Economic Dynamics and Control*, 26(7-8), 1159–1193.

Arditti, F. D. (1975): "Skewness and Investors Decisions: A Reply," *Journal of Financial and Quantitative Analysis*, 10(1), 173 –176.

Asmild, M., J. C. Paradi, V. Aggarwall, and C. Schaffnit (2004): "Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry," *Journal of Productivity Analysis*, 21(1), 67–89.

Babalos, V., G. M. Caporale, and N. Philippas (2012): "Efficiency Evaluation of Greek Equity Funds," *Research in International Business and Finance*, 26(2), 317–333.

Brandouy, O., W. Briec, K. Kerstens, and I. Van de Woestyne (2010): "Portfolio Performance Gauging in Discrete Time Using a Luenberger Productivity Indicator," *Journal of Banking & Finance*, 34(8), 1899–1910.

Brandouy, O., K. Kerstens, and I. Van de Woestyne (2015): "Frontier-

based vs. Traditional Mutual Fund Ratings: A First Backtesting Analysis," *European Journal of Operational Research*, 242(1), 332–342.

Briec, W., C. Comes, and K. Kerstens (2006): "Temporal Technical and Profit Efficiency Measurement: Definitions, Duality and Aggregation Results," *International Journal of Production Economics*, 103(1), 48–63.

Briec, W., and K. Kerstens (2009a): "Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator," *Journal of Optimization Theory and Applications*, 141(1), 55.

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

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

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

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

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

Chambers, R. G., R. Färe, and S. Grosskopf (1996): "Productivity Growth in APEC countries," *Pacific Economic Review*, 1(3), 181–190.

Cooper, W. W., L. M. Seiford, and K. Tone (2007): *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer, Berlin.

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

Dittmar, R. F. (2002): "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns," *Journal of Finance*, 57(1), 369–403.

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

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

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

Gardijan, M., and T. Škrinjarić (2015): "Equity portfolio optimization: A DEA based methodology applied to the Zagreb Stock Exchange," *Croatian Operational Research Review*, 6(2), 405–417.

Germain, L., N. Nalpas, and A. Vanhems (2011): "Nonparametric Hedge Funds and Replication Indices Performance Analysis: A Robust Directional

Application," in *Hedge Funds Replication*, ed. by G. Gregoriou, and M. Kooli, pp. 90–105. Palgrave Macmillan, Basingstoke.

Gong, X., C. Yu, L. Min, and Z. Ge (2021): "Regret Theory-based Fuzzy Multi-objective Portfolio Selection Model Involving DEA Cross-efficiency and Higher Moments," *Applied Soft Computing*, 100, 106958.

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

Gregoriou, G. N., and J. Moschella (2017): "The Efficiency and Productivity of Hedge Fund Families," *Journal of Wealth Management*, 20(3), 101–112.

Guidolin, M., and A. Timmermann (2002): "International Asset Allocation with Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Guo, X., R. H. Chan, W. K. Wong, and L. Zhu (2019): "Mean-Variance, Mean-VaR, and Mean-CVaR Models for Portfolio Selection with Background Risk," *Risk Management*, 21(2), 73–98.

Hanoch, G., and H. Levy (1969): "The Efficiency Analysis of Choices Involving Risk," *The Review of Economic Studies*, 36(3), 335–346.

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

Hwang, S., and S. Satchell (1998): "Evaluation of Mutual Fund Performance in Emerging Markets," *Emerging Markets Quarterly*, 2(23), 39–50.

Jondeau, E., and M. Rockinger (2006): "Optimal Portfolio Allocation Under Higher Moments," *European Financial Management*, 12(1), 29–55.

Kerstens, K., A. Mounir, and I. Van de Woestyne (2011a): "Geometric Representation of the Mean-Variance-Skewness Portfolio Frontier Based upon the Shortage Function," *European Journal of Operational Research*, 210(1), 81–94.

——— (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.

Konno, H., H. Shirakawa, and H. Yamazaki (1993): "A Mean-Absolute Deviation-Skewness Portfolio Optimization Model," *Annals of Operations Research*, 45(1), 205–220.

Krüger, J. (2021): "Nonparametric Portfolio Efficiency Measurement with Higher Moments," *Empirical Economics*, 61(3), 1435–1459.

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

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

Leland, H. E. (1999): "Beyond Mean–Variance: Performance Measurement in a Nonsymmetrical World (corrected)," *Financial Analysts Journal*, 55(1), 27–36.

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

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

Lin, R., Z. Chen, Q. Hu, and Z. Li (2017): "Dynamic Network DEA Approach with Diversification to Multi-period Performance Evaluation of Funds," *OR Spectrum*, 39(3), 821–860.

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

Luenberger, D. G. (1995): *Microeconomic Theory*. Mcgraw-Hill College.

Lwin, K. T., R. Qu, and B. L. MacCarthy (2017): "Mean-VaR Portfolio Optimization: A Nonparametric Approach," *European Journal of Operational Research*, 260(2), 751–766.

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

Matallín-Sáez, J., A. Soler-Domínguez, and E. Tortosa-Ausina (2014): "On the Informativeness of Persistence for Evaluating Mutual Fund Performance using Partial Frontiers," *Omega*, 42(1), 47–64.

Mehlawat, M. K., A. Kumar, S. Yadav, and W. Chen (2018): "Data Envelopment Analysis based Fuzzy Multi-objective Portfolio Selection Model Involving Higher Moments," *Information sciences*, 460, 128–150.

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

Nalpas, N., L. Simar, and A. Vanhems (2017): "Portfolio Selection in a Multi-Moment Setting: A Simple Monte-Carlo-FDH Algorithm.," *European Journal of Operational Research*, 263(1), 308–320.

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

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

Tobin, J. (1958): "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, 25(2), 65–86.

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

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

Zeitun, R., K. S. Abdulqader, and K. A. Alshare (2013): "On the Relative Efficiency of Conventional and Islamic Banks: A DEA-window Approach," *International Journal of Financial Services Management*, 6(3), 236–251.

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

CHAPTER

4

Multi-Time and Multi-Moment Nonparametric Frontier-Based Fund Rating: A Nonconvex Metafrontier Approach

Abstract

This contribution combines the nonparametric metatechnology with the shortage function to define a comparable metafrontier-based fund rating across classifications accounting for heterogeneity. Without limiting ourselves to the traditional mean-variance portfolio setting, we develop the nonparametric metafrontier rating methods with multiple times and multiple moments, aiming to allow for the fund evaluation compatible with general investor preferences. The empirical part employs a large database of nonhomogeneous funds covering five different types not only to offer extensive tests of the specification factors considered in these models, but also to test the performance of different rating methods in a simple backtesting setup. First, the evaluation results show that multiple moments, multiple times, and convexity assumption for each GTPPS have an impact on the estimates of

MTE and its components, respectively. Second, the backtesting results indicate that the strategies based on the multi-time and multi-moment metafrontier ratings exhibit a comparable performance to those based on the financial ratings, whereas the strategies based on the other metafrontier ratings maintain a poor performance.

KEYWORDS: Shortage function; Fund rating; Metafrontier.

4.1 Introduction

With the development of global financial markets, the number and diversity of funds (e.g., hedge funds, equity funds, bond funds, etc.) available to investors are steadily increasing. Therefore, it has been a closely watched concern in both academia and industry to gauge the fund efficiency based on available information and to identify the funds worthwhile investing in. Obviously, an effective fund evaluation and rating can provide not only strategic support for investors' fund screening, but also investment benchmarking for fund managers to improve the performance of their managed funds. To this end, numerous business magazines and private firms now specialize in giving regular, exhaustive rankings and ratings of funds (such as Lipper, Morningstar, Standard & Poor's, Fitch, etc.).

A fundamental principle followed by these rating agencies in assigning ratings is that funds are evaluated within a homogeneous (peer) group with comparable investment objectives. These homogeneous groups are divided depending on the fund classification/category defined by each rating agency. For example, the Lipper Leaders uses the Lipper Global Classification (LGC) as the primary peer group for a number of calculations on the fund efficiency and assigns ratings to each fund within a specific classification accordingly. The LGC creates homogeneous groups

of funds whose members are similar enough in their risk factor exposures for the return comparisons between them. One of the key classification attributes is the underlying asset markets in which a fund strategically invests. According to the LGC, funds classified in the same group are required to maintain at least 75% of their exposure to one or more identical underlying markets. For instance, funds with the primary objective of investing in the stock markets are placed into the equity group. In general, the LGC maps into eight broad fund groups that include equity, bond, commodity, money markets, real estate, mixed asset, alternative (for instance, hedge funds are classified in this group), and others. This contribution in particular focuses on the assessment with respect to these typical groups of fund performance defined by the LGC.

Ratings depending on the fund types reveal a wide consensus in finance: funds in different groups adopt distinct styles and exhibit heterogeneous performances, making their efficiency not directly comparable. In essence, this heterogeneity in fund performance across groups is attributed to regulatory constraints and strategic investment reasons. Specifically, funds with a certain regulatory and strategic framework are subject to limitations on the choice of underlying asset types (e.g., equities, bonds, money markets, etc.), investment regions (e.g., UK, Europe, emerging markets, etc.), rules on portfolio composition and management mandates. Fund managers often have limited direct control over these factors. For instance, the money market fund (also called the money market mutual fund) are limited to investing in money market instruments, such as cash, cash equivalent securities, and high-credit-rating, debt-based securities with a short-term maturity. In the finance literature, some research has been devoted to explore the relationship between the fund with regulatory and strategic limitations and its performance (e.g., Alexander and Stover (1980), Brown and Goetzmann (1997), Del Guercio and Tkac (2002), Gerlach and Maurer (2015), Ewen (2018), among

others). These research contributions, in general, provide empirical evidence for the argument that heterogeneity in fund performance can be due to differences in both the regulatory and strategic frameworks.

In recent years, the successful nonparametric frontier estimation methodologies from production theory have been gradually and partially transposed to fund performance assessment. Intuitively, based on a sample of observed units, one estimates nonparametric frontiers of any multi-dimensional choice set and uses an efficiency measure to position the benchmark of each observation on the boundary of such choice set (see Ray (2004)). In the context of production, to the best of our knowledge, homogeneity of the observed units is a common assumption made in standard applications of nonparametric frontier methods. It implies that the operating technologies or environments of all units are sufficiently similar for the purposes of efficiency evaluation (see Dyson, Allen, Camanho, Podinovski, Sarrico, and Shale (2001)). Besides, this assumption further indicates that the units under evaluation should be regarded as members with the same production technology, and allows these units to be benchmarked only against the common production frontier generated by all observed units in the peer group (see Cooper, Seiford, and Tone (2007)).

The assumption of homogeneity is naturally inherited when the nonparametric frontier methods is applied to fund evaluation. In this context, the efficiency is evaluated according to the distance between the fund being evaluated and its projection on the frontier generated by all homogeneous funds, that is the distance relative to the peer group between its current position and the frontier projection it can achieve with minimum effort. For example, the seminal contribution of Murthi, Choi, and Desai (1997) apply the traditional convex nonparametric frontier methods to estimate the mutual funds (MF) efficiency within the same

classification, where the variance and transaction cost indicators are treated as the inputs, and the expected return is regarded as the output. Kerstens, Mounir, and Van de Woestyne (2011) adapt the convex/nonconvex nonparametric frontier methods combining with the shortage function to gauge funds from European and United States market, and systematically discuss the specifications of the nonparametric frontier estimates. For more studies with respect to performance evaluation of homogeneous traditional MFs (e.g., equity funds, bond funds and mixed asset funds/ balanced funds, etc.), the reader can may refer to Basso and Funari (2001), Choi and Murthi (2001), Anderson, Brockman, Christos, and McLeod (2004), Chen and Lin (2006), Daraio and Simar (2006), Galagedera, Roshdi, Fukuyama, and Zhu (2018) among others.

Meanwhile, some burgeoning literature focusing on the evaluation of other special MF types can be found. For instance, similar models have also been proposed for the evaluation of hedge funds (Gregoriou (2003), Gregoriou, Sedzro, and Zhu (2005), Darolles and Gourieroux (2010), Eling and Faust (2010)), pension funds (Miszczyńska and Miszczyński (2007), Andreu, Sarto, and Vicente (2014)), ethical mutual funds (Basso and Funari (2003), Basso and Funari (2014)). Strictly speaking, the fund efficiency defined by the existing rating methodologies can be regarded as the group-specific efficiency of a fund, which evaluates how well this fund performs well relative to the peer group.

For the investor adhering to a certain specific fund type (e.g., hedge funds, pension funds, etc.), the funds' performance that is rated by the group-specific efficiency provides fairly useful information for selecting best-performing funds from the specific group. However, for most typical investors without a fixed target investment fund type, as they may change their preference for fund type according to fund market movements (see, e.g., when investors expect the equity market

to enter a bear market, they may sell their equity funds and choose to hold bond funds or money funds, in which case their preference changes from equity to bond or money funds), the group-specific efficiency may hardly offer sufficient information for the cross-group selection of funds. The aim of our contribution is to develop an evaluation procedure for making the efficiency comparisons across groups of funds. Generally speaking, the additional information is revealed in this cross-group comparison of funds compared to the within-group comparison: (i) this result allows investors, who have no preference for certain specific fund types, to select well performing funds among a broad fund universe in line with their own risk-aversion utility functions; (ii) as for investors, who adhere to a specific fund type, this result can indicate the potential gains they may experience if the investment preferences on fund types would be changed. To validly conduct a cross-group comparison for funds, this evaluation procedure is designed in a way that accounts for the heterogeneity of the different groups of funds, on the one hand, and that the performance of funds is assessed on the same benchmark for different classes, on the other hand. The concept of a metafrontier representing an unrestricted technology from production theory opens a new sight to handle this problem and allows for the evaluation and comparison across groups of funds with heterogeneity.

The metafrontier, originated back to the concept of meta-production function proposed by Hayami and Ruttan (1970), accounts for heterogeneity when estimating production relations. Those authors call “the envelope of all known and potentially discoverable activities a secular or meta-production function." (p. 898). Their meta-production function gives the maximum output obtainable from given inputs and a given set of production technologies. Different units may choose a different specific technology from the set of available technologies depending on a variety of circumstances. This basic logic has initially been transposed into a

production frontier framework using both stochastic parametric and deterministic nonparametric estimation approaches by Battese and Rao (2002), Battese, Rao, and O'Donnell (2004), and O'Donnell, Rao, and Battese (2008).

O'Donnell, Rao, and Battese (2008) defines a production possibilities, metaset as the union of underlying group-specific production possibilities sets. The boundary of this metaset is referred to as a production metafrontier, and the boundaries of the group-specific sets are called group-specific production frontiers (or group frontiers). Thereafter, this metafrontier approach has been amply applied across sectors and disciplines. Examples are production studies from agriculture (e.g., Latruffe, Fogarasi, and Desjeux (2012)), banking (e.g., Casu, Ferrari, and Zhao (2013)), hotels (e.g., Huang, Ting, Lin, and Lin (2013)), schools (e.g., Thieme, Prior, and Tortosa-Ausina (2013)), metal industry (e.g., Feng, Huang, and Wang (2018)), and wastewater treatment plants (e.g., Sala-Garrido, Molinos-Senante, and Hernández-Sancho (2011)) to name but a few. This basic metafrontier concept has found its way in a variety of other literatures: One example is its transposition to a cost frontier framework (e.g., Huang and Fu (2013)); another example is the computation of productivity indices relative to metafrontiers (see, e.g., Casu, Ferrari, and Zhao (2013) and Huang, Juo, and Fu (2015)); a final example is the development of more elaborate efficiency decompositions in the metafrontier evaluation framework (see Kounetas, Mourtos, and Tsekouras (2009) and Tsekouras, Chatzistamoulou, and Kounetas (2017)). However, to the best of our knowledge, none of these existing articles systematically investigates the application of the metafrontier approach in the assessment of funds accounting for heterogeneity.[1]

In this paper, we tackle the problem of assessing heterogeneous funds belonging

[1]Even though Makni, Benouda, and Delhoumi (2015) use the meta-frontier approach with data envelopment analysis (DEA) to compare the relative efficiency of Islamic equity funds. However, the main specification issues surrounding the application of metafrontier methods for assessing the financial performance of funds are not discussed in their research.

to different groups by borrowing from recent developments in the nonparametric metafrontier approaches. Employing the distance (shortage) function measure (Luenberger (1995)), a series of nonparametric metafrontier-based rating methods are established that allow for evaluating the fund performance across groups along a multitude of dimensions, and that offer a comparable measure against the corresponding nonparametric metafrontier. The nonparametric metafrontier can envelop all observed funds of multi-dimensional metasets that include each of the group-specific sets as a subset, and position each of the funds relative to the boundary of this metaset using the shortage function. Clearly, this metafrontier is viewed as representing the benchmark that can be achieved in principle for each of the observed funds with respect to the available fund universes.

