Université catholique de Louvain

Louvain Institute of Data Analysis and Modeling in economics and statistics - Louvain Finance

Doctoral Thesis

Optimal Portfolio Selection under Parameter Uncertainty

Rodolphe Vanderveken

Thesis submitted in partial fulfillment of the requirements of the degree of *Docteur en sciences économiques et de gestion*

Jury composition:
Prof. Frédéric Vrins (UCLouvain, BE), Co-Supervisor.
Prof. Nathan Lassance (UCLouvain, BE). Co-Supervisor.
Prof. Majeed Simaan (Stevens Institute of Technology, USA)
Prof. Guofu Zhou (Washington University, USA)
Prof. Marco Saerens (UCLouvain, BE), President.

Louvain-la-Neuve, June 2025

To my family, Lilian, Annick and Grégoire

'The more they study, the more stupid they become'

My late great-aunt

Table of Contents

Acknowledgments

According to a well-known saying, *there is always someone stupid in a meeting, and if you don't know who it is, then it's you.* Well, in addition to the many contributions to portfolio theory made in this thesis, I'd like to offer one more – arguably even more groundbreaking: this saying is wrong. And I have a solid counterexample to prove it: the countless hours of meetings with Frédéric, Nathan, and myself. I always knew exactly who the "stupid one" was, and it certainly wasn't either of them.

Frédéric, Nathan, working with you both has been an incredible experience. A bit disorienting at first, coming from the private sector. And yes, it was intense at times, your standards are undeniably high, but it was always deeply rewarding. I have been constantly challenged and I have learned so much under your joint mentorship. I'm proud of what we've accomplished together, and of the quality of this thesis. You both belong to a rare and elite species: PhD supervisors who are not only deeply knowledgeable and supportive, but also open to ideas, genuinely kind, fun to be around, and perhaps most importantly, appreciative of a good old Orval when the occasion calls for it.

Nathan, please don't forget about me when you hit your 40,000th citation and begin levitating *à la* Harry Markowitz. Jokes aside, you are truly a research powerhouse, and it has been a privilege to work alongside you. You've consistently been patient, supportive, deeply involved, and incredibly accessible: always up for a "quick" Teams call (which, let's be honest, rarely stayed quick), always willing to share some code or dive into the weeds of some obscure and not-so-thrilling implementation detail, even late in the evening. And yes, I know . . . I should have just used MATLAB.

Frédéric, I honestly don't know how you do it. By "it", I mean: publishing in top journals, actively mentoring PhD students (including me), writing a great textbook on derivatives pricing, leading the LFIN research center as president, handling your teaching duties with your legendary enthusiasm, taking on external consulting activities, and still somehow making time for your family. Truly, you seem to do it all. Well... almost all. Let's be honest: you're not exactly a force to be reckoned with on the padel court. So hey, at least I've got that going for me!

I remember saying before I even began my PhD that the topic didn't matter much, as long as I could work with both of you. Looking back, I couldn't have been more right. I consider myself fortunate to have had the opportunity to work alongside you. Thank you.

I wish to warmly thank the members of my jury. Prof. Marco Saerens (UCLouvain), thank you not only for presiding over the jury, but also for your thoughtful and insightful feedback during and after my confirmation step. Although portfolio selection is not your primary area of research, your questions and remarks were remarkably relevant. Prof. Majeed Simaan (Stevens Institute of Technology), I am sincerely grateful that you agreed to join my jury. Your insightful comments during my confirmation step, as well as during your visit to Louvain-la-Neuve, were greatly appreciated. You are a recognized expert in finance and portfolio selection, yet still particularly accessible and one of the nicest scholars I have met. Prof. Guofu Zhou (Washington University), thank you for accepting the invitation to be part of my jury. Our work in this thesis is heavily based on your foundational contributions, in particular your 2007 (JFQA, with R. Kan) and 2011 (JFE, with J. Tu) papers. Therefore, it goes without saying that it is a great honor for me to have you participating in my jury.

I was proud to be affiliated with the LIDAM institute and to be a fellow of the Louvain Finance (LFIN) research center. Being part of such an intellectually stimulating environment was truly motivating. A special shout-out to Arnaud – it's been a genuine pleasure to have you by my side across my PhD. I'll miss our discussions about politics, research, the latest achievements of our beloved students, and, of course, the occasional dissection of LinkedIn wisdom.