In the existing production literature, the metafrontier should be estimated correctly as the boundary of a nonconvex metaset, which is defined as the union of either convex or nonconvex group-specific sets. However, there are still a lot of articles that ignore the nonconvex nature of the metafrontier and adopt a convexification strategy, i.e. estimating the metafrontier as the boundary of a convex metaset(e.g., O'Donnell, Rao, and Battese (2008), Kontolaimou and Tsekouras (2010), Portela, Thanassoulis, Horncastle, and Maugg (2011)). It is clear that the assumption on convexity between groups needs normally not to be valid, "convexifying" estimator of the metafrontier risk being biased. Kerstens, O'Donnell, and Van de Woestyne (2019) develop some new results on the union operation on sets under a variety of assumptions and deliver convincing empirical evidence that a convexification strategy yields statistically significant biases. Therefore, the metafrontier for fund rating in our contribution is estimated using the nonparametric nonconvex metafrontier, which is the only correct estimator in the production literature (see also Afsharian and Podinovski (2018), Huang, Ting, Lin, and Lin (2013), and Jin, Kerstens, and Van de Woestyne (2020), among others.).

Distinct from the production area, the performance evaluation of funds/porfolios mainly focuses on the tradeoff between their returns and risks, among which the most classical measures of return and risk are respectively the mean and variance (see Markowitz (1952)). In this paper, apart from developing the metafrontier methodology in a basic mean-variance (MV) evaluation framework solely for a single period in a time horizon, we further explore the application of this metafrontier method in the general frameworks consistent with mixed risk-aversion investor preferences. Instead of extensively discussing each of these modelling approaches, we offer some arguments to narrow down the number of potential models worthwhile considering. In particular, we focus on two factors for extending nonparametric metafrontier rating models to gauge heterogeneous funds: (i) consideration of multiple times, and (ii) inclusion of multiple moments (i,e., mean, variance, skewness and kurtosis).

The first factor to be considered aims at addressing a fund's total risk and total return performance over different times. Morey and Morey (1999) are the first to propose two types of efficiency measures from a multi-time perspective for assessing MV portfolio performance: The first one contracts all risk dimensions proportionally and the second one focuses on augmenting all return dimensions as much as possible in a proportional way. Briec and Kerstens (2009) extend the analysis of multi-time MV portfolio analysis by Morey and Morey (1999) and obtain benchmark portfolios by simultaneously considering risk contraction and mean return expansions using the discounted temporal shortage function within a multi-time framework. These various time performances available to investors can provide considerable additional information over that given by a fund's performance over just a single time (see Ren, Zhou, and Xiao (2021)).

The second factor accounted for in the rating models to handle mixed risk-

aversion preferences of investors, where investors exhibit preference (aversion) for odd (even) moments of the probability distribution of asset returns (see Trautmann and van de Kuilen (2018) for a review). In fact, while the MV rating methods remain a popular reference among practitioners and academics, this evaluation criterion is only consistent with the quadratic utility function of the investor. For the widely accepted mixed risk-aversion utility functions, one needs to use higher-order moments as input-output specification allowing for gauging the performance of funds compatible with these general investor preferences (see Briec and Kerstens (2010)). Therefore, enlarging the classical framework with multi-moment models, i,e., mean-variance-skewness (MVS) and mean-variance-skewness-kurtosis (MVSK) models, is a potentially interesting improvement for finance. An extensive discussion can be also found on the inclusion of higher moments in fund evaluation. Examples of these studies include, e.g., Matallín-Sáez, Soler-Domínguez, and Tortosa-Ausina (2014), Brandouy, Kerstens, and Van de Woestyne (2015), Nalpas, Simar, and Vanhems (2017), Mehlawat, Kumar, Yadav, and Chen (2018), Krüger (2021), Lin and Li (2020), Gong, Yu, Min, and Ge (2021). Naturally, the nonparametric metafrontier methods developed in this contribution can be directly extended to consider even higher moments without increasing the computational cost excessively, as these nonparametric models can normally be solved using linear or mixed linear integer programming.

In summary, the main objectives of this contribution are fourfold. First, we apply the most recent metafrontier approach to the fund evaluation accounting for heterogeneity, and define the metatechnology efficiency of funds by combining the shortage function with the nonparametric metafrontier models. This metatechnology efficiency can be decomposed into two components: the group-specific efficiency and the technology difference gap. The group-specific efficiency of funds allows for evaluating the performance of a fund relative to the peer group that has

opted to invest in similar underlying asset markets due to an identical regulatory and strategic framework. The technology difference gap measures the gap between the metafrontier and the group-specific frontier, which evaluates how well this fund has initially opted to invest in the underlying asset markets among the potentially available options. Second, we extend the metafrontier evaluation procedure from the basic single-time MV framework to the multi-time and multi-moment framework, which is compatible with general investor preferences regarding multiple times and multiple moments. Third, applying the actual data from different types of funds, we employ a Li-test approach to test for the impact of multiple times and multiple moments separately and jointly on the metatechnology efficiency and the corresponding component. Finally, a backtesting analysis is presented to test the potential benefits of the proposed metafrontier evaluation methods.

The remainder structure of this contribution is as follows. Section 4.2 presents the geometric intuition on the portfolio metafrontier and the corresponding nonparametric estimators. Section 4.3 develops a series of nonparametric metafrontier methodologies with a special focus on measuring the efficiency of funds across groups. Section 4.4 presents an empirical illustration using actual fund data. Finally, Section 4.5 summarizes our key results and issues for future work.

4.2 Portfolio Metafrontier and Nonparametric Estimators: Graphical Illustration

We start by reminding the reader about the intuition underlying the metafrontier approach. O'Donnell, Rao, and Battese (2008) develop the metafrontier approach that enables analysis of efficiency of heterogeneous decision making units (DMUs). According to this contribution, the metafrontier approach can be ad-

opted whenever DMUs should be classified into groups according to the technologies they use, and when DMUs in different groups choose input-output combinations from potentially different production possibility sets (PPS). DMUs in the same group are considered sufficiently homogeneous and represent the same group-specific technology. The group-specific technology is generally characterized by the group-specific technology PPS (GTPPS), and the boundary of the GTPPS is referred to as the group-specific frontier. The metatechnology includes all production possibilities achievable using different group-specific technologies. In particular, the metatechnology-specific PPS (MTPPS) is the set containing each of the above GTPPSs as a subset. The boundary of the MTPPS is referred to as the metafrontier. The metafrontier can be considered as representing the best production possibility set that can be achieved in principle, and allows for assessing whether DMUs have used the best operating technology among the available options to transforming inputs into outputs.

We now explore how to adapt the matafrontier approach from the original production context to handle the evaluation of heterogeneous funds/portfolios. In a production context, DMUs are commonly classified into groups according to the production technology in which they operate. Following the definition on technology (i.e., a technique, method or system for transforming inputs into outputs) in O'Donnell (2016, p. 328), this concept in the portfolio evaluation context can be considered as a selection for the underlying asset markets, which is generally described by the generated input-output combinations (risk-return and similar for higher moments). In essence, funds are classified into several groups based on the regulatory and strategic framework in which they operate, and these objective factors are largely linked to the underlying asset markets of investment that they have chosen at the outset. For instance, Euro bond funds are required to primarily invest in Euro-denominated fixed-income markets due to certain regulatory and

strategic limitations (normally at least two thirds of shares), in which case their selection on fixed-income markets is treated as the technology they use. This classification criterion is well understood in analogy to the classification criterion of a technology in a production context. Consequently, we generally refer to the selection of similar underlying markets as the group-specific technology for funds that has a specific regulatory and strategic framework, and the set of all these available selections across different regulatory and strategic frameworks as the metatechnology. Correspondingly, the boundaries of both are called the portfolio group-specific frontier and the portfolio metafrontier.

To illustrate the concept of a portfolio metatfrontier, we consider a simple case regarding a MV portfolio depicted in Figure 4.1, where only two portfolio group-specific frontiers are available. In other words, this case only contains two types of funds that invest in different underlying markets (e.g., equity markets and bond markets). Let $G1$ denote the set of MV combinations for funds that use group-specific frontier 1, i.e., invest in the equity markets. The MV portfolio frontier obtained from the funds in $G1$ is marked in blue. Similarly, let $G2$ denote the set of MV combinations for funds that use group-specific frontier 2, e.g., invest in the bond markets, and the corresponding MV portfolio frontier is marked in red. Let Γ denote the set of MV combinations for all funds that use the two available technologies, where the MV combinations generated by funds investing in the two underlying asset markets are included. This metatechnology set is given by the union of $G1$ and $G2$. The boundary of this set is the defined portfolio metafrontier in our work, as depicted by the solid line in Figure 4.1.

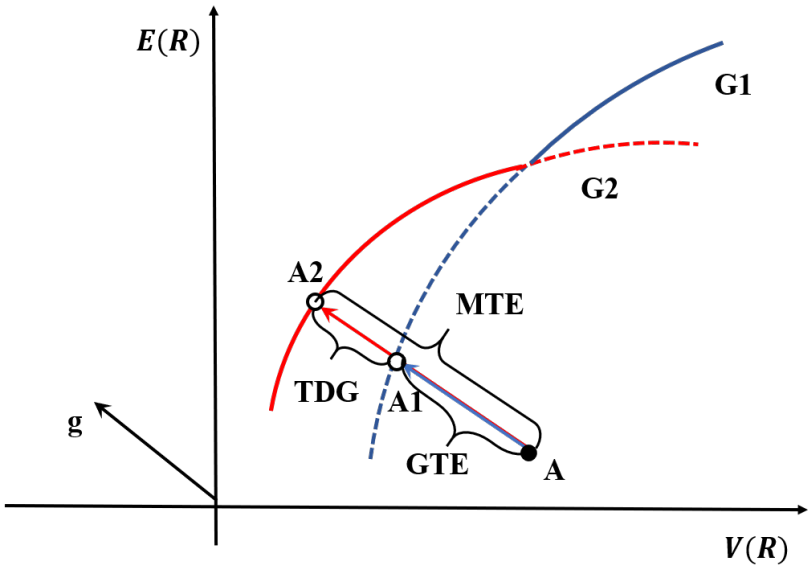

Figure 4.1: Portfolio group-specific frontiers and metafrontier in MV space

Clearly, this MTPSS of portfolios is nonconvex.[2] Note that assuming convexity for observations between different groups is a rather strong premise, since there are two distinct technologies which may have different characteristics. One should realize that this convexity assumption applied to the MTPPS is to some extent self-contradictory: it runs counter the very idea of distinguishing between different group-specific technologies. Even if the GTPPS could be assumed to satisfy convexity, there is no reason why the union of these sets should be convex. Otherwise stated, the union operator on GTPPSs does not normally preserve the convexity

[2]We restrict our portfolio metafrontier analysis to the assumption that the diversification effect exists only among the funds within the same group (or classification). The basic idea for this assumption is that a general diversified portfolio is likely to be difficult to implement among the financial products across groups given the specific regulatory and strategic limitations (i.c., the investment mandates). Therefore, the portfolio metafrontier in this contribution is defined as the union of the group-specific portfolio frontiers. This allows to retrospectively assess whether a given fund could have benefited if it would have been allowed to invest in different asset classes compared to its own current asset class. However, it is worth noticing that if diversification is allowed among all funds across groups, then the corresponding portfolio metafrontier should be generated by all observed funds from different groups: it would remain convex as in the case of the group-specific portfolio frontiers.

assumption on the resulting MTPPS. More importantly, the convex combinations between observations using different techniques create some virtual units, while in most cases, these virtual units need not be feasible in reality. As stated earlier, the group-specific portfolio frontiers are also distinguished because of the regulatory and strategic reasons, which lead to clear limitations on the selection of underlying asset markets, composition rules and investment mandates, etc. These virtual units imply that funds counterfactually could have been invested in other underlying financial markets, which in most cases would not be permitted as it violates relevant regulatory and strategic constraints (or the initial mandate). However, the efficiency determined using the metatechnology frontier may be influenced by these virtual units such that the results become unreliable. As a consequence, we use the nonconvex union of either convex or nonconvex GTPPS to form the MTPPS for gauging funds across different groups. This nonconvex MTPPS is also recommended in the production literature: examples include Afsharian and Ahn (2015), Afsharian and Podinovski (2018), Kerstens, O'Donnell, and Van de Woestyne (2019), Jin, Kerstens, and Van de Woestyne (2020).

Let us consider the measurement of efficiency regarding these frontiers in Figure 4.1. Consider the fund operating at point A with a given MV level, and recall that the fund can only use the technology 1, i.e., can only invest in the equity markets. Recall that in this case the portfolio metafrontier is given by the solid line in Figure (4.1). Assume that this fund with the current MV level uses technology 1 can increase the return and decrease the risk by moving to point $A1$ on the portfolio group-specific frontier with technology 1 along the given direction g. If instead it had been able to use technology 2, i.e., invest in the bond markets, then it could have improved the return and risk by moving to point $A2$ on the portfolio metafrontier. The overall efficiency determined by the portfolio metafrontier requires that retrospectively it should have opted to use technology

2: This fund should have opted to invest in the bond markets rather than in the equity markets. As presented in Figure 4.1, the metatechnology efficiency (MTE) of a fund is a direction-oriented measure of the distance from an observed point to a projection point on the portfolio metafrontier (i.e., the distance between A and $A2$). The MTE of a fund consists of two components: (i) the group-specific technology efficiency (GTE) is a direction-oriented measure of the distance from an observed point to a projection point on the portfolio group-specific frontier (i.e., the distance between A and $A1$), and (ii) the technology difference gap (TDG) is a direction-oriented measure of the distance between the portfolio metafrontier and the portfolio group-specific frontier (i.e., the distance between $A1$ and $A2$). The GTE component is the common measure of fund efficiency within a homogeneous group that invests in the same underlying markets, while the TDG component reveals the eventual restrictive nature of the underlying asset markets in which the fund has initially made its choice.[3]

Considering that the diversified portfolio frontier is difficult to obtain, especially when the time dimension and the higher-order moments are both included, in this contribution, we employ the nonparametric production frontier specifications to estimate the GTPPSs and the group-specific frontiers in the various frameworks (i.e., the single/multi-time and multi-moment frameworks with convexity and nonconvexity). Figure 4.2 illustrates the nonparametric estimators using the above simple MV case when only two technologies exist. In this figure, the left Figure 4.2(a) plots the estimated GTPPSs using the convex nonparametric frontier method (DEA estimator), and the right Figure 4.2(b) plots the inner estimate using the nonconvex nonparametric frontier method (FDH estimator). Whether the estimated GTPPSs is convex or nonconvex, the estimated MTPPS that is defined

[3]In this contribution, we only analyze the relative performance of the fund under evaluation by the comparison between funds, but not with respect to the efficiency measure of the financial markets in which this fund operate.

as the union of all the estimated GTPPSs is clearly nonconvex. In the following section, we introduce a series of nonparametric matafrontier approaches for the fund evaluation in different frameworks.

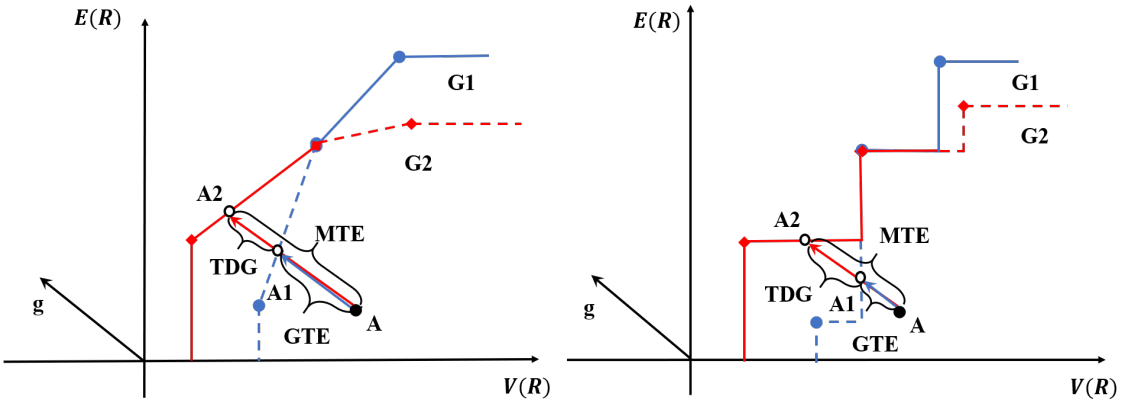

Figure 4.2: Nonparametric frontier estimates in MV space

4.3 Nonparametric Metafrontier Rating: Methodology

As introduced earlier, O'Donnell (2016, p. 328) defines a technology as ' 'a technique, method or system for transforming inputs into outputs. For most practical intents and purposes, it is convenient to think of a technology as a book of instructions, or recipe". Following this work, we view each group of funds as operating in a different technology condition. This definition is adopted here: we perceive a technology (group-specific technology) as a selection for (one or more) available underlying markets for funds that have a specific regulatory and strategic framework.

Assume that there are N observed funds over a given time horizon T. At

time t in this time horizon ($t \in \{1, \ldots, T\}$), the j-th fund ($j \in \{1, \ldots, N\}$) is characterized by m input-like values and s output-like values. Input-like variables need to be minimized and output-like variables need to be maximized. Suppose that these funds operate with different technologies, i.e., they invest in different underlying markets or specific segments of those markets. In this contribution, depending on the underlying asset universe, these may include alternative assets, short or long-term fixed income markets, equities, mixed assets, etc. To represent such differences, we assume that all funds under evaluation can be partitioned into $\Gamma > 1$ distinct groups, so that the funds in the same group $k \in \{1, \ldots, \Gamma\}$ operate in the same underlying markets.

Before proposing the following methodology, we first introduce some notations. At time t in a given time horizon T, all the observed input-output combinations of funds, i.e., $(x_1^t, y_1^t), \ldots, (x_N^t, y_N^t) \in \mathbb{R}^m \times \mathbb{R}^s$, are used to estimate MTPPS. The nonparametric estimator of the k-th ($k \in \{1, \ldots, \Gamma\}$) GTPPS only uses $n_k \leq N$ (where $\sum_{k=1}^{\Gamma} n_k = N$) for these observations of funds that invest one or more specific asset markets. To identify these particular observations, consider the one-to-one index function $\phi_k : \{1, \ldots, n_k\} \to \{1, \ldots, N\}$. Then $(x_{\phi_k(j)}, y_{\phi_k(j)})$ denotes the j-th observation in the set of funds used to estimate the k-th GTPPS. For example, consider the case where the nonparametric estimator of the k-th GTPPS only uses the four observations $(x_2, y_2), (x_4, y_4), (x_5, y_5)$ and (x_7, y_7). It means that $n_k = 4$ and $\phi_k(1) = 2, \phi_k(2) = 4, \phi_k(3) = 5$ and $\phi_k(4) = 7$.

4.3.1 Group-specific Technology and Group-specific Shortage Function

The group-specific technology is represented by the GTPPS, which is a set containing all input-output combinations that are possible using a given technology.

Transposing this concept into the context of fund evaluation, this GTPPS can be represented as the set consisting of input-output combinations of observed funds with a specific regulatory and strategic framework. As explained earlier, this specific framework largely determines certain underlying asset markets in which funds can select to invest (this selection on available underlying markets is regarded as the technology that funds use). The nonparametric frontier-based rating methods have been widely used to gauge the financial performance of funds and these evaluations are done mostly using frontier-based models which originate from production theory. We begin by considering the nonparametric frontier estimations of these GTPPSs under the assumption of variable returns to scale (VRS). Following Briec, Kerstens, and Vanden Eeckaut (2004), a unified algebraic representation of the k-th ($k \in \{1, \ldots, \Gamma\}$) GTPPS under the VRS assumption for the set of n_k funds using technology k at time t is:

$$P_{\Lambda}^{k,t} = \{(x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s \mid \forall i \in \{1, \ldots, m\} : x^t \geq \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} x^t_{i\phi_k(j)}, \\ \forall r \in \{1, \ldots, s\} : y^t \leq \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} y^t_{r\phi_k(j)}, \lambda_{\phi_k} \in \Lambda\}, \quad (4.3.1)$$

where, $\Lambda \equiv \Lambda^C = \{\lambda_{\phi_k} \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} = 1 \text{ and } \forall j \in \{1, \ldots, n\} : \lambda_{\phi_k(j)} \geq 0\}$ if convexity is assumed, and
$\Lambda \equiv \Lambda^{NC} = \{\lambda_{\phi_k} \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} = 1 \text{ and } \forall j \in \{1, \ldots, n\} : \lambda_{\phi_k(j)} \in \{0, 1\}\}$ if nonconvexity is assumed.

At time t, if there exists an input-output combination $(\sum_{j=1}^{n} \lambda_{\phi_k(j)} x^t_{i\phi_k(j)}, \sum_{j=1}^{n} \lambda_{\phi_k(j)} y^t_{i\phi_k(j)})$ in the convex or nonconvex GTPPS using less inputs and producing more outputs than the observed fund, then this fund is considered group-specific technology inefficient since it can improve its inputs and/or outputs against the k-th group-specific frontier. Based on the above GTPPS, the k-th group-specific shortage function of any observed fund at time t is now defined as follows:

Definition 4.3.1. *At time t, let $g^t = (-g_x^t, g_y^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ and $g^t \neq 0$. For any observation $z^t = (x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s$, the group-specific shortage function $S_\Lambda^{k,t}$ when using technology k in the direction of vector g^t is defined as:*

$$S_\Lambda^{k,t}(z^t; g^t) = \sup\{\beta \in \mathbb{R} \mid z^t + \beta g^t \in P_\Lambda^{k,t}\}.$$

For a fund with index $o \in \{1, \ldots, N\}$, the k-th group-specific shortage function with direction vector $g_o^t = (-g_{xo}^t, g_{yo}^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ can be determined from the following program:

$$\begin{aligned}
\max \quad & \beta \\
\text{s.t.} \quad & \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} x_{i\phi_k(j)}^t \leq x_{io}^t - \beta g_{io}^t, \quad i = 1, \ldots, m, \\
& \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} y_{i\phi_k(j)}^t \geq y_{ro}^t + \beta g_{ro}^t, \quad r = 1, \ldots, s, \\
& \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} = 1, \quad \beta \geq 0, \\
& \forall j = 1, \ldots, n_k : \begin{cases} \lambda_{\phi_k(j)} \geq 0, & \text{under convexity,} \\ \lambda_{\phi_k(j)} \in \{0, 1\}, & \text{under nonconvexity.} \end{cases}
\end{aligned} \tag{4.3.2}$$

Note that model (4.3.2) results in a linear programming (LP) problem under convexity and a binary integer linear programming (MBILP) problem under nonconvexity. Then both models allow to project the input-output combination of the fund o in such a way that inputs are decreased and outputs are increased in the direction g_o^t against the k-th group-specific frontier. The optimal solution β^* of model (4.3.2) measures how many times the direction vector g_o^t fits in the line segment from the input-output combination of the fund o to this frontier in the direction of g_o^t. Remarkably, if $(x_o^t, y_o^t) \in P_\Lambda^{k,t}$, then model (4.3.2) has a finite op-

timal solution and we have $\beta \geq 0$. If $(x_o^t, y_o^t) \notin P_\Lambda^{k,t}$, then either model (4.3.2) is infeasible or its optimal value is strictly less than 0.

Differing from fund ratings in a single-time framework, fund ratings in a multi-time framework consider performance over a time horizon consisting of multiple discrete time periods. The logic is now for observed funds to simultaneously expand its multiple outputs and decrease its multiple inputs over all discrete times. To develop the nonparametric frontier rating models in this multi-time framework, some definitions and properties are presented. Let T denote the number of consecutive times in a time horizon of interest. In addition, define a multi-time path of inputs and outputs as $Z_j = (x_j^t, y_j^t)_{t=1}^T$ for fund j, $(j = 1, \ldots, N)$, where $x_j^t = (x_{1j}^t, \ldots, x_{mj}^t)$ and $y_j^t = (y_{1j}^t, \ldots, y_{sj}^t)$ represent m inputs and s outputs at time t, respectively. Assuming VRS for all times $t \in \{1, \ldots, T\}$ and strong free disposability of all inputs and outputs, the k-th ($k \in \{1, \ldots, \Gamma\}$) multi-time GTPPS for the set of n_k funds using technology k can be defined as:

$$\mathbf{P}_\Lambda^{k,T} = P_\Lambda^{k,1} \times \cdots \times P_\Lambda^{k,T} \subset (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m \times T} \times \mathbb{R}^{s \times T}, \tag{4.3.3}$$

where $P_\Lambda^{k,t}$ $(t = 1, \ldots, T, k = 1, \ldots, \Gamma)$, is the k-th GTPPS at time t mentioned previously in (4.3.1).

In particular, this multi-time GTPPS is mathematically represented as follows

$$\mathbf{P}_{\Lambda}^{k,T} = \{(x^t, y^t)_{t=1}^T \in \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T} \mid \forall i \in \{1,\ldots,m\}, \forall t \in \{1,\ldots,T\} :$$
$$x^t \geq \sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t x_{i\phi_k(j)}^t,$$
$$\forall r \in \{1,\ldots,s\}, \forall t \in \{1,\ldots,T\} :$$
$$y^t \leq \sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t y_{r\phi_k(j)}^t,$$
$$\forall t \in \{1,\ldots,T\} : \lambda_{\phi_k}^t \in \Lambda\}, \tag{4.3.4}$$

where $\Lambda \equiv \Lambda^C = \{\lambda_{\phi_k}^t \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t = 1 \text{ and } \forall j \in \{1,\ldots,n\} : \lambda_{\phi_k(j)}^t \geq 0\}$ if convexity is assumed, and
$\Lambda \equiv \Lambda^{NC} = \{\lambda_{\phi_k}^t \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t = 1 \text{ and } \forall j \in \{1,\ldots,n\} : \lambda_{\phi_k(j)}^t \in \{0,1\}\}$ if nonconvexity is assumed.

To allow for a general definition for the multi-time shortage function, we first introduce some abbreviating notations. The time dependent direction vector denoted by $G = (g^1,\ldots,g^T) \in (\mathbb{R}_-^m \times \mathbb{R}_+^s)^T \cong \mathbb{R}_-^{m\times T} \times \mathbb{R}_+^{s\times T}$ represents a given multi-time direction path, where $g^t = (-g_x^t, g_y^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ represents the direction vector at time $t \in \{1,\ldots,T\}$. In addition, we denote $\Theta = (\beta_1,\ldots,\beta_T) \in \mathbb{R}^T$ and $\Theta \cdot G = (\beta_1 g^1,\ldots,\beta_T g^T) \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$. Considering the time preference of an investor in a portfolio context, we introduce a time discounting factor denoted ξ $(0 < \xi < 1)$ to weight the efficiency measures over the time horizon.

Then, based on the logic of the multi-time rating, the k-th multi-time group-specific shortage function assuming convexity or nonconvexity is defined as follows:

Definition 4.3.2. *With the notations introduced above, for any observation* $Z \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$*, the multi-time group-specific shortage function* S_Λ^T

when using technology k in the multi-time direction path of G is defined as:

$$S_{\Lambda}^{k,T}(Z;G) = \sup\left\{\frac{1}{T}\sum_{t=1}^{T}\xi^{T-t}\beta_t \mid Z + \Theta \cdot G \in \boldsymbol{P}_{\Lambda}^{k,T}\right\}.$$

For a given time horizon T, this amounts to looking for the largest arithmetic mean of time discounted distances over all times in a given time horizon of the input-output combinations of an observed fund to boundary of $\mathbf{P}_{\Lambda}^{k,T}$, which is called the multi-time group-specific frontier. If the value of the multi-time group-specific shortage function $S_{\Lambda}^{k,T}(Z;G) > 0$ for the input-output path Z of the observed fund, then it means that its inputs and outputs can be reduced and improved simultaneously in one or more time periods against the k-th multi-time group-specific frontier.

For a given multi-time direction path $G = (g^t)_{t=1}^{T} \in \mathbb{R}_{-}^{m\times T} \times \mathbb{R}_{+}^{s\times T}$, the k-th multi-time group-specific shortage function of a fund with index $o \in \{1,\ldots,N\}$ can be determined as the optimal value of the following programming:

$$\begin{aligned}
\max \quad & \frac{1}{T}\sum_{t=1}^{T}\xi^{T-t}\beta_t \\
\text{s.t.} \quad & \sum_{j=1}^{n_k}\lambda_{\phi_k(j)}^{t}x_{i\phi_k(j)}^{t} \le x_{io}^{t} - \beta_t g_{io}^{t}, \quad i=1,\ldots,m, \quad t=1,\ldots,T, \\
& \sum_{j=1}^{n_k}\lambda_{\phi_k(j)}^{t}y_{i\phi_k(j)}^{t} \ge y_{ro}^{t} + \beta_t g_{ro}^{t}, \quad r=1,\ldots,s, \quad t=1,\ldots,T, \\
& \sum_{j=1}^{n_k}\lambda_{\phi_k(j)}^{t} = 1, \quad \beta_t \ge 0, \quad t=1,\ldots,T, \\
& \forall j=1,\ldots,n_k: \begin{cases} \lambda_{\phi_k(j)}^{t} \ge 0, \quad t=1,\ldots,T, & \text{under convexity,} \\ \lambda_{\phi_k(j)}^{t} \in \{0,1\}, \quad t=1,\ldots,T, & \text{under nonconvexity.} \end{cases}
\end{aligned} \tag{4.3.5}$$

We add two remarks on computational issues. First, while in principle several options are available for the choice of direction vector, we opt here to employ the observation under evaluation itself, that is, $g^t = (-|x_{1o}^t|,\ldots,-|x_{mo}^t|,|y_{1o}^t|,\ldots,|y_{so}^t|)$.

In this case, the shortage function measures the maximum percentage of simultaneous risk reduction and expected return augmentation. Second, the block-diagonal structure of the above mathematical programming is a consequence of the time separability assumption, since there are no temporal linkages between the estimated fund problems for each time period. This structure basically allows us to solve the single-time mathematical programming for each time separately and to compute the objective function of the above problem based on the optimal solutions of these T sub-problems at the end (see also Briec, Comes, and Kerstens (2006) for a similar case in a production context).

4.3.2 Metatechnology and Metatechnology Shortage Function

The metatechnology is characterised by the set of all technologies that exist in a given time (e.g. O'Donnell (2016, p. 87)). In the metafrontier literature, technology sets are more often referred to as metatechnology. The set of all input and output vectors that are feasible using a given technology set (i.e., using some technology that is contained in a metatechnology) is labelled a MTPPS. As stated previously, a group-specific technology of funds is considered as a kind of selection possibility for the underlying markets. The metatechnology for funds contains all these selection possibilities for the available underlying markets being invested.