This thesis would not have been possible without the financial support I received. I am deeply grateful to the Fonds de la Recherche Scientifique (F.R.S.-FNRS) for the grant which enabled me to conduct my research under great conditions.

To my friends: thank you for being there through it all. Whether it was grabbing a drink after a long day, listening to me vent about research stuff you never asked to hear about, pretending to understand what I was working on, or just helping me disconnect when I needed it, you made this whole ride a lot more bearable (and way more fun). Your support, humor, and friendship reminded me that there is more to life than Python errors and conference deadlines. I'm really lucky to have you around.

My family has always been the most important part of my life and a constant source of inspiration. Some might tease me by suggesting that I pursued a PhD just to follow in my older brother's footsteps. And while I'd love to deny it, there's more truth to that than I'd like to acknowledge. To my parents, thank you from the bottom of my heart for your unwavering support throughout this journey – especially in the final months. Finishing this thesis would simply not have been possible without you. I dedicate this work to the three of you.

Abstract

Portfolio selection lies at the heart of financial decision-making for a wide range of economic agents. Building upon the classical mean-variance framework of Markowitz (1952), this thesis addresses the challenge of *optimal portfolio selection under parameter uncertainty*, i.e., when the required inputs – expected returns and covariances – are unknown and must be estimated from historical data. This estimation process inevitably introduces estimation errors which, if not accounted for properly, lead to poor out-of-sample performance.

To address this issue, we propose three different yet complementary approaches to deal with parameter uncertainty in the portfolio selection context, all based on the expected out-of-sample utility (EU) criterion introduced by Kan and Zhou (2007).

In Chapter 2, we study how to best combine the sample mean-variance portfolio with the naive equally weighted portfolio to optimize out-of-sample performance. We show that the seemingly natural convexity constraint that Tu and Zhou (2011) impose – the two combination coefficients must sum to one – is undesirable because it severely constrains the allocation to the risk-free asset relative to the unconstrained portfolio combination. However, we demonstrate that relaxing the convexity constraint inflates estimation errors in combination coefficients, which we alleviate using a shrinkage estimator of the unconstrained combination scheme. Empirically, the constrained combination outperforms the unconstrained one in a range of generally small degrees of risk aversion, but severely deteriorates otherwise. In contrast, the shrinkage unconstrained combination enjoys the best of both strategies and performs consistently well.

In Chapter 3, we introduce a method to determine the investor's optimal portfolio size that maximizes the expected out-of-sample utility. This portfolio size trades off between accessing investment opportunities and limiting the number of parameters to be estimated. Unlike sparse methods such as lasso that exclude assets during the optimization step, our approach fixes the optimal number of assets *before* computing the portfolio weights, which improves robustness and provides greater flexibility in practical implementations. Empirically, our restricted portfolios outperform their counterparts applied to all available assets. Our methodology renders portfolio theory valuable even when the dataset dimension and sample size are comparable.

In Chapter 4, we introduce analytical linear and nonlinear shrinkage estimators of the sample covariance matrix that are optimal for mean-variance portfolio choice. Unlike

classical methods based on statistical loss functions like the mean squared error, our shrinkage covariance matrices optimize the expected out-of-sample portfolio utility and account for estimation errors stemming from the mean vector. Our estimators shrink the sample eigenvalues more intensively than in conventional methods, and they especially shrink the contribution of principal components with small squared Sharpe ratios. Overall, our methodology estimates the covariance matrix and the optimal mean-variance portfolio jointly, which we show delivers significant empirical gains relative to the conventional two-step approach and recently proposed regularized mean-variance portfolios.

Research output

The work featured in this thesis includes: (a) publications in scientific journals, (b) working papers, and (c) presentations in international scientific conferences.

(a) *Publications*:

- Lassance, N., Vanderveken, R. and Vrins, F. (2024b). On the combination of naive and mean-variance portfolio strategies. *Journal of Business & Economic Statistics*, *42*(3), pp. 875–889

(b) *Working Papers*:

- Lassance, N., Vanderveken, R. and Vrins, F. (2025a). Optimal portfolio size under parameter uncertainty. *SSRN working paper*.
  This paper received the 2024 Best Academic Paper Award from the French Association of Institutional Investors (AF2I).

- Lassance, N., Vanderveken, R. and Vrins, F. (2025b). Optimal shrinkage of the covariance matrix for portfolio selection. *Work in progress*.

(c) *Conference presentations*:

- The **12$^{\text{th}}$ General AMaMeF (Advanced Mathematical Methods for Finance) conference**, event organised by the *University of Verona* (Verona, Italy, June 23-27 2025).