Let Γ represent the number of group-specific technologies contained in the metatechnology. Following the above notations, this associated MTPPS is mathematically defined as the union of all possible GTPPSs for a given time t, i.e.,

$M_\Lambda^{\Gamma,t} = \cup_{k\in\Gamma} P_\Lambda^{k,t}$. Equivalently,

$$M_\Lambda^{\Gamma,t} = \{(x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s \mid \forall i \in \{1,\dots,m\} : x^t \geq \sum_{k\in\Gamma}\sum_{j=1}^{n_k} \lambda_{\phi_k(j)} x^t_{i\phi_k(j)},$$
$$\forall r \in \{1,\dots,s\} : y^t \leq \sum_{k\in\Gamma}\sum_{j=1}^{n_k} \lambda_{\phi_k(j)} y^t_{r\phi_k(j)},$$
$$\sum_{k\in\Gamma}\sum_{j=1}^{n_k} \lambda_{\phi_k(j)} = 1, \lambda_{\phi_k} \in \Lambda\}, \tag{4.3.6}$$

where:
$\Lambda \equiv \Lambda^C = \{\lambda_{\phi_k} \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} = 1 \text{ and } \forall j \in \{1,\dots,n_k\} : \lambda_{\phi_k(j)} \geq 0\}$ if convexity is assumed for all GTPPSs, and
$\Lambda \equiv \Lambda^{NC} = \{\lambda_{\phi_k} \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)} = 1 \text{ and } \forall j \in \{1,\dots,n_k\} : \lambda_{\phi_k(j)} \in \{0,1\}\}$ if nonconvexity is assumed for all GTPPSs.

Even though each GTPPS is either convex or nonconvex, the MTPPS is generally not a convex set. Hence, this MTPPS does not inherit the convexity property of the GTPPSs. The boundary of the above MTPPS is called a metafrontier at a fixed time t.

We further define the shortage function based on the MTPPS for any observed fund, which is considered as an important extension of the shortage function for measuring the performance of nonhomogeneous funds. Using the MTPPS $M_\Lambda^{\Gamma,t}$, the metatechnology shortage function for any observed fund with the set of technology at time t is mathematically defined as follows.

Definition 4.3.3. *At time t, let $g^t = (-g_x^t, g_y^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ and $g^t \neq 0$. For any observation $z^t = (x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s$, the metatechnology shortage function $S_\Lambda^{\Gamma,t}$ in the direction of vector g^t is defined as:*

$$S_\Lambda^{\Gamma,t}(z^t; g^t) = \sup\{\delta \in \mathbb{R} \mid z^t + \delta g^t \in M_\Lambda^{\Gamma,t}\}.$$

From the definitions of the group-specific and metatechnology shortage functions in the single-time rating framework, one can easily conclude the following two properties. Firstly, the value of every group-specific shortage function is less than or equal to the value of the metatechnology shortage function. This is because every GTPPS is contained in the MTPPS. Secondly, it suffices to get the minimum value of all available group-specific shortage functions to determine the metatechnology shortage function. Referring to Kerstens, O'Donnell, and Van de Woestyne (2019, p. 783), this latter result is simply a transposition of the well-known results in the metafrontier context. Thus, we have:

$$S_{\Lambda}^{\Gamma,t}(z^t; g^t) = \max_{k \in \Gamma} \{S_{\Lambda}^{k,t}(z^t; g^t)\}. \tag{4.3.7}$$

For the input-output combination $z_o^t = (x_o^t, y_o^t)$ of an observed fund at time t, equation (4.3.7) suggests an enumeration algorithm for calculating the value of the metatechnology shortage function. For each $k (k \in 1, \ldots, \Gamma)$, we solve model (4.3.2) to identify the corresponding group-specific shortage function $S_{\Lambda}^{k,t}(z_o^t; g_o^t)$, and subsequently take the maximum of all the obtained group-specific shortage functions across all groups k, that is, the metatechnology shortage function $S_{\Lambda}^{\Gamma,t}(z^t; g^t)$ for this fund.

In the following, we develop the metatechnology shortage function in a multi-time framework. First, analogous to the idea of defining the multi-time GTPPS, the multi-time MTPPS is mathematically defined in the following formulation:

$$\mathbf{M}_{\Lambda}^{\Gamma,T} = M_{\Lambda}^{\Gamma,1} \times \cdots \times M_{\Lambda}^{\Gamma,T} \subset (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m \times T} \times \mathbb{R}^{s \times T}, \tag{4.3.8}$$

where $M_{\Lambda}^{\Gamma,t}$, $(t = 1, \ldots, T)$, is the MTPPS at time t mentioned previously in (4.3.6).

Equally, the specific formulation of this multi-time MTPPS is expressed as

$$\begin{aligned}\mathbf{M}_{\Lambda}^{\Gamma,T} = \big\{(x^t, y^t)_{t=1}^{T} \in \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T} \mid \forall i \in \{1,\ldots,m\}, \forall t \in \{1,\ldots,T\} :\\ x^t \geq \sum_{k\in\Gamma}\sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t x_{i\phi_k(j)}^t,\\ \forall r \in \{1,\ldots,s\}, \forall t \in \{1,\ldots,T\} :\\ y^t \leq \sum_{k\in\Gamma}\sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t y_{r\phi_k(j)}^t,\\ \forall t \in \{1,\ldots,T\} : \sum_{k\in\Gamma}\sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t = 1, \lambda_{\phi_k}^t \in \Lambda\big\}, \quad (4.3.9)\end{aligned}$$

where:

$\Lambda \equiv \Lambda^C = \{\lambda_{\phi_k}^t \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t = 1 \text{ and } \forall j \in \{1,\ldots,n_k\} : \lambda_{\phi_k(j)}^t \geq 0\}$ if convexity is assumed for all GTPPSs, and

$\Lambda \equiv \Lambda^{NC} = \{\lambda_{\phi_k}^t \in \mathbb{R}^{n_k} \mid \sum_{j=1}^{n_k} \lambda_{\phi_k(j)}^t = 1 \text{ and } \forall j \in \{1,\ldots,n_k\} : \lambda_{\phi_k(j)}^t \in \{0,1\}\}$ if nonconvexity is assumed for all GTPPSs.

The definition of a multi-time metatechnology shortage function does the same as a single-time shortage function, but it spans over multiple times: it simultaneously seeks to expand multiple return dimensions and contract multiple risk dimensions over all times for all funds with the technology set. Based on the multi-time MTPPS, the multi-time metatechnology shortage function for any observed fund in the given time horizon T is defined as follows:

Definition 4.3.4. *With the notations introduced above, for any observation* $Z \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$*, the multi-time metatechnology shortage function* S_{Λ}^{T} *in the direction path of* G *is defined as:*

$$S_{\Lambda}^{\Gamma,T}(Z;G) = \sup\left\{\frac{1}{T}\sum_{t=1}^{T} \xi^{T-t}\delta_t \mid Z + \Theta \cdot G \in \boldsymbol{M}_{\Lambda}^{k,T}\right\}.$$

This multi-time metatechnology shortage function is easily calculated, because

it simply corresponds to the largest arithmetic mean of time discounted single-time metatechnology shortage functions (see (4.3.7)) over the whole time horizon. It is clear that the value of every multi-time group-specific shortage function (see Definition 4.3.2) is less than or equal to the value of the multi-time metatechnology shortage function. This result is consistent with the single-time context.

Referring to Briec and Kerstens (2009), we also allow the virtual weights $\lambda^t_{\phi_k(j)}(t = 1, \ldots, T)$ (see Definition 4.3.4 and formulae (4.3.9)) to vary over time to look for a periodical benchmark over the given time T horizon, taking the short-run investment perspective. If one prefers to seek a stable benchmark in this time horizon following a long-term investment perspective, then one can simply impose the constraint on constant virtual weights (i.e., $\lambda^1_{\phi_k(j)} = \cdots = \lambda^T_{\phi_k(j)}$) across time periods.

4.3.3 Technology Efficiency and Technology Difference Gap

Recall from Section (4.3.2) that in the single-time framework, the fund under assessment is efficient with respect to the metatechnology that contains Γ available underlying markets, if and only if the metatechnology shortage function $S^{\Gamma,t}_{\Lambda} = 0$ for its input-output combination $z^t = (x^t, y^t)$ at a fixed time t. While in the multi-time framework, this fund is efficient with the metatechnology if and only if the multi-time metatechnology shortage function $S^{\Gamma,T}_{\Lambda} = 0$ for its multi-time path of input-output combinations $Z = (x^t, y^t)^T_{t=1}$ in the give time horizon T.

More specially, the single-time metatechnology efficiency (MTE) of the fund under evaluation at time t is defined as the metatechnology shortage function (see Definition 4.3.3):

$$\mathrm{MTE}^{\Gamma,t}(z^t; g^t) = S^{\Gamma,t}_{\Lambda}(z^t; g^t). \tag{4.3.10}$$

This single-time MTE measure lies in the closed unit interval (i.e., $\text{MTE}^{\Gamma,t}(z^t; g^t) \in [0, 1]$) and indicates the maximum proportion for a fund with respect to reduction in x^t and extension in y^t along the direction of g^t using a set of group-specific technologies Γ (i.e., all available investment asset markets across different regulatory and strategic frameworks). We can also measure the single-time group-specific technology efficiency (GTE) for this fund with respect to the k group-specific frontier based on the group-specific technology shortage function (see Definition 4.3.1). The associated measure of the single-time GTE is:

$$\text{GTE}^{k,t}(z^t; g^t) = S_{\Lambda}^{k,t}(z^t; g^t). \tag{4.3.11}$$

This measure also lies in the closed unit interval and indicates the maximum proportion of input reductions and output increases when using a group-specific technology k at time t. Following O'Donnell, Rao, and Battese (2008), the translated metatechnology frontier envelops translated group-specific frontiers for all $k = 1, \ldots, \Gamma$. Hence, the value of single-time GTE has to always be less than or equal to that of single-time MTE along the prespecified direction g^t. Based on this property, we define the difference between both of them as the single-time technology difference gap (TDG):

$$\text{TDG}^{k,t}(z^t; g^t) = \text{MTE}^{\Gamma,t}(z^t; g^t) - \text{GTE}^{k,t}(z^t; g^t), \quad k = 1, \ldots, \Gamma. \tag{4.3.12}$$

The single-time TDG also lies in the closed unit interval. The TDG in the production context is considered as a direction-oriented measure of whether a unit has chosen the best group-specific technology that is available at time t. Naturally, this indicator shifted to the context of portfolio evaluation can be regarded as a direction-oriented measure of how well the fund has initially chosen its underlying investment markets among the available options across groups that have distinct

regulatory and strategic limitations. TDG = 0 indicates that the fund's performance would not be restricted by its regulatory and strategic framework, as it has opted for the best possible investment compared to the other groups of funds. TDG > 0 means that the current regulatory and strategic framework somewhat limits the fund to realize counterfactual or potential gains.[4]

By analogy with the single-time case, the multi-time MTE of the observed fund with the multi-time path of input-output combinations $Z_j = (x_j^t, y_j^t)_{t=1}^T$ for the given time horizon T is defined as the multi-time metatechnology shortage function (see Definition 4.3.4):

$$\mathrm{MTE}^{\Gamma,T}(Z;G) = S_\Lambda^{\Gamma,T}(Z;G). \tag{4.3.13}$$

This multi-time MTE measure indicates the largest arithmetic mean of simultaneous reductions in inputs and expansions in outputs into a path of direction G such that an observed input-output path of the fund is projected on the corresponding metatechnology frontier at each time t $(t = 1, \ldots, T)$. In theory, the multi-time MTE is lager than the single-time MTE due to the single-time MTPPS is a subset of the multi-time MTPPS. The multi-time MTE allows for testing whether the fund has opted for the best possible investment among available options across groups that have certain regulatory and strategic limitations at multiple times in this horizon T. Correspondingly, the multi-time group-specific GTE for a fund under evaluation is defined as the multi-time group-specific technology shortage

[4]The TDG only measures the counterfactual benefit assuming that the fund had been able to invest in different underlying financial markets. However, this indicator is not sufficient for analyzing the efficiency of the underlying financial markets in which different groups of funds invest, since we analyze the relative performance of the fund with respect to other funds rather than with respect to the underlying financial markets.

function with respect to group k (see Definition 4.3.2):

$$\mathrm{GTE}^{k,T}(Z;G) = S_{\Lambda}^{k,T}(Z;G). \tag{4.3.14}$$

This measure indicates the largest arithmetic mean of time discounted distances over all times in the given time horizon T of the input-output combinations of this observed fund to the k-th group-specific frontier at each time t ($t = 1, \ldots, T$). It is easily proven that the multi-time GTE is lager than the single-time GTE due to the single-time GTPPS is a subset of the multi-time GTPPS. Like in the single-time framework, the multi-time TDG as the gap between the multi-time MTE and multi-time GTE can be represented as follows:

$$\mathrm{TDG}^{k,T}(Z;G) = \mathrm{MTE}^{\Gamma,T}(Z;G) - \mathrm{GTE}^{k,T}(Z;G), \quad k = 1, \ldots, \Gamma. \tag{4.3.15}$$

Since the metafrontier envelops the k-th ($k \in 1, \ldots, \Gamma$) group-specific frontier for each time t within the time horizon T, one can easily obtain that the multi-time TDG is non-negative. The TDG determined in the multi-time framework is in theory either higher or lower than the one determined in the single-time framework. The multi-time TDG measures how well the fund has initially chosen its investment asset universe among the available options over multiple times in this time horizon T.

The setting in the previous discussion is general and flexible and can thus handle a large choice of inputs and outputs. We now particularize the above formulation to characterize the metafrontier in the MV, MVS and MVSK spaces. Suppose that there are N funds under evaluation. At time t, let $R_1^t, \ldots, R_N^t$ denote the random returns of the N funds, which are characterized by the mean $E(R_j^t)$, variance $V(R_j^t)$, skewness $S(R_j^t)$ and kurtosis $K(R_j^t)$ for $j \in \{1, \ldots, N\}$. Here, the

calculations of variance, skewness and kurtosis are expressed as follows: $V(R_j^t) = E[(R_j^t - E(R_j^t))^2]$, $S(R_j^t) = E[(R_j^t - E(R_j^t))^3]$, and $K(R_j^t) = E[(R_j^t - E(R_j^t))^4]$.

To obtain detailed specifications of the MTPSS defined in (4.3.6) and GTPPSs defined in (4.3.1), we need to classify the different goals of the investor in terms of either inputs (i.e., objectives to minimize), or outputs (i.e., objectives to maximize). As introduced earlier, the need for multiple moments is apparent to assess funds whose return distributions may exhibit strong asymmetry and fat tails. Given mixed risk-aversion utility functions, investors express a preference for odd moments and a dislike for even moments of the distribution of asset returns. Therefore, when the MVSK framework is considered, we can define the first and second inputs of funds as $x_{1j}^t = V(R_j^t)$ and $x_{2j}^t = K(R_j^t)$, and the first and second outputs as $y_{1j}^t = E(R_j^t)$ and $y_{2j}^t = S(R_j^t)$ for $j \in \{1, \ldots, n\}$. Obviously, for the MVS case only the first input is considered. In the multi-time framework, we select variance and kurtosis of each time t $(t = 1, \ldots, T)$, as inputs and mean and skewness as outputs, whereas for the MVS case only variance for each t is considered as inputs.

Finally, the above three indicators MTE, GTE and TDG can be determined based on these specifications of MTPPS and GTPPS. It is crucial to note that since the MTE of funds is measured with respect to the metafrontier, common to all funds in different groups (i.e., operating with distinct investment asset universes), their performance can be compared directly depending on this measurement. By means of MTE in the shortage function framework, the fund under evaluation can improve its mean and skewness dimensions and reduce its variance and kurtosis dimensions along a given direction g at a single time or along a direction path G over multiple times.

4.4 Empirical Illustration

In this section, we use actual data for different types (groups) of funds, which invest in distinct underlying markets, to investigate the impact of multiple moments and multiple times on the MTE estimate and its decompositions. Then, with the help of backtesting analysis, we aim to further illustrate the use of the proposed multi-time and multi-moment metafrontier rating methods for selecting funds across groups.

4.4.1 Sample Description

For the empirical analysis, we use the Lipper for Investment Management database. In particular, we extract a nonhomogeneous set of 717 funds composed of five types. These five distinct groups consist of hedge funds (HF, 141 observations), bond mutual funds (BMF, 175 observations), equity mutual funds (EMF, 180 observations), mix-asset mutual funds (MixMF, 86 observations), and money mutual funds (MonMF, 135 observations). To maximize the credibility of the empirical results, we choose these samples with a consistent geographic attribute, all belonging to the European universe. According to the classification criteria regarding asset types designed by the Lipper database, each group of funds is assigned an asset type based on the definitions below in Table (4.1).

Table 4.1: Description for the classification of funds

Asset Type	Description
HF	Funds using derivative instruments as the source of return.
BMF	Funds investing in fixed income markets.
EMF	Funds investing in stock markets.
MixMF	Funds investing in fixed income markets, with an average maturity > one year.
MonMF	Funds investing in fixed income markets, with an average residual life to maturity < 12 months.

For these funds, we collect 179 monthly returns from May 2006 to March 2021 with the availability of continuous data for at least 15 years, which is required for the use of our backtesting analysis in the newly proposed multi-time and multi-moment metafrontier ratings. It needs to be stated that we initially specify these nonparametric group-specific frontier estimates following the idea of Kerstens, Mounir, and Van de Woestyne (2011) that higher-order moments and cost components are included. However, since HF cost data is unavailable in this database, our empirical analysis is limited to focus on the characteristics of the return distributions for all these funds without considering cost factors. In the following, we make a basic comparative analysis of the monthly return characteristics of funds across the five groups for the entire sample period.

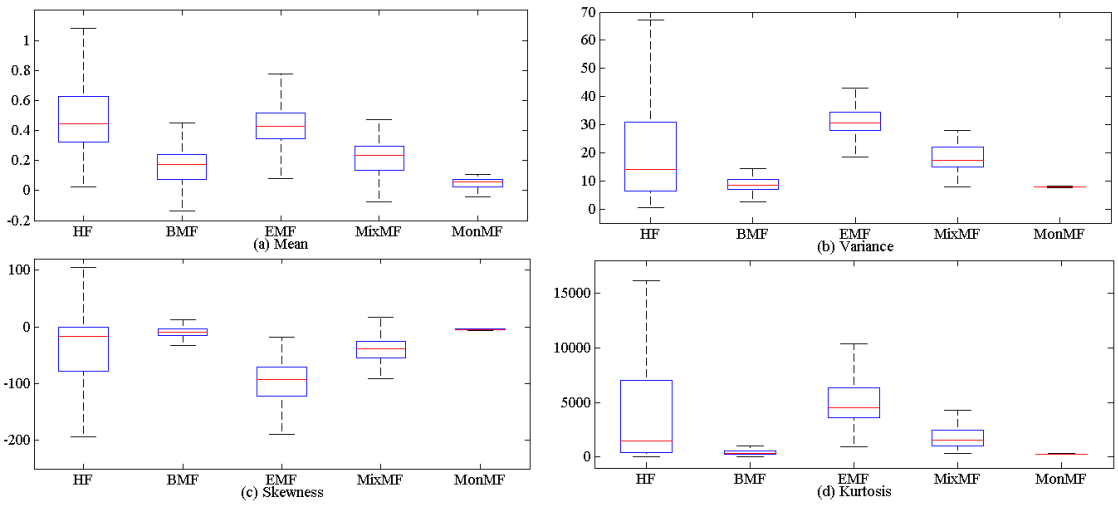

Figure 4.3: Box-plots of moments distribution for different types of funds

Figure 4.3 propose box-plots to intuitively summarize the entire distribution of different types of funds for each of the four moments. The sub-figures (a) to (d) correspond to the mean, variance, skewness and kurtosis of these funds. In these box-plots, the box indicates the interquartile range where the small vertical lines reporting the location of the median. Striking features are twofold. First, generally speaking, most funds are characterized by high volatility, very negative skewness and positive kurtosis, except for the BMF and MonMF. This indicates that the multi-moment metafrontier rating methods can be of great importance to select well-performing funds from a large fund universe. Second, relatively speaking, different types of funds show a marked difference regarding the four moments distribution. The return distribution of the HF generally exhibits fairly obvious moment characteristics. Specifically, the interquartile ranges of the four moments for 141 HF samples are quite large, and the distributions of mean and variance but especially skewness and kurtosis are quite asymmetric and have long tails. In contrast, the MonMF present the most insignificant moment characteristics.

Compared to funds investing in the other four asset markets, funds investing in the money markets exhibit consistently a low return-risk level, especially in terms of the two moments of skewness and kurtosis. As to the BMF, EMF and MixMF clearly experience a slightly asymmetric, a very asymmetric, and a mildly asymmetric distribution.

4.4.2 Evaluation Results

In this subsection, we shed light on whether the multiple moments and multiple times have an impact on the MTE and it decomposition (GTE and TDG) for these different types of funds. First, the three indices are estimated in the single- and multi-time frameworks to illustrate the effect of considering multiple times. Second, the distribution of differences of the three indices are examined separately when adding multiple moments. In the single-time framework, we extract the monthly returns of these samples for the past 5 years to date to calculate the estimates of MTE, GTE and TDG. While in the multi-time framework, the monthly returns for the past 1 year, 3 years and 5 years to date are integrated and applied to estimate the three indices.

Table 4.2 contains basic descriptive statistics of the MTE estimates and its components (i.e., GTE and TDG estimates) with the unbalanced panel data of funds across the five groups in 12 various rating frameworks. Here, let us introduce the abbreviations used in Table 4.2. The acronym nc-c indicates that the estimated GTPPSs are convex and the estimated MTPPS is not, while the acronym nc-nc indicates that both the estimated TPPSs and the estimated MTPPS are nonconvex. The acronyms ST and MT indicate that the estimates are calculated in the single- and multi-time frameworks, respectively. Similarly, the acronyms MV, MVS and MVSK indicate that the estimates are calculated in the MV, MVS and MVSK

frameworks, respectively.

Remarkably, even though we collect these nonparametric metafrontier rating methods corresponding to the estimated GTPPSs using both nonparametric convex and nonconvex estimators, it should be realized that imposing the convexity on the GTPPS is not always useful. Liu, Zhou, Liu, and Xiao (2015) state that a convex VRS nonparametric frontier estimator provides an inner approximation to the portfolio frontier derived from the traditional MV diversified model. In the more general case where we want to explore a nonconvex higher moment portfolio model, then one can argue that the nonconvex nonparametric frontier estimators with VRS (already advocated by Kerstens, Mounir, and Van de Woestyne (2011)) provides a conservative underestimation of the corresponding nonconvex diversified portfolio model within some common subspace of moments (see also Germain, Nalpas, and Vanhems (2011)). By contrast, the more widely used convex nonparametric frontier specification may in this case overestimate the corresponding nonconvex diversified portfolio model within the common subspace of moments. The latter argument seems to have escaped attention so far: This explains why most nonparametric frontier rating models with higher moments do impose convexity (for instance, Gregoriou, Sedzro, and Zhu (2005)). In this contribution, we are also particularly interested in the effects of the convexity assumption of the GTPPS on the MTE and it corresponding components.

Table 4.2 is structured as follows: First, we discuss the columns in Table 4.2. Both nc-c and nc-nc results are reported in two parallel blocks of six columns. Within each of both results, the first, second and last two columns report the results in the MV, MVS and MVSK frameworks, respectively. A further distinction is related to whether multiple time periods are considered: ST indicates a single time is considered to calculate these estimates, while MT refers to multiple times

being considered. Second, turning to the explanation of the rows in Table 4.2, the first horizontal block of results contains summary statistics regarding the MTE estimates for all 717 observations in the sample. The following two horizontal blocks of rows report the GTE and TDG estimates. The first row in each horizontal block reports the number of efficient observations (i.e., the number of times the relevant measure is estimated to be 0). The next three rows in each horizontal block report the arithmetic means, standard deviations, and medians of the relevant estimates.

Table 4.2: Descriptive statistics and Li-test comparing the estimates of MTE and decompositions in single- and multi-time frameworks

		nc-c						nc-nc					
		MV		MVS		MVSK		MV		MVS		MVSK	
		ST	MT	ST	MT	ST	MT	ST	MT	ST	MT	ST	MT
MTE	#Ef. Obs.	12	0	18	2	30	6	24	3	78	12	134	20
	Arith.Mean	0.6308	0.5703	0.4924	0.4535	0.4182	0.3726	0.5985	0.5125	0.3101	0.2574	0.3049	0.2489
	Stand.Dev.	0.3223	0.2094	0.2706	0.1748	0.2961	0.1813	0.3387	0.2251	0.3203	0.2083	0.3242	0.2105
	Med.	0.7055	0.6414	0.4658	0.4691	0.3175	0.3040	0.6448	0.5951	0.1924	0.1863	0.1887	0.1774
	Li-test		111.43		13.86		49.86		109.70		59.90		71.20
	p-value		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
GTE	#Ef. Obs.	22	3	34	9	55	19	63	14	181	59	211	84
	Arith.Mean	0.3799	0.3535	0.3301	0.2793	0.3172	0.2561	0.3191	0.2913	0.2257	0.1754	0.2244	0.1723
	Stand.Dev.	0.3101	0.2358	0.2867	0.2028	0.2902	0.1977	0.3117	0.2258	0.2889	0.1958	0.2894	0.1956
	Med.	0.3047	0.3079	0.2876	0.2565	0.2715	0.2343	0.2321	0.2410	0.1351	0.1138	0.1294	0.1117
	Li-test		32.44		28.88		32.14		54.84		21.03		17.31
	p-value		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
TDG	#Ef. Obs.	299	55	280	58	327	53	299	121	339	123	416	130
	Arith.Mean	0.2509	0.2168	0.1623	0.1742	0.1010	0.1165	0.2795	0.2212	0.0844	0.0819	0.0805	0.0767
	Stand.Dev.	0.3332	0.2482	0.1793	0.1612	0.1260	0.1068	0.3464	0.2423	0.1399	0.0870	0.1414	0.0853
	Med.	0.0856	0.1160	0.0915	0.1428	0.0205	0.0828	0.1271	0.1427	0.0085	0.0652	0.0000	0.0589
	Li-test		110.34		79.56		121.66		41.01		63.18		91.04
	p-value		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)

Several conclusions can be drawn from the descriptive statistics results reported in Table 4.2. First, the basic descriptive statistics for MTF, GTF and TDG estimates in the single-time framework versus those in the multi-time framework

all show some differences. For the MTF estimates, the multi-time MTF estimates are on average lower than the single-time MTF estimates in all settings (nc-c and nc-nc for MV & MVS & MVSK). Combining the numbers of efficient observations, the multi-time MTF estimates clearly yield fewer efficient observations compared to those based on the single-time MTF estimates. For the GTF estimations, this coincides with the above finding that the multi-time GTFs all show lower mean values and fewer efficient observations than the single-time ones. Theoretically, estimates of these two indices derived in the multi-time framework can be no higher than the ones derived in the single-time framework. This is consistent with the observations for the MTF and GTF estimates. As for the TDG estimate that is the gap between the corresponding MTE and GTE, this indicator obtained in the multi-time framework can in theory be either higher or lower than those obtained in the single-time framework. However, it is clear that the TDGs obtained in the multi-time framework all yield fewer efficient observations compared to those obtained in the single-time framework. Since the TDG captures the difference between the metafrontier and the group-specific frontier, this result implies that major funds in the multi-time rating framework would have the possibility to improve their performance if they could invest in other available underlying asset markets across different regulatory and strategic limitations.

Second, looking at the descriptive statistics for the estimates of three indices in the traditional MV framework versus those in the multi-moment (MVS & MVSK) framework, it is clear to observe that all of them show some differences when adding the higher-order moments. For the MTF estimates, the MTF obtained in the multi-moment case yield lower mean values and more efficient observations than those obtained in the MV case for all settings (nc-c and nc-nc for both single- and multi-times). For the GTF estimations, the multi-moment GTF estimations display lower mean value and more efficient observations compared to the MV

ones. By construction, the estimates of MTE and GTE obtained in the multi-moment case can be no higher than the estimates obtained using the MV case. This is validated by our observations for the MTF and GTF estimates. For the TDG estimates, while the TDG estimated in the multi-moment case show lower mean values and lager efficient observations compared to that estimated in the MV case for our data, in theory the former can be either higher or lower than the latter.

Third, in theory, the estimates of MTE and GTE obtained using the nc-nc model cannot be higher than those obtained using the nc-c model. From the comparison of the two parallel blocks (nc-c verse nc-nc) in Table 4.2, it reveals that in our application estimates of MTE obtained using the nc-nc model yield lower mean values and more efficient observations than those obtained using the nc-n model for all settings (MV & MVS &MVSK for both single- and multi-times). In addition, the difference between the MTE estimates obtained using the two models tend to be more pronounced when multi-moments and multi-times are considered than when only single-time mean and variance are considered. Similar results can be observed for the comparison between estimates of GTF using the nc-nc and nc-c models. The TDG estimated using the nc-nc model can in theory be either higher or lower than those obtained estimated using the nc-c model. In the single-time and MV framework, little difference can be found on the estimated TDG using the nc-nc and nc-c models in terms of the mean value and the number of efficient observations. On the contrary, when multiple moments and multiple times are included separately or jointly, a clear difference between the two TDG estimates is revealed for our data.

To formally assess the reported differences in the estimates, we employ a non-parametric test initially developed by Li (1996) and refined by Fan and Ullah

(1999) and most recently by Li, Maasoumi, and Racine (2009). This nonparametric test assesses the eventual statistical significance of differences between two kernel-based estimates of density functions f and g of a random variable x. The null hypothesis maintains the equality of both density functions almost everywhere: $H_0 : f(x) = g(x)$ for all x. The alternative hypothesis negates this equality of both density functions: $H_1 : f(x) \neq g(x)$ for some x. This test is valid for both dependent and independent variables: observe that dependency is a characteristic of frontier estimators (i.e., efficiency levels depend on sample size, among others).[5] The results are reported in the last rows of each horizontal block in Table 4.2.

We start by conducting the Li-tests to test the null hypothesis that the distributions of the single-time MTE vs. the multi-time MTE, as well as its two components GTF and TDG are equal. The Li-test result for the MTF estimates reveals that the multiple times lead to a significant impact in the distribution of the MTE estimates. The MTE estimates obtained in the multi-time framework and those obtained in the single-time framework are significantly different at the 1% significance level for all settings. A similar result can also be observed for both the GTE and TDG estimates. This suggests that the multiple times have a clear effect on the three estimates of MTE, GTE and TDG.

Keeping the impact of multiple times in mind, we now turn to test for the effect of adding multiple moments on the three indices. The Li-test result of adding higher-order moments in some basic models is reported in Table 4.3. To eliminate the possible disruptions from other variables (times and convexity), this table summarises the results in terms of four blocks: nc-c and nc-nc, with both single time and multiple times. From Table 4.3, it can be concluded that moments greater than two (mean and variance) all contribute significantly to the differences

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

in the distribution of MTE, GTE and TDG estimates at a 1% significance level, except for adding skewness and kurtosis jointly in the case of nc-c with single time. Turning to a comparison between these indices in the MVS versus MVSK framework, the MTE estimates computed in the MVS and MVSK frameworks yield significantly different efficiency distributions. This somewhat indicates that adding kurtosis has a significant impact on the estimates of MTE. A similar conclusion is found for the TDG estimates. In contrast, for the GTE estimates, adding kurtosis does not contribute in a significant way to the distribution of GTE estimates.