- The **14$^{\text{th}}$ International conference of the Financial Engineering and Banking Society (FEBS)**, event organised by *MBS, University of Surrey and University of Southampton* (Montpellier, France, June 12-13 2025).

- The **17$^{\text{th}}$ annual SoFiE (Society for Financial Econometrics) conference**, event organised by *ESSEC Business School* (Paris, France, June 9-12 2025).

- The **7$^{\text{th}}$ Quantitative Finance and Financial Econometrics (QFFE) international conference**, event organised by the *Aix-Marseille Université School of Economics* (Marseille, France, June 5-6 2025).

- The **41th International Conference of the French Finance Association (AFFI)**, event organised by the *IAE Dijon* (Dijon, France, May 26-28 2025).
- The **International Finance and Banking Society (IFABS) Oxford Conference**, event co-organised by the *Bank of England, European Central Bank and Oxford University* (Oxford, UK, April 15-17, 2025).
- The **17th German Probability and Statistics Days (GPSD) 2025**, event organised by *TU Dresden* (Dresden, Germany, March 11-14 2025).
- The **16th Actuarial and Financial Mathematics Conference (AFMath)**, event co-organised by all Belgian universities and the *FNRS, FWO* (Brussels, Belgium, February 3-4, 2025).
- The **17th International Risk Management Conference (IRMC)**, event organised by *Bocconi University* and the *New York University Stern School of Business* (Milan, Italy, June 24-25 2024).
- The **6th Quantitative Finance and Financial Econometrics (QFFE) international conference**, event organised by the *Aix-Marseille Université School of Economics* (Marseille, France, June 6-7 2024).
- The **40th International Conference of the French Finance Association (AFFI)**, event organised by the *IAE Lille University School of Management* (Lille, France, May 27-29 2024).
- The **18th meeting of the Belgian Financial Research Forum (BFRF)**, event co-organised by all Belgian universities and the *National Bank of Belgium* (Brussels, Belgium, May 13, 2024)
- The **10th International Conference on Computational Finance (ICCF)**, event organised by *Utrecht University* (Amsterdam, The Netherlands, April 2-5 2024).
- The **17th meeting of the Belgian Financial Research Forum (BFRF)**, event co-organised by all Belgian universities and the *National Bank of Belgium* (Brussels, Belgium, April 20-21, 2023).
- The **14th Actuarial and Financial Mathematics Conference (AFMath)**, event co-organised by all Belgian universities and the *FNRS, FWO* (Brussels, Belgium, February 9-10, 2023).
- The **16th International Conference on Computational and Financial Econometrics (CFE)**, event organised by the *King's College London* (London, UK, December 17-19, 2022).

- The **International Finance and Banking Society (IFABS) Naples Conference**, event co-organised by the *Bank of England, European Central Bank, University of Leicester, and University of Naples Federico II* (Naples, Italy, September 7-9, 2022).

List of Figures

List of Tables

List of Acronyms

CRSP	Center for Research in Security Prices
ESR	Expected out-of-sample Sharpe Ratio
EU	Expected out-of-sample Utility
EW	Equally Weighted
GMV	Global Minimum Variance
i.i.d.	Independent and Identically Distributed
LASSO	Least Absolute Shrinkage and Selection Operator
MCS	Model Confidence Set
MPT	Modern Portfolio Theory
MSE	Mean Squared Error
MV	Mean-Variance
OLS	Ordinary Least Squares
PC	Principal Component
PCA	Principal Component Analysis
RF	Risk-Free
RMSE	Root Mean Squared Error
SDF	Stochastic Discount Factor
SGMV	Sample Global Minimum Variance
SMV	Sample Mean-Variance

Chapter 1

Introduction

How should investors allocate capital across different investment opportunities? This question is not only at the core of academic finance but also a practical concern for a wide range of economic agents. From individual savers planning for retirement to institutional investors managing vast pools of capital, and even to those seemingly passive participants – through pension plans, mutual funds, or exchange-traded funds – the challenge of allocating capital, i.e., performing *portfolio selection*, is universal.