Table 4.3: Li-test comparing the estimates of MTE and decompositions with different moments

nc-c				nc-nc			
ST				ST			
Order	MV → MVS	MV → MVSK	MVS → MVSK	Order	MV → MVS	MV → MVSK	MVS → MVSK
MTE	62.24	-2.17	37.83	MTE	69.17	83.94	20.18
	(0.000)	(0.993)	(0.000)		(0.000)	(0.000)	(0.000)
GTE	32.97	35.28	0.25	GTE	24.89	28.94	-1.49
	(0.000)	(0.000)	(0.358)		(0.000)	(0.000)	(1.000)
TDG	11.73	7.05	10.10	TDG	10.91	16.50	7.35
	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
MT				MT			
Order	MV → MVS	MV → MVSK	MVS → MVSK	Order	MV → MVS	MV → MVSK	MVS → MVSK
MTE	56.49	59.26	50.32	MTE	92.29	92.13	6.37
	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
GTE	48.69	46.73	-5.75	GTE	65.05	73.56	-1.19
	(0.000)	(0.000)	(1.000)		(0.000)	(0.000)	(0.984)
TDG	41.70	36.67	24.91	TDG	19.30	18.96	2.26
	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.020)

Having explored the impact of multiple times and multiple moments on the distribution of MTE, GTE and TDG estimates, we now focus specifically on testing for the effects of the traditional convexity assumption for each GTPPS on the distribution of these three indices, respectively. Table 4.4 reports the Li-test statistics representing the distributional differences between nc-c and nc-nc models

in six different maintained settings (both single- and multi-time framework with MV & MVS &MVSK moments). One can clearly observe from Table 4.4 that the convexity axiom for each GTPPS does have a significant impact on the distribution of the MTE, GTE and TDG estimates, except for the MTE estimated in the single-time MV framework.

Table 4.4: Li-test comparing the estimates of MTE and decompositions computed by nc-c and nc-nc models

Indices	ST			MT		
	MV	MVS	MVSK	MV	MVS	MVSK
	nc-c vs. nc-nc			nc-c vs. nc-nc		
MTE	-0.07	70.30	60.72	42.08	68.34	42.38
	(0.451)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GTE	41.82	106.41	85.87	32.41	83.40	75.58
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TDG	5.08	14.87	8.72	45.30	38.42	34.26
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

From these analyses, we can basically conclude that the multiple moments and multiple times both separately and jointly have an impact on the estimates of MTE and its decomposition for our data. Furthermore, this impact is more significant when the two factors are considered jointly. Finally, the convexity assumption for each GTPPS is proven to have a clear impact on the estimated values of these three indices.

4.4.3 Backtesting Analysis

In this subsection, we use the 717 nonhomogeneous funds with 15-year continuous return data to test the application of the proposed metafrontier methods for rating and selecting funds across different groups. To this end, a comparative approach based on a backtesting methodology is adopted. Backtesting refers to executing fictitious investment strategies using historical data to simulate how these strategies would have performed if they had actually been adopted by fund managers in the past. It is powerful for evaluating and comparing the performance of different investment strategies without using real capital. Some examples of a backtesting approach are found in DeMiguel, Garlappi, and Uppal (2009), Tu and Zhou (2011), Brandouy, Kerstens, and Van de Woestyne (2015), Zhou, Xiao, Jin, and Liu (2018) and Lin and Li (2020), among others.

Apart from the 12 nonparametric matafrontier-based ratings discussed in the previous section, we also test some traditional financial indicators: Sharpe ratio, Sortino ratio and Omega ratio for comparison. The exact definition for the Sharpe, Sortino and Omega ratios are found in Feibel (2003, p. 187 and p. 200) and Eling and Schuhmacher (2007, p. 2635), respectively. In total, there are 15 fund rating methods (3 traditional financial rating methods plus 12 metafrontier-based rating methods) collected in our work to operate the backtesting exercises. Table 4.5 lists the rating methods considered in our work and their abbreviations.

Table 4.5: List of various rating models compared: 3 financial rating methods and 12 matafrontier rating methods

Models	Abbreviation
Traditional financial measures	
Sharpe ratio	Sharpe
Sortino ratio	Sortino
Omega ratio	Omega
Nonconvex MTPPSs and Convex GTPPSs	
Single-time and MV framework	STMVnc-c
Single-time and MVS framework	STMVSnc-c
Single-time and MVSK framework	STMVSKnc-c
Multi-time and MV framework	MTMVnc-c
Multi-time and MVS framework	MTMVSnc-c
Multi-time and MVSK framework	MTMVSKnc-c
Nonconvex MTPPSs and Nonconvex GTPPSs	
Single-time and MV framework	STMVnc-nc
Single-time and MVS framework	STMVSnc-nc
Single-time and MVSK framework	STMVSKnc-nc
Multi-time and MV framework	MTMVnc-nc
Multi-time and MVS framework	MTMVSnc-nc
Multi-time and MVSK framework	MTMVSKnc-nc

We consider a simple buy-and-hold backtesting strategy consisting of buying in each time the 10, 20 and 30 best performing funds ranked by any given rating method, respectively. Our work now is to empirically test the out-of-sample performance of these 15 buy-and-hold strategies. Since the Sharpe ratio and other relative performance measures are only suitable for the MV world, we opt for the shortage function as an absolute performance measure that is capable to assess

the performance of these strategies in multiple dimensions simultaneously (i.e., mean, variance, skewness and kurtosis). Hence, the 15 buy-and-hold backtesting strategies are compared based on the MVSK performance of their holding values evaluated by combining shortage functions with the single-time and multi-moment frontiers (with convexity and nonconvexity).

Backtesting setup: The details

Using 5 years of data to obtain our first metafrontier-based rankings and finance-based rankings, we start backtesting from May 2006 onwards. The first buy-and-hold strategy being made on that date, this strategy is repeated each month thereafter with an updated set of ratings to select the best funds. Thus, we use a rolling window of 5 years with a step of a single month to compute the ratings. This rating and the ensuing buy-and-hold process based on these selected funds is repeated 60 times (months) till the end of March 2016, where the 5 years before the end of the sample period are kept apart to test the long-term holding performance of the strategy in the last backtesting period. For each time window or each backtesting event, the steps can be detailed as follows:

(1) Adopt the 5-year time window of data to compute the single-time metafrontier ratings, as well as the traditional financial ratings. In combination with the other two time periods (i.e., 1-year and 3-year) of data from this time window, the multi-time metafrontier ratings are computed.

(2) Depending on the rating computed by this time window of data for each method, the 10, 20 or 30 best performing funds which may consist different type of funds are selected for the backtesting exercise, and then one holds these selected funds for 1 year, for 3 years, for 5 years, and till the end of

the whole sample period, respectively.

(3) In each of the above three selecting scenarios, we compute and store the complete historical track record of the holding values per buy-and-hold backtesting strategy, and then we calculate the mean, variance, skewness and kurtosis of these holding value series.

The above steps for backtesting are repeated over 60 time windows in total. For each of the four holding period scenarios, the performance of these MVSK observations (15 times 60 observations) that are generated by the 15 strategies over 60 backtesting exercises are all evaluated by the shortage function based on the single-time and MVSK frontier (with convexity and nonconvexity). In particular, we first establish the convex and nonconvex VRS nonparametric frontiers in the single-time and MVSK framework for these MVSK observations, and then measure their (in)efficiency scores using the shortage functions. Clearly, each buy-and-hold strategy yields the efficiency scores of 60 MVSK observations. The average efficiency score and the number of efficient units, as well as the distribution of inefficiency scores across these 60 observations, are adopted to evaluate the 15 strategies.

Backtesting Results

We begin by reporting the Kendall rank correlations to test the degree of concordance in rankings depending on the MTE estimated by 15 rating methods for each backtesting exercise, which delivers a 15 times 15 dimensions. Then, we have aggregated the Kendall rank correlations using a simple arithmetic mean of the computations in 60 time windows to report their overall degree of concordance in ranking. Table 4.6 shows the aggregated rank correlation between different rat-

ings. It is noted that the correlation coefficients between the different rankings are significantly different from zero at the 1 % significance level for each time window calculation.

Table 4.6: Kendall rank correlations comparing the rankings computed by different rating methods (3 financial rating methods and 12 metafrontier rating methods)

	Sortino	Omega	STMVnc-c	STMVSnc-c	STMVSKnc-c	MTMVnc-c	MTMVSnc-c	MTMVSKnc-c	STMVnc-nc	STMVSnc-nc	STMVSKnc-nc	MTMVnc-nc	MTMVSnc-nc	MTMVSKnc-nc
Sharpe	0.983	0.962	0.489	0.482	0.476	0.449	0.432	0.392	0.485	0.475	0.473	0.447	0.400	0.396
Sortino		0.952	0.490	0.487	0.480	0.449	0.435	0.394	0.487	0.481	0.479	0.447	0.403	0.399
Omega			0.478	0.472	0.467	0.440	0.422	0.383	0.476	0.465	0.463	0.439	0.390	0.386
STMVnc-c				0.943	0.924	0.481	0.311	0.264	0.971	0.897	0.895	0.470	0.275	0.265
STMVSnc-c					0.979	0.465	0.350	0.305	0.920	0.951	0.949	0.455	0.316	0.307
STMVSKnc-c						0.451	0.341	0.313	0.902	0.959	0.959	0.441	0.318	0.311
MTMVnc-c							0.743	0.647	0.482	0.448	0.446	0.925	0.619	0.608
MTMVSnc-c								0.882	0.314	0.342	0.340	0.705	0.829	0.824
MTMVSKnc-c									0.267	0.308	0.308	0.614	0.882	0.890
STMVnc-nc										0.918	0.915	0.477	0.279	0.270
STMVSnc-nc											0.995	0.442	0.325	0.318
STMVSKnc-nc												0.440	0.325	0.318
MTMVnc-nc													0.646	0.634
MTMVSnc-nc														0.972

The main observations resulting from Table 4.6 can be summarised as follows. First, the ratings for these nonhomogeneous funds based on the traditional financial indicators present a consistently low correlation on average with all the metafrontier-based ratings (around 0.38-0.49). Second, when one moves to the comparison between metafrontier ratings, it is clear that the single-time metafrontier rating and multi-time metafrontier rating show a low correlation on average (around 0.32-0.48). Third, the MV metafrontier rating exhibits a low correlation on average with the multi-moment rating in the multi-time framework (around 0.64-0.75). In the single-time framework, the metafrontier ratings with different moments show a strong internal consistency: all the correlation coefficients within this framework are highly positive and significant. Finally, the metafrontier ratings based on the nc-c model show a high average correlation with those depending on the nc-nc model. This correlation is lower in the multi-time and multi-moment framework compared to the traditional single-time and MV framework.

Clearly, the fund selection process differentiates the 15 buy-and-hold backtesting strategies, since the best-performing funds are identified according to the above ratings. Table 4.7 presents an overall analysis with respect to the performances of the MVSK observations generated by per buy-and-hold strategy held until the end of the whole sample period (March 2021). This table is structured as follows: The first series of four columns list the results with regard to the 10 best funds selected for the backtesting exercise, and the second and third series of four columns present the results for selecting the 20 and the 30 best funds, respectively. Within each selecting (buying) scenario, the first two columns report the average inefficiency scores and the number of efficient units for each strategy when evaluated using the convex VRS frontier in the single-time and multi-moment framework (VRSc), while the last two columns report these results in the nonconvex case (VRSnc).

Table 4.7: Performance results for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 metafrontier rating methods): Descriptive statistics of the shortage function values

Methods	Fund(10)				Fund(20)				Fund(30)			
	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
Sharpe	0.0309	1	0.0184	12	0.0340	2	0.0258	10	0.0299	0	0.0212	8
Sortino	0.0246	5	0.0159	17	0.0353	0	0.0264	10	0.0302	3	0.0212	7
Omega	0.0314	1	0.0193	11	0.0341	1	0.0256	8	0.0286	2	0.0199	12
STMVnc-c	0.0715	4	0.0512	18	0.0719	0	0.0547	3	0.0543	0	0.0347	9
STMVSnc-c	0.0730	2	0.0528	11	0.0737	0	0.0533	2	0.0646	0	0.0482	1
STMVSKnc-c	0.0568	2	0.0406	6	0.0531	8	0.0384	14	0.0496	4	0.0352	9
MTMVnc-c	0.0691	1	0.0522	12	0.0503	4	0.0380	16	0.0432	7	0.0287	14
MTMVSnc-c	0.0831	1	0.0676	5	0.0614	2	0.0481	11	0.0461	5	0.0305	16
MTMVSKnc-c	0.0637	3	0.0497	9	0.0408	7	0.0303	15	0.0328	3	0.0203	15
STMVnc-nc	0.0659	2	0.0510	12	0.0643	2	0.0471	7	0.0549	0	0.0358	8
STMVSnc-nc	0.0901	4	0.0738	7	0.0542	1	0.0403	7	0.0448	2	0.0303	9
STMVSKnc-nc	0.0751	2	0.0570	11	0.0584	4	0.0463	9	0.0374	5	0.0271	14
MTMVnc-nc	0.0703	2	0.0529	12	0.0521	4	0.0359	15	0.0445	3	0.0286	14
MTMVSnc-nc	0.0589	6	0.0466	9	0.0478	2	0.0344	11	0.0360	1	0.0229	14
MTMVSKnc-nc	0.0524	5	0.0418	11	0.0376	4	0.0260	12	0.0346	1	0.0232	14

Analyzing Table 4.7 yields the following key conclusions. First, when only selecting the 10 best funds, the buy-and-hold strategies based on traditional financial indicators all outperform those based on the metafrontier ratings. From the first parallel block in Table 4.7, the average inefficiency scores of all strategies based on the financial ratings are lower than those of the metafrontier ratings. From

the second and third parallel block in Table 4.7, which report the performance results for these strategies when selecting 20 and 30 best funds. One can find that as the number of selected funds increases, the strategies based on the multi-time and multi-moment metafrontier ratings exhibit a comparable performance to those based on the financial ratings, whereas the strategies based on the other metafrontier ratings maintain a poor performance.

Second, the strategies driven by the multi-time metafrontier ratings present superior results compared to those driven by the single-time metafrontier ratings. It is clearly observed that the strategies in the multi-time framework in most cases exhibit lower average inefficiency scores and have a larger number of efficient units than those in the single-time framework. This result is valid when selecting the 10, 20 and 30 funds from the set of nonhomogeneous funds. Third, in the majority of cases, the metafrontier-based strategies with higher-order moments perform better than those with MV solely: This is valid for both single-time and multi-time frameworks. Again, this result is confirmed when buying the 10, 20 and 30 best funds.

The final observation relates to the comparison between the strategies based on the nc-c and nc-nc metafrontier ratings. Combining the two evaluation indicators of average inefficiency scores and the number of efficient units, it is found that with the consideration of multiple times and multiple moments, the buy-and-hold strategies consisting of the funds selected by the nc-nc rating models always outperform strategies consisting of the funds selected by the nc-c rating models. This result remains valid when buying the 10, 20 and 30 best funds.

To compare the 15 buy-and-hold strategies intuitively, Figure 4.4 provides boxplots to describe the entire distributions of the inefficiency scores per strategy held to the whole of the sample period. In each figure, the sub-figures (a) to (c) corres-

pond to the performance results of the buying scenarios with 10, 20 and 30 best funds selected, whereby the performance of strategies based on the convex VRS frontier are depicted in blue, and those based on the nonconvex VRS frontier are displayed in red. As introduced in subsection 4.4.1, the box of these box-plots indicates the interquartile range where the small vertical line reports the location of the median. Straightforwardly, the location of the median closer to the left indicates that the entire distribution of inefficiency scores for one strategy is somewhat skewed to the left, which signals that the strategy performs better in the backtesting analysis because the probability mass of the inefficiency is closer to zero. As we can observe from Figure 4.4, while the buy-and-hold strategies constructed by the multi-moment and multi-time metafrontier ratings do not outperform those constructed by the financial ratings, these strategies exhibit superiority compared to the strategies constructed by the single-time and multi-time MV metafrontier ratings.