At its heart, the objective of portfolio selection is to allocate capital across available assets in a way that balances return and risk in a rational and informed manner. This objective is as central today as it was when Harry Markowitz introduced his groundbreaking *mean-variance* framework in the 1950s (Markowitz, 1952, 1959), which was the first to provide a mathematical foundation for thinking about this risk and return trade-off. His insights laid the foundation for what became known as Modern Portfolio Theory (MPT), and later awarded him the Nobel Prize in Economics in 1990.[1] The mean-variance framework is still widely used nowadays, both in academic research and professional practice, and is valued for its clarity, simplicity, and flexibility.

The mean-variance framework is normative, as it prescribes how investors should allocate their capital to optimize the risk-return trade-off. It identifies a set of *mean-variance portfolios* forming together the *efficient frontier*: Those that offer the lowest risk for a given expected return, or the highest expected return for a given level of risk, where risk is measured as the variance of portfolio returns. Any portfolio not lying on this frontier is suboptimal, as there exists at least one alternative portfolio that either delivers a higher return for the same level of risk, or achieves the same return with less risk. The specific portfolio an investor should then choose from the efficient frontier will depend on the investor's individual degree of risk aversion.

The mean-variance framework was groundbreaking in its treatment of both *uncertainty*

[1] In 1989, Harry Markowitz was also awarded the *John von Neumann Theory Prize*, recognizing not only his foundational work in portfolio theory but also his significant contributions to sparse matrix techniques and the development of simulation programming languages.

and *diversification*. Uncertainty is an inherent aspect of portfolio selection, reflecting the fact that investors cannot know future asset returns with certainty when constructing a portfolio. If future returns were known, the optimal portfolio would be trivial: The investor would allocate all capital to the asset with the highest return, and diversification would never be a rational choice. Yet, diversification has always been widely practiced and recognized as a sound investment practice. The mean-variance framework provides a theoretical foundation for this intuition by incorporating uncertainty into the portfolio selection process. Specifically, future asset returns are considered random variables and follow a multivariate probability distribution which depends on key parameters such as the vector of expected asset returns and their covariance matrix. The covariance matrix captures not only the individual risk (variance) of each asset but also the correlation structure among them. In this setting, diversification becomes the rational choice: By combining assets that are not perfectly correlated, investors can reduce overall portfolio variance while maintaining expected returns.

However, the very uncertainty that rationalizes diversification also introduces significant challenges when applying the mean-variance framework in practice. Indeed, the key parameters of the distribution of the future asset returns – the expected returns and the covariance matrix – are themselves unobservable and must be estimated using a sample of historical data. This estimation inevitably introduces *estimation errors*, in the sense that the estimated parameters differ from their true, but unobservable, values. These estimation errors can significantly distort portfolio weights and, as a result, degrade portfolio performance. Thus, while the mean-variance framework embraces uncertainty to rationalize diversification, it is also inherently exposed to a second layer of uncertainty – *parameter uncertainty* – which complicates its practical implementation.

Of course, the investor can choose to completely ignore parameter uncertainty, and thus to rely on the estimates as if they were the true parameters. This is known as the *plug-in approach*, which yields for instance the *sample mean-variance* portfolio when the corresponding estimates are the sample estimates. However, substantial evidence has shown over the years that that this approach yields very poor performance in practice and that investors are then often better off not investing at all. In particular, "naive" strategies, i.e., theoretically sub-optimal from a mean-variance perspective, but with little or no exposure to parameter uncertainty, are typically more efficient in practice.

The central question then becomes: Is it possible to account properly for parameter uncertainty, and thus, to perform *optimal portfolio selection under parameter uncertainty*?

Given the nature of the problem, it is difficult to identify *a priori* a single, universal method that consistently yields the best portfolio performance under parameter uncertainty. As a result, the scientific literature has expanded considerably over the years, proposing a wide range of approaches. These include, but are not limited to: robust and shrinkage estimators of key parameters (Stein, 1956; James and Stein, 1961; Ledoit and Wolf, 2004,

2012; Bodnar et al., 2014), constraints in the portfolio selection problem (Frost and Savarino, 1988; Jagannathan and Ma, 2003; Behr et al., 2013), norm-based regularization (DeMiguel et al., 2009a), sparse portfolio strategies (Gao and Li, 2013; Ao et al., 2019; Kremer et al., 2020), Bayesian methods (Jorion, 1986; Black and Litterman, 1992), and portfolio combinations (Kan and Zhou, 2007; Tu and Zhou, 2011; Kan et al., 2021).