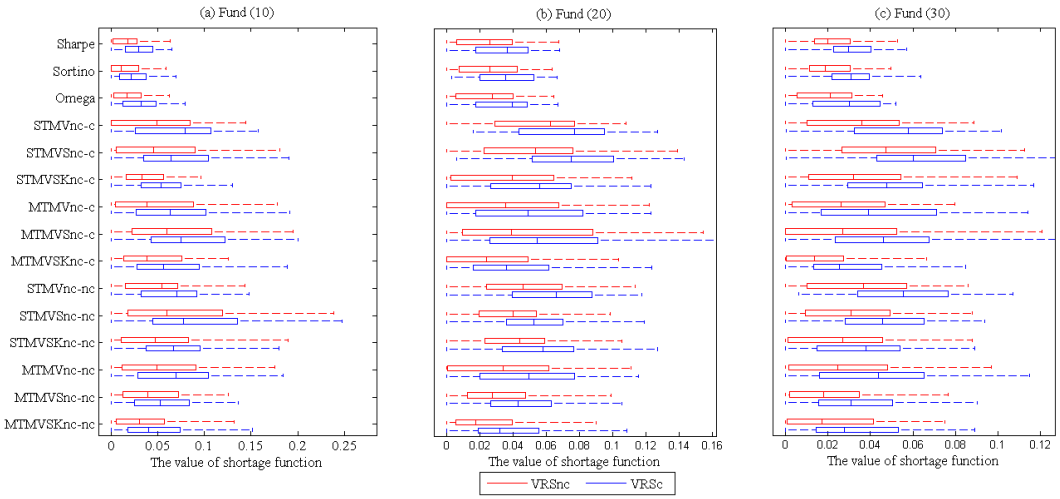

Figure 4.4: Inefficiency distributions for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 metafrontier rating methods)

As a sensitivity analysis, we calculate the performance of the above 15 buy-and-

holding strategies held for 1 year, 3 years and 5 years, respectively. This can be regarded as testing the short-term holding performance, the medium-term holding performance and the long-term holding performance of an investment strategy. Table 4.8 reports the summarized results with respect to the performance per buy-and-hold backtesting strategy held for 1, 3 and 5 years evaluated by the convex and nonconvex VRS frontiers, respectively. Equally so, Figures 4.5, 4.6 and 4.7 report the entire distributions of the inefficiency scores for the 15 strategies held for 1, 3 and 5 years, respectively. These results are presented in the Appendix. From the results of this sensitivity analysis, the above four findings are also evident in most cases for the three holding period scenarios. Moreover, the buy-and-hold backtesting strategies consisting of the best HFs rated by the multi-moment and multi-time performance measure tend to show consistent performance over different holding periods.

4.5 Conclusion

This contribution introduces the metafrontier approach from the production theory to the portfolio evaluation for assessing the efficiency of funds across groups with heterogeneity. Following recent developments on the metatechnology (see Afsharian and Podinovski (2018); Kerstens, O'Donnell, and Van de Woestyne (2019)), we estimate the nonparametric metafrontier in a nonconvex way for the fund rating, where the metafrontier are defined as the boundary of the union of all group-specific sets. To the best of our knowledge, this contribution is the first to systematically discuss the application of this metafrontier in the portfolio analysis and it proposes a series of nonparametric metafrontier-based methods for handling the appraisal of nonhomogeneous funds. We now summarize the main contributions in terms of both methodologies and empirical investigations.

In the methodology section, with the help of the shortage function, we establish the nonparametric metafrontier rating method that allows for evaluating the performance of funds across groups along a multitude of dimensions, and provide an MTE measure against the metafrontier for comparing the efficiency of these heterogeneous funds directly. This MTE measure is decomposed into GTE and TDG. The first component allows for gauging funds within a specific group, and the latter component measures the gap between the metafrontier and this group-specific frontier. Then, we extend this metafrontier rating procedure from the classic MV framework to the multi-time and multi-moment framework, which aims to be compatible with the general mixed risk-aversion preferences of investors regarding multiple times and multiple moments.

In the empirical section, the proposed rating methods are applied to a non-homogeneous set of actual funds covering the five distinct types of HF, BMF, EMF, MixMF and MonMF. In particular, our empirical investigation focuses on two key issues surrounding the application of the proposed methods for fund assessment and selection. First, we identify and discuss whether the specification factors (i.e., multiple times, multiple moments and the convexity of GTPPS) considered in the nonparametric metafrontier models have an impact on the MTE and its decomposition (GTE and TDG). By using the Li-test to compare the MTE (GTE and TDG) estimated by different metafrontier models, we find that the multiple moments and multiple times both separately and jointly have an impact on the estimates of MTE and its component for our data, and this impact is more significant when the two factors are considered jointly. In addition, the convexity assumption for each GTPPS is proven to have a clear impact on these three indices. Second, we design a simple buy-and-hold backtesting strategy to compare the performances of these proposed metafrontier ratings and traditional financial ratings for selecting best performing funds across groups. According to

the backtesting results, even though the buy-and-hold strategies constructed by the multi-moment and multi-time metafrontier ratings do not outperform those constructed by the financial ratings, these strategies exhibit superiority compared to the strategies determined by the single-time and multi-time MV metafrontier ratings.

To conclude, the proposed metafrontier methods provide an alternative procedure to evaluate funds across groups accounting for heterogeneity. This procedure is generalized and extended in the multi-time and multi-moment framework allowing for gauging heterogeneous fund performances in line with general investor preferences. An interesting perspective for future research is to extend the proposed nonparametric metafrontier rating methods with the Luenberger productivity indicator for estimating the performance changes of heterogeneous funds across groups over time. Since these indicators of different groups are evaluated relative to the same metafrontier, they are comparable and able to provide insightful information, or more specifically, whether performance change is driven by the efficiency change or the frontier change has different implications to fund managers and individual investors.

References

Afsharian, M., and H. Ahn (2015): "The Overall Malmquist Index: A New Approach for Measuring Productivity Changes over Time," *Annals of Operations Research*, 226(1), 1–27.

Afsharian, M., and V. Podinovski (2018): "A Linear Programming Approach to Efficiency Evaluation in Nonconvex Metatechnologies," *European Journal of Operational Research*, 268(1), 268–280.

Alexander, G. J., and R. D. Stover (1980): "Consistency of Mutual Fund Performance during Varying Market Conditions," *Journal of Economics and Business*, 32(3), 219–226.

Anderson, R., C. Brockman, G. Christos, and R. McLeod (2004): "A Non-parametric Examination of Real Estate Mutual Fund Efficiency," *International Journal of Business and Economics*, 3(3), 225–238.

Andreu, L., J. Sarto, and L. Vicente (2014): "Efficiency of the Strategic Style of Pension Funds: An Application of the Variants of the Slacks-based Measure in DEA," *Journal of the Operational Research Society*, 65(12), 1886–1895.

Basso, A., and S. Funari (2001): "A Data Envelopment Analysis Approach to Measure the Mutual Fund Performance," *European Journal of Operational Research*, 135(3), 477–492.

——— (2003): "Measuring the Performance of Ethical Mutual Funds: A DEA Approach," *Journal of the Operational Research Society*, 54(5), 521–531.

——— (2014): "Constant and Variable Returns to Scale DEA Models for Socially Responsible Investment Funds," *European Journal of Operational Research*, 235(3), 775–783.

Battese, G., and D. Rao (2002): "Technology Gap, Efficiency and a Stochastic Metafrontier Function," *International Journal of Business and Economics*, 1(2), 87–93.

Battese, G., D. Rao, and C. O'Donnell (2004): "A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating Under Different Technologies," *Journal of Productivity Analysis*, 21(1), 91–103.

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

Briec, W., C. Comes, and K. Kerstens (2006): "Temporal Technical and Profit Efficiency Measurement: Definitions, Duality and Aggregation Results," *International Journal of Production Economics*, 103(1), 48–63.

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

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

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

Brown, S. J., and W. N. Goetzmann (1997): "Mutual Fund Styles," *Journal of financial Economics*, 43(3), 373–399.

Casu, B., A. Ferrari, and T. Zhao (2013): "Regulatory Reform and Productivity Change in Indian Banking," *Review of Economics and Statistics*, 95(3), 1066–1077.

Chen, Z., and R. Lin (2006): "Mutual Fund Performance Evaluation using Data Envelopment Analysis with New Risk Measures," *OR Spectrum*, 28(3), 375–398.

Choi, Y., and B. Murthi (2001): "Relative Performance Evaluation of Mutual Funds: A Nonparametric Approach," *Journal of Business Finance and Accounting*, 28(7–8), 853–876.

Cooper, W. W., L. M. Seiford, and K. Tone (2007): *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer, Berlin.

Daraio, C., and L. Simar (2006): "A Robust Nonparametric Approach to Evaluate and Explain the Performance of Mutual Funds," *European Journal of Operational Research*, 175(1), 516–542.

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

Del Guercio, D., and P. A. Tkac (2002): "The Determinants of the Flow

of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds," *Journal of Financial and Quantitative Analysis*, 37(4), 523–557.

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

Dyson, R. G., R. Allen, A. S. Camanho, V. V. Podinovski, C. S. Sarrico, and E. A. Shale (2001): "Pitfalls and Protocols in DEA," *European Journal of operational research*, 132(2), 245–259.

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

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

Ewen, M. (2018): "Where is the Risk Reward? The Impact of Volatility-based Fund Classification on Performance," *Risks*, 6(3), 80.

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

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

Feng, C., J. Huang, and M. Wang (2018): "Analysis of green total-factor productivity in China's regional metal industry: A meta-frontier approach," *Resources Policy*, 58, 219–229.

Galagedera, D. U., I. Roshdi, H. Fukuyama, and J. Zhu (2018): "A New Network DEA Model for Mutual Fund Performance Appraisal: An Application to US Equity Mutual Funds," *Omega*, 77, 168–179.

Gerlach, P., and R. Maurer (2015): "Return-based Classification of Absolute Return Funds," *Journal of Asset Management*, 16(2), 117–130.

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

Gong, X., C. Yu, L. Min, and Z. Ge (2021): "Regret Theory-based Fuzzy Multi-objective Portfolio Selection Model Involving DEA Cross-efficiency and Higher Moments," *Applied Soft Computing*, 100, 106958.

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

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

Hayami, Y., and V. Ruttan (1970): "Agricultural Productivity Differences among Countries," *American Economic Review*, 60(5), 895–911.

Huang, C.-W., C.-T. Ting, C.-H. Lin, and C.-T. Lin (2013): "Measuring Non-Convex Metafrontier Efficiency in International Tourist Hotels," *Journal of the Operational Research Society*, 64(2), 250–259.

Huang, M.-Y., and T.-T. Fu (2013): "An Examination of the Cost Efficiency

of Banks in Taiwan and China using the Metafrontier Cost Function," *Journal of Productivity Analysis*, 40(3), 387–406.

Huang, M.-Y., J.-C. Juo, and T.-T. Fu (2015): "Metafrontier Cost Malmquist Productivity Index: An Application to Taiwanese and Chinese Commercial Banks," *Journal of Productivity Analysis*, 44(3), 321–335.

Jin, Q., K. Kerstens, and I. Van de Woestyne (2020): "Metafrontier Productivity Indices: Questioning the Common Convexification Strategy," *European Journal of Operational Research*, 283(2), 737–747.

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

Kerstens, K., C. O'Donnell, and I. Van de Woestyne (2019): "Metatechnology Frontier and Convexity: A Restatement," *European Journal of Operational Research*, 275(2), 780–792.

Kontolaimou, A., and K. Tsekouras (2010): "Are Cooperatives the Weakest Link in European Banking? A Non-parametric Metafrontier Approach," *Journal of Banking & Finance*, 34(8), 1946–1957.

Kounetas, K., I. Mourtos, and K. Tsekouras (2009): "Efficiency Decompositions for Heterogeneous Technologies," *European Journal of Operational Research*, 199(1), 209–218.

Krüger, J. (2021): "Nonparametric Portfolio Efficiency Measurement with Higher Moments," *Empirical Economics*, 61(3), 1435–1459.

Latruffe, L., J. Fogarasi, and Y. Desjeux (2012): "Efficiency, Productivity and Technology Comparison for Farms in Central and Western Europe: The Case of Field Crop and Dairy Farming in Hungary and France," *Economic Systems*, 36(2), 264–278.

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

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

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

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

Luenberger, D. G. (1995): *Microeconomic Theory*. Mcgraw-Hill College.

Makni, R., O. Benouda, and E. Delhoumi (2015): "Large Scale Analysis of Islamic Equity Funds using a Meta-Frontier Approach with Data Envelopment Analysis," *Research in International Business and Finance*, 34, 324–337.

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

Matallín-Sáez, J., A. Soler-Domínguez, and E. Tortosa-Ausina (2014): "On the Informativeness of Persistence for Evaluating Mutual Fund Performance using Partial Frontiers," *Omega*, 42(1), 47–64.

Mehlawat, M. K., A. Kumar, S. Yadav, and W. Chen (2018): "Data Envelopment Analysis based Fuzzy Multi-objective Portfolio Selection Model Involving Higher Moments," *Information sciences*, 460, 128–150.

Miszczyńska, D., and M. Miszczyński (2007): "Application of DEA Method to the Evaluation of the Efficiency of Polish Open Pension Funds in the Years 2004-2006," *Multiple Criteria Decision Making*.

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

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

Nalpas, N., L. Simar, and A. Vanhems (2017): "Portfolio Selection in a Multi-Moment Setting: A Simple Monte-Carlo-FDH Algorithm.," *European Journal of Operational Research*, 263(1), 308–320.

O'Donnell, C. (2016): "Using Information About Technologies, Markets and Firm Behaviour to Decompose a Proper Productivity Index," *Journal of Econometrics*, 190(2), 328–340.

O'Donnell, C., D. Rao, and G. Battese (2008): "Metafrontier Frameworks for the Study of Firm-Level Efficiencies and Technology Ratios," *Empirical Economics*, 34(1), 231–255.

Portela, M. C. A. S., E. Thanassoulis, A. Horncastle, and T. Maugg (2011): "Productivity change in the water industry in England and Wales: application of the meta-Malmquist index," *Journal of the Operational Research Society*, 62(12), 2173–2188.

Ray, S. C. (2004): *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. Cambridge university press.

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

Sala-Garrido, R., M. Molinos-Senante, and F. Hernández-Sancho (2011): "Comparing the Efficiency of Wastewater Treatment Technologies through a DEA Metafrontier Model," *Chemical Engineering Journal*, 173(3), 766–772.

Thieme, C., D. Prior, and E. Tortosa-Ausina (2013): "A Multilevel Decomposition of School Performance Using Robust Nonparametric Frontier Techniques," *Economics of Education Review*, 32, 104–121.

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

Tsekouras, K., N. Chatzistamoulou, and K. Kounetas (2017): "Productive Performance, Technology Heterogeneity and Hierarchies: Who to Compare with Whom," *International Journal of Production Economics*, 193, 465–478.