In this thesis, we follow the framework introduced by Kan and Zhou (2007), who propose portfolio strategies aimed at maximizing a specific objective: the mean-variance *expected out-of-sample utility* (EU). Intuitively, the EU is a measure of portfolio *out-of-sample performance*, i.e., of how well a portfolio performs outside of the sample of data on which it was estimated. This is a meaningful objective, as investors should be concerned with how their portfolio will perform in the future (out-of-sample performance), rather than how it performed on past data (in-sample performance). Specifically, the EU measures the average out-of-sample performance of a portfolio strategy estimated on a sample of data, averaged over all possible sample realizations. This approach accounts for parameter uncertainty by penalizing the EU when the parameter estimates that are used may be far off the true values. A portfolio that maximizes the EU is thus *optimal under parameter uncertainty* in the sense that there is no other portfolio of that form delivering a higher expected out-of-sample performance.

We use this framework for several reasons. First, it is particularly effective at producing analytical insights on the impact of parameter uncertainty on portfolio selection, and is thus also computationally efficient. Second, it delivers overall strong and robust performance in empirical settings. Notably, the portfolio strategies introduced by Kan and Zhou (2007) have become over the years standard benchmarks in the literature. Third, it is highly flexible, as it applies to any portfolio based on on sample estimates and extends beyond portfolio weights computation. For instance, we also leverage it to calibrate other quantities such as portfolio sizes and shrinkage intensities, as developed in Chapters 3 and 4.

In this thesis, we introduce three different yet complementary ways to perform optimal portfolio selection under parameter uncertainty. This document is thus organized as a collection of three research papers, one per chapter. While each method is different, they all share a common foundation in the mean-variance framework, and the same objective: maximizing the EU. We now briefly present each of them.

In Chapter 2, we study how to best combine the sample mean-variance portfolio with the naive equally weighted portfolio to optimize out-of-sample performance. We show that the seemingly natural convexity constraint that Tu and Zhou (2011) impose – the two combination coefficients must sum to one – is undesirable because it severely constrains the allocation to the risk-free asset relative to the unconstrained portfolio combination. However, we demonstrate that relaxing the convexity constraint inflates estimation errors in combination coefficients, which we alleviate using a shrinkage estimator of the unconstrained combination scheme. Empirically, the constrained combination outperforms

the unconstrained one in a range of generally small degrees of risk aversion, but severely deteriorates otherwise. In contrast, the shrinkage unconstrained combination enjoys the best of both strategies and performs consistently well.

In Chapter 3, we introduce a method to determine the investor's optimal portfolio size that maximizes the expected out-of-sample utility. This portfolio size trades off between accessing investment opportunities and limiting the number of parameters to be estimated. Unlike sparse methods such as lasso that exclude assets during the optimization step, our approach fixes the optimal number of assets *before* computing the portfolio weights, which improves robustness and provides greater flexibility in practical implementations. Empirically, our restricted portfolios outperform their counterparts applied to all available assets. Our methodology renders portfolio theory valuable even when the dataset dimension and sample size are comparable.

In Chapter 4, we introduce analytical linear and nonlinear shrinkage estimators of the sample covariance matrix that are optimal for mean-variance portfolio choice. Unlike classical methods based on statistical loss functions like the mean squared error, our shrinkage covariance matrices optimize the expected out-of-sample portfolio utility and account for estimation errors stemming from the mean vector. Our estimators shrink the sample eigenvalues more intensively than in conventional methods, and they especially shrink the contribution of principal components with small squared Sharpe ratios. Overall, our methodology estimates the covariance matrix and the optimal mean-variance portfolio jointly, which we show delivers significant empirical gains relative to the conventional two-step approach and recently proposed regularized mean-variance portfolios.

In each chapter, a significant effort is dedicated to validating the robustness of our results, ensuring that our portfolio strategies are effective across various settings and applicable in real-world scenarios. To achieve this, we rely on extensive simulation exercises to verify the reliability of our theoretical findings. Additionally, each chapter includes an comprehensive empirical analysis based on real-world datasets that are standard in the literature and that span a range of assets including characteristic, industry and anomaly portfolios, as well as individual stocks. The performance of our methods is consistently compared to state-of-the-art benchmarks, transaction costs are always accounted for, and various statistical tests are implemented to confirm the significance of our results.

Chapter 5 concludes the thesis and offers some avenues for future research based on this work. To improve readability, each chapter begins and ends with its own abstract and conclusion, and relevant supplementary material is provided at the end of each chapter.