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

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

4.A Additional Table and Figures

Table 4.8: Performance results for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 metafrontier rating methods) held for 1, 3 and 5 years: Descriptive statistics of the shortage function values

		Hold 1 year				Hold 3 years				Hold 5 years			
	Methods	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
		Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
	Sharpe	0.0327	1	0.0198	12	0.0430	0	0.0198	10	0.0383	0	0.0253	4
	Sortino	0.0333	1	0.0203	10	0.0415	1	0.0204	11	0.0363	2	0.0225	13
	Omega	0.0333	5	0.0201	12	0.0441	0	0.0189	6	0.0393	0	0.0251	4
	STMVnc-c	0.0667	2	0.0535	6	0.0519	2	0.0257	17	0.0669	1	0.0444	14
	STMVSnc-c	0.0655	1	0.0526	3	0.0510	4	0.0252	13	0.0579	4	0.0401	14
	STMVSKnc-c	0.0551	2	0.0379	7	0.0514	4	0.0293	16	0.0599	1	0.0420	5
	MTMVnc-c	0.0622	2	0.0467	8	0.0502	1	0.0228	20	0.0606	4	0.0386	17
Fund(10)	MTMVSnc-c	0.0606	3	0.0409	11	0.0488	2	0.0238	19	0.0656	5	0.0458	17
	MTMVSKnc-c	0.0558	2	0.0365	11	0.0507	0	0.0233	18	0.0604	4	0.0391	15
	STMVnc-nc	0.0561	0	0.0415	11	0.0479	5	0.0253	20	0.0596	2	0.0409	12
	STMVSnc-nc	0.0631	2	0.0495	7	0.0616	0	0.0398	11	0.0656	1	0.0498	8
	STMVSKnc-nc	0.0630	0	0.0461	11	0.0585	1	0.0366	14	0.0616	2	0.0428	11
	MTMVnc-nc	0.0600	3	0.0446	8	0.0525	1	0.0244	16	0.0633	3	0.0416	16
	MTMVSnc-nc	0.0512	2	0.0355	13	0.0518	0	0.0310	16	0.0559	3	0.0373	14
	MTMVSKnc-nc	0.0447	2	0.0326	8	0.0452	4	0.0262	21	0.0497	2	0.0311	12
	Sharpe	0.0295	3	0.0177	15	0.0336	0	0.0153	13	0.0363	0	0.0263	6
	Sortino	0.0295	4	0.0202	16	0.0354	5	0.0153	17	0.0373	1	0.0278	9
	Omega	0.0310	0	0.0200	11	0.0342	1	0.0158	16	0.0376	0	0.0282	3
	STMVnc-c	0.0484	1	0.0343	10	0.0377	4	0.0199	19	0.0559	2	0.0450	8
	STMVSnc-c	0.0469	2	0.0298	11	0.0356	5	0.0211	21	0.0575	2	0.0418	10
	STMVSKnc-c	0.0459	3	0.0322	11	0.0373	2	0.0223	21	0.0526	3	0.0400	9
	MTMVnc-c	0.0502	0	0.0359	6	0.0395	4	0.0231	11	0.0490	1	0.0360	12
Fund(20)	MTMVSnc-c	0.0508	3	0.0367	8	0.0415	1	0.0254	14	0.0564	1	0.0408	10
	MTMVSKnc-c	0.0414	3	0.0260	14	0.0368	3	0.0193	24	0.0432	6	0.0314	11
	STMVnc-nc	0.0477	2	0.0358	8	0.0372	4	0.0190	22	0.0550	1	0.0415	13
	STMVSnc-nc	0.0431	2	0.0318	5	0.0377	2	0.0200	19	0.0464	0	0.0333	8
	STMVSKnc-nc	0.0425	2	0.0308	5	0.0356	6	0.0223	20	0.0441	4	0.0309	12
	MTMVnc-nc	0.0503	1	0.0374	8	0.0414	1	0.0246	9	0.0519	3	0.0359	14
	MTMVSnc-nc	0.0473	0	0.0359	4	0.0397	1	0.0202	18	0.0493	3	0.0386	6
	MTMVSKnc-nc	0.0417	3	0.0314	7	0.0361	0	0.0174	22	0.0413	3	0.0310	11
	Sharpe	0.0319	1	0.0225	6	0.0383	2	0.0176	13	0.0325	0	0.0237	11
	Sortino	0.0295	3	0.0201	11	0.0374	4	0.0194	14	0.0289	5	0.0205	9
	Omega	0.0321	1	0.0231	8	0.0366	2	0.0189	13	0.0337	2	0.0245	9
	STMVnc-c	0.0399	5	0.0309	10	0.0385	3	0.0191	19	0.0432	2	0.0327	7
	STMVSnc-c	0.0414	5	0.0276	10	0.0373	5	0.0213	20	0.0496	1	0.0380	8
	STMVSKnc-c	0.0384	4	0.0273	11	0.0355	6	0.0216	23	0.0430	3	0.0329	11
	MTMVnc-c	0.0449	2	0.0351	7	0.0414	3	0.0242	16	0.0417	4	0.0311	13
Fund(30)	MTMVSnc-c	0.0405	4	0.0310	8	0.0363	4	0.0205	18	0.0429	4	0.0303	19
	MTMVSKnc-c	0.0362	7	0.0283	17	0.0385	3	0.0201	20	0.0375	3	0.0278	10
	STMVnc-nc	0.0419	2	0.0312	13	0.0392	4	0.0218	11	0.0467	2	0.0351	11
	STMVSnc-nc	0.0366	1	0.0258	9	0.0372	1	0.0204	13	0.0365	1	0.0259	7
	STMVSKnc-nc	0.0326	3	0.0204	9	0.0312	7	0.0164	22	0.0335	4	0.0225	13
	MTMVnc-nc	0.0437	2	0.0331	8	0.0452	0	0.0275	11	0.0434	3	0.0302	17
	MTMVSnc-nc	0.0400	1	0.0306	10	0.0399	0	0.0201	16	0.0401	0	0.0299	8
	MTMVSKnc-nc	0.0384	1	0.0295	11	0.0385	1	0.0210	20	0.0383	1	0.0278	11

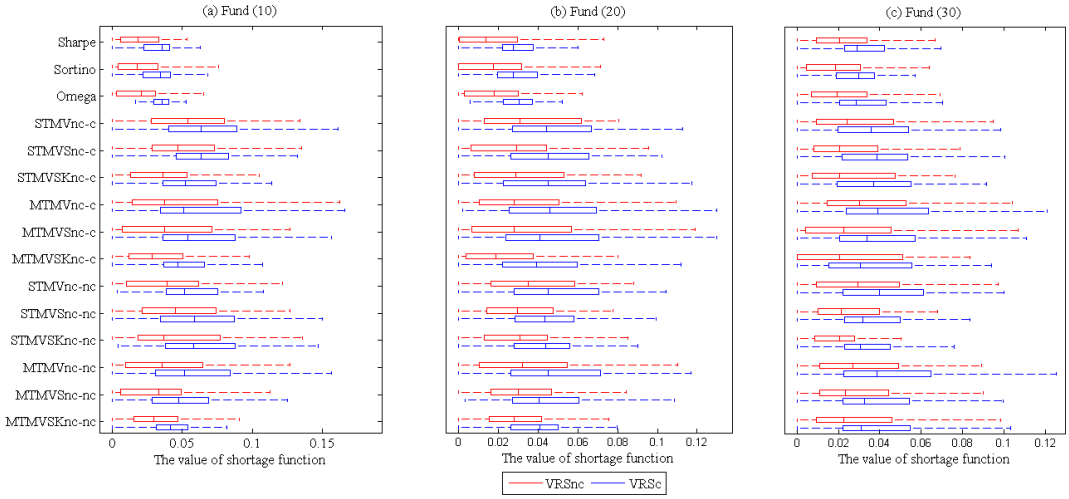

Figure 4.5: Inefficiency distributions for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 metafrontier rating methods): held for 1 year

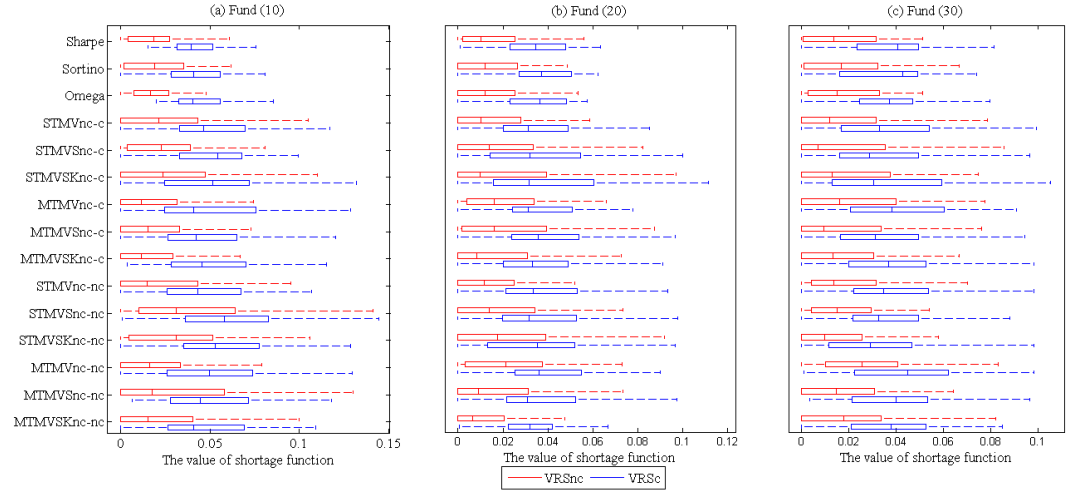

Figure 4.6: Inefficiency distributions for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 metafrontier rating methods): held for 3 years

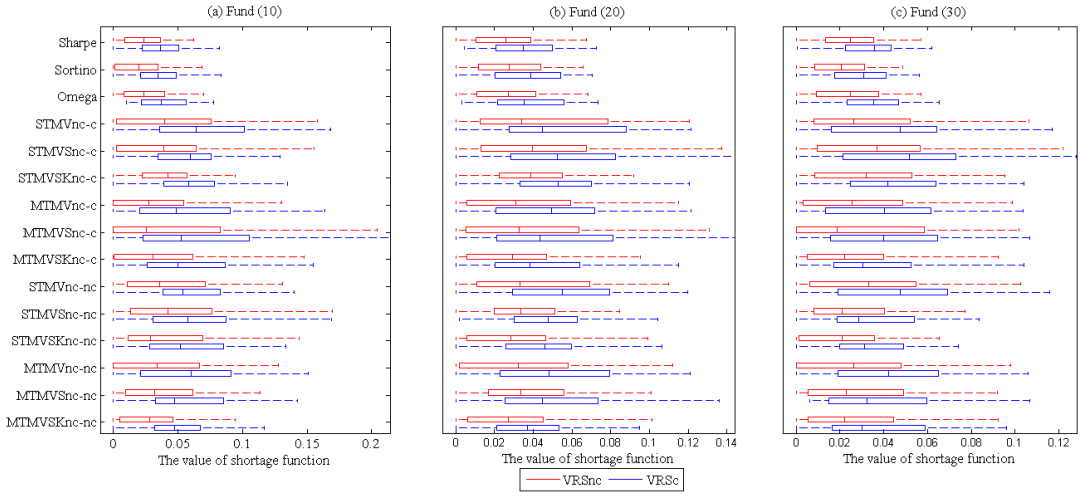

Figure 4.7: Inefficiency distributions for 15 buy-and-hold backtesting strategies (3 financial rating methods and 12 metafrontier rating methods): held for 5 years

CHAPTER 5 General Conclusion

This thesis has aimed to propose novel nonparametric frontier-based/metafrontier-based rating methods in the multi-time and multi-moment framework by integrating recent developments of nonparametric production frontier/metafrontier into the portfolio performance evaluation. Using the existing shortage functions, these proposed methods position each of the observed fund relative to the nonparametric frontier/metafrontier that can envelop the observations of any multi-dimensional choice set, and define a new performance measure that can simultaneously handle both multiple moments and multiple times. Furthermore, we have proposed a new Luenberger portfolio productivity indicator and its decomposition in the multi-time and multi-moment frontier to trace the performance change of funds over time. In the empirical analysis, we have explored the potential benefits of these newly proposed rating methods for selecting the best performing funds compared to the traditional finance measures and some existing frontier rating methods. We are now in a position to summarize the main contributions and discuss some future avenues for research.

Chapter 2 has proposed a series of nonparametric convex and nonconvex

frontier rating methods with multiple moments and multiple times, which are suitable to handle mixed risk-aversion preferences of investors which aim at time persistence. The proposed rating methodologies have been empirically applied to the HFs ratings, and the key findings can be summarized in three major elements. First, the multiple moments and multiple times both separately and jointly have an impact on the HF efficiency and ranking, and this impact is more significant when the two factors are considered jointly. Second, the buy-and-hold strategies based on the multi-time and multi-moment ratings exhibit a superiority over those based on traditional financial ratings and single-time MV ratings. Lastly, the nonconvex rating models have stronger discriminatory power with respect to the effect of multiple moments over the convex rating models. Importantly, the former display a more significant advantage in HF selection over the latter.

A general method for measuring the evolution of fund performance in the multi-time and multi-moment framework has been explored in **Chapter 3** by using some extensions of the shortage functions combined with the Luenberger productivity indicator. The proposed methodology has innovated in three main ways. First, the new Luenberger portfolio productivity indicators allow to trace the evolution of fund performance with multiple times and multiple moments. Second, these portfolio productivity indicators based on the nonparametric frontier methods are all simply computed by the linear (or binary mixed integer) programming, which are much easier and time-saving when applied in the large-scale and multi-dimensional evaluation. Third, the decomposition of these indicators distinguishes whether changes in fund performance over time are due to the efficiency change or the frontier shift. The efficiency change component allows testing in an alternative, but conceptually promising way the eventual ability of fund managers, since this measurement is not contaminated by any changes in the financial market itself (as measured by the frontier change component). With the same HF data, we have

tested the potential of the proposed indicators and their components in HF selection. The backtesting results point to a remarkable superior performance of the multi-time and multi-moment Luenberger-based ratings compared to traditional finance-based ratings and the single-time MV Luenberger-based ratings for the actual HF selection.

The nonparametric metafrontier-based rating methods have been proposed in **Chapter 4** where the homogeneity assumption of funds required in the standard frontier methods is relaxed. This metafrontier method allows for evaluating the performance of funds across groups along a multitude of dimensions, and define a metatechnology efficiency (MTE) measure against the metafrontier for gauging these heterogeneous funds directly. Without limiting ourselves to the traditional mean-variance portfolio setting, we have developed the nonparametric metafrontier rating methods in the multi-time and multi-moment framework, aiming to allow for the fund evaluation compatible with the general investor preferences. The empirical part employs a large database of heterogeneous funds covering five different groups (types) not only to offer extensive tests of the specification factors considered in these metafrontier-based models, but also to test the performance of different rating methods in a simple backtesting setup. Specifically, the multiple moments and multiple times both separately and jointly have an impact on the estimate of MTE, and imposing the convexity assumption for each GTPPS is proven to have a clear impact on the MTE in any rating framework. The backtesting analysis indicates that even though the buy-and-hold strategies constructed by the multi-moment and multi-time metafrontier ratings do not outperform the traditional financial ratings, these strategies exhibit superiority compared to the other strategies from nonparametric metafrontier rating family.

Overall, the proposed methodologies and the empirical findings in this thesis

provide several novel ideas into the important finance topic of rating and selecting funds. Future work can expand on this work in multiple directions. First, one could further account for other commonly used risk measures that allow for assessing the tail risk as an alternative to variance (or the second moment of the return distribution), such as Value-at-Risk (VaR) or the Conditional Value-at-Risk (CVaR) (see Alexander and Baptista (2002), Lwin, Qu, and MacCarthy (2017), Guo, Chan, Wong, and Zhu (2019)). The fundamental methodologies regarding the portfolio efficiency and the portfolio productivity indicator in our work could be extended to the mean-VaR (or CVaR)-skewness-kurtosis framework. Second, the portfolio evaluation by using the shortage functions or distance functions relies on the selection of the direction vectors along which the distance from the efficient frontier is measured. It could be interesting to test for the impact of different directions determined by different direction-selecting techniques (see Wang, Xian, Lee, Wei, and Huang (2019) for a review) on the portfolio efficiency and the portfolio productivity indicator. Third, our backtesting analysis is limited to a simple buy-and-hold backtesting setting. Future work could consider several alternative backtesting strategies to test the potential benefits of the newly proposed ratings in funds selection. Lastly, it is worthwhile looking for further developments in investigating the impact of unforeseen events (e.g., the 2008 financial crisis and the Covid 19 pandemic) on fund ratings and selection, as well as in testing the performance of the multi-time and multi-moment frontier rating models in periods of high market volatility.

References

Alexander, G. J., and A. M. Baptista (2002): "Economic Implications of Using a Mean-VaR Model for Portfolio Selection: A Comparison with Mean-Variance Analysis," *Journal of Economic Dynamics and Control*, 26(7-8), 1159–1193.

Guo, X., R. H. Chan, W. K. Wong, and L. Zhu (2019): "Mean-Variance, Mean-VaR, and Mean-CVaR Models for Portfolio Selection with Background Risk," *Risk Management*, 21(2), 73–98.

Lwin, K. T., R. Qu, and B. L. MacCarthy (2017): "Mean-VaR Portfolio Optimization: A Nonparametric Approach," *European Journal of Operational Research*, 260(2), 751–766.

Wang, K., Y. Xian, C. Lee, Y. Wei, and Z. Huang (2019): "On Selecting Directions for Directional Distance Functions in a Non-parametric Framework: A Review," *Annals of Operations Research*, 278(1), 43–76.