Chapter 2

On the Combination of Naive and Mean-Variance Portfolio Strategies

Abstract

We study how to best combine the sample mean-variance portfolio with the naive equally weighted portfolio to optimize out-of-sample performance. We show that the seemingly natural convexity constraint that Tu and Zhou (2011) impose—the two combination coefficients must sum to one—is undesirable because it severely constrains the allocation to the risk-free asset relative to the unconstrained portfolio combination. However, we demonstrate that relaxing the convexity constraint inflates estimation errors in combination coefficients, which we alleviate using a shrinkage estimator of the unconstrained combination scheme. Empirically, the constrained combination outperforms the unconstrained one in a range of generally small degrees of risk aversion, but severely deteriorates otherwise. In contrast, the shrinkage unconstrained combination enjoys the best of both strategies and performs consistently well for all levels of risk aversion.

Reference

This chapter is based on:

Lassance, N., Vanderveken, R. and Vrins, F. (2024b). On the combination of naive and mean-variance portfolio strategies. *Journal of Business & Economic Statistics*, *42*(3), pp. 875–889.

2.1 Introduction

The sample mean-variance (SMV) portfolio of Markowitz (1952) performs notoriously poorly out of sample because it is highly sensitive to estimation errors in the mean and covariance matrix of asset returns (Barroso and Saxena, 2021). One solution to alleviate this problem is to *combine* the SMV portfolio with another portfolio less sensitive to estimation risk.

In this chapter, we study how to best combine the SMV portfolio with the naive equally weighted (EW) portfolio so as to optimize out-of-sample performance. Specifically, we consider the portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) = \kappa_1 \hat{\boldsymbol{w}}^\star + \kappa_2 \boldsymbol{w}_{ew}$, where $\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ is the SMV portfolio obtained from a sample of size T, $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$ is the EW portfolio, and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$ are the combination coefficients. As in Kan and Zhou (2007), we calibrate $\boldsymbol{\kappa}$ to optimize the performance, measured by the expected out-of-sample utility, $\mathbb{E}\big[\hat{\boldsymbol{w}}'(\boldsymbol{\kappa})\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}'(\boldsymbol{\kappa})\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\big]$. Several combinations with naive strategies can be used to improve the performance of the SMV portfolio, the most famous ones being the EW and the sample global-minimum-variance (SGMV) portfolios. Although SGMV is used in many studies such as Kan and Zhou (2007), DeMiguel et al. (2015), and Kan et al. (2021), we focus on EW in this chapter for several reasons.[1] First, DeMiguel et al. (2009b) show that, albeit naive, the EW portfolio often compares favourably to mean-variance models. Second, unlike the SGMV portfolio that depends on the inverse covariance matrix, the EW portfolio is not subject to any estimation error. Thus, our results indicate that when the number of assets is sizable, combining with EW delivers superior performance and lower turnover than combining with SGMV. Third, we demonstrate theoretically and empirically that the SMV and EW portfolios have a smaller out-of-sample return correlation than SMV and SGMV, and thus deliver larger diversification gains when combined together. Finally, in addition to these performance-related arguments, when combining the SMV and EW portfolios the relevant literature is the paper by Tu and Zhou (2011). However, we show that Tu and Zhou (2011) combine the SMV and EW portfolios in a suboptimal way, which calls for further analysis of this particular combination.

We contribute to the portfolio combination literature in four ways. First, we analyze the impact of the *convexity constraint* that Tu and Zhou (2011) impose on combination coefficients, i.e., $\kappa_1 + \kappa_2 = 1$, which gives rise to the *constrained strategy*. This constraint suffices for the constrained strategy to outperform the SMV and EW portfolios, due to the concavity of expected utility. However, we show that the constrained strategy *underperforms* the third underlying fund—the risk-free asset—once the risk-aversion coefficient γ exceeds a given threshold. This can be explained because the convexity constraint prevents the constrained strategy to invest in the risk-free asset directly; it can only do so via the SMV

[1] Although we focus on combining the SMV and EW portfolios, in Section 2.C.3 of the Supplementary Material we also derive counterparts of our main results for the combination of the SMV and SGMV portfolios.

portfolio.[2] In this case, a higher exposure to the risk-free asset comes at the cost of a higher exposure to the tangent portfolio, which is not desirable for highly risk-averse investors.

To address this issue, we propose an *unconstrained strategy* that combines the SMV and EW portfolios without imposing the convexity constraint. We theoretically compare the properties of the portfolios $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^c)$ and $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)$, which are the *oracle* constrained and unconstrained strategies, i.e., the combination coefficients $\boldsymbol{\kappa}^c$ and $\boldsymbol{\kappa}^u$ use the true values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We show that the unconstrained strategy *always outperforms* not just the SMV and EW portfolios, but also the risk-free asset and the constrained strategy of Tu and Zhou (2011). One striking shortcoming of the constrained strategy $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^c)$ is that *it always overinvests in the SMV portfolio* relative to $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)$, and this effect is substantial for highly risk-averse investors.[3] Another drawback we prove theoretically is that for most degrees of risk aversion, the weights $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^c)$ are more extreme (i.e., have higher L_2 norm) than $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)$.

Our second contribution is to study the impact of the convexity constraint when the combination coefficients must be estimated and become $\hat{\boldsymbol{\kappa}}^c$ and $\hat{\boldsymbol{\kappa}}^u$. Although relaxing the convexity constraint helps improve the performance of the oracle portfolio combination, it inflates the sensitivity of combination coefficients to estimation errors in mean returns. We show analytically and via simulations that this larger sensitivity makes the unconstrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^u)$ *underperform* the constrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^c)$ in an interval of risk aversion γ. As a remedy, we introduce a *shrinkage estimator* of the unconstrained strategy. We derive the shrinkage intensity minimizing the mean-squared error and show that it shrinks intensively exactly for those levels of risk aversion where estimation errors hurt performance most. This estimator offers an appealing tradeoff between the constrained and unconstrained strategies.

Our third contribution is to demonstrate *theoretically* the robustness of our results to the i.i.d. normality assumption that is used to derive the optimal portfolio combinations. Specifically, we consider a model in which asset returns are *elliptically* distributed. We exploit results in El Karoui (2010, 2013) to derive the expected out-of-sample utility loss incurred when asset returns are elliptically distributed but the investor relies on the normal combination coefficients, and we show that this loss is typically very small.[4] These results give theoretical foundation to the empirical findings of Tu and Zhou (2004), who conclude that "[...] the certainty–equivalent losses associated with ignoring fat tails are small. This suggests that the normality assumption works well in evaluating portfolio performance for

[2] In contrast with EW that never invests in the risk-free asset, SMV does so whenever $\gamma \neq \mathbf{1}'\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$.

[3] For the dataset of 25 size-and-book-to-market portfolios (25SBTM) spanning July 1926 to December 2022, a sample size $T = 120$ months, and a risk aversion $\gamma = 10$, the constrained and unconstrained strategies allocate 57% and 20% to SMV, respectively.

[4] Suppose asset returns are t distributed with six degrees of freedom, i.e., a realistic excess kurtosis of three for monthly returns. For the 25SBTM dataset, $N/T = 0.2$, and a risk aversion $\gamma = 5$, the unconstrained combination coefficients are $(\kappa_1, \kappa_2) = (0.19, 0.29)$, while those under normality are $(\kappa_1, \kappa_2) = (0.20, 0.29)$. The expected out-of-sample utility loss by relying on the normal coefficients is a mere 0.007% per month.

a mean-variance investor."

Our fourth contribution is to compare empirically the performance of the three competing approaches to combine the SMV and EW portfolios: constrained, unconstrained, and shrinkage unconstrained. Moreover, we benchmark them against relevant combination strategies in Kan and Zhou (2007), Kan et al. (2021), and Bodnar et al. (2023), and against the EW and SGMV portfolios. We compare the strategies in terms of out-of-sample utility net of transaction costs, across three datasets of characteristic-sorted portfolios, one dataset of industry-sorted portfolios, and two datasets of 50 and 100 U.S. individual stocks.

The empirical observations match our theory. The constrained strategy can outperform the unconstrained strategy for a range of generally small γ's but quickly deteriorates otherwise, in which case it often underperforms the risk-free asset and the two-fund rule of Kan and Zhou (2007) that does not invest in the EW portfolio. In contrast, the unconstrained strategy performs consistently well, always outperforms the risk-free asset, and also the two-fund rule in most cases. Moreover, it delivers less turnover and transaction costs. Finally, the shrinkage unconstrained strategy closely matches the performance of either the constrained or the unconstrained strategy depending on which one outperforms given the risk aversion γ.

The two larger datasets of 50 and 100 individual stocks uncover additional insights on portfolio combination. First, combining the SMV portfolio with EW rather than with SGMV delivers much better performance and lower turnover, because EW suffers no estimation error unlike SGMV. This is true even if one uses shrinkage estimates of the covariance matrix. Second, one might deem it difficult to outperform the EW portfolio for such datasets, particularly when $N = 100$. Still, our proposed unconstrained strategies do not lose much relative to EW when $T = 120$, and actually outperform EW when $T = 240$, unlike the constrained strategy. Third, the value of the risk-aversion coefficient γ for which the constrained and unconstrained strategies coincide is larger than that for the characteristic and industry-sorted datasets, and thus the unconstrained and shrinkage unconstrained strategies also deliver a large improvement relative to the constrained strategy when γ is small and equal to one.[5]

2.2 The portfolio combination problem

Let $\boldsymbol{r}_t$ be the $N \times 1$ vector of asset excess returns at time t. As in Kan and Zhou (2007), Tu and Zhou (2011), and Ao et al. (2019), we assume that $\boldsymbol{r}_t \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is positive definite. In Section 2.5, we relax the normality assumption and only assume that $\boldsymbol{r}_t$ is elliptically distributed. At time T, the investor collects a sample of T i.i.d. returns, $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$. The investor then chooses her portfolio $\boldsymbol{w} = (w_1, ..., w_N)'$ on the risky assets,

[5]A detailed literature review is available in Section 2.A of the Supplementary Material.

the weight on the risk-free asset being $w_0 = 1 - \mathbf{1}_N'\boldsymbol{w}$ with $\mathbf{1}_N$ a $N \times 1$ vector of ones. We assume that $N + 4 < T < \infty$.[6]

When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, the optimal *mean-variance portfolio* maximizes utility,

$$U(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}, \tag{2.1}$$

where $\gamma > 0$ is the risk-aversion coefficient. Note that the risk-free asset has a utility of zero. The well-known solution to (2.1) and its utility are

$$\boldsymbol{w}^\star = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \quad \text{and} \quad U(\boldsymbol{w}^\star) = U^\star = \frac{\theta^2}{2\gamma}, \tag{2.2}$$

where $\theta = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$ is the maximum Sharpe ratio. Note that by investing in the mean-variance portfolio $\boldsymbol{w}^\star$, one invests in two funds: the fully invested tangent portfolio $\boldsymbol{w}_{tan} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/(\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$ and the risk-free asset. Specifically,

$$\boldsymbol{w}^\star = \frac{\gamma_{tan}}{\gamma}\boldsymbol{w}_{tan} \quad \text{and} \quad w_0^\star = 1 - \mathbf{1}_N'\boldsymbol{w}^\star = 1 - \frac{\gamma_{tan}}{\gamma} \quad \text{with} \quad \gamma_{tan} = \mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{2.3}$$

2.2.1 Parameter uncertainty and expected out-of-sample utility

The moments $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown in practice. As in Kan and Zhou (2007), Tu and Zhou (2011), and Kan et al. (2021), we rely on sample estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, which are

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{T-N-2}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})'. \tag{2.4}$$

The *sample mean-variance (SMV)* portfolio, which exploits $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, is defined as

$$\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}. \tag{2.5}$$

Under i.i.d. normality, $\widehat{\boldsymbol{\Sigma}}^{-1}$ is unbiased and independent of $\hat{\boldsymbol{\mu}}$, and thus the SMV portfolio is unbiased: $\mathbb{E}[\hat{\boldsymbol{w}}^\star] = \boldsymbol{w}^\star$. To account for estimation errors, we measure the performance of a sample portfolio $\hat{\boldsymbol{w}}$ via its *expected out-of-sample utility* (EU) as in Kan and Zhou (2007):[7]

$$EU(\hat{\boldsymbol{w}}) = \mathbb{E}[U(\hat{\boldsymbol{w}})] = \mathbb{E}\left[\hat{\boldsymbol{w}}'\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}\right]. \tag{2.6}$$

2.2.2 Combining the SMV and EW portfolios

We study the combination of the SMV portfolio $\hat{\boldsymbol{w}}^\star$ with the EW portfolio $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$:

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) = \kappa_1\hat{\boldsymbol{w}}^\star + \kappa_2\boldsymbol{w}_{ew}, \tag{2.7}$$

[6] In Section 2.C.2 of the Supplementary Material, we derive theoretical results in the high-dimensional asymptotic regime, i.e., $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, which is often considered in recent literature; see, e.g., Bodnar et al. (2018), Ao et al. (2019), Ledoit and Wolf (2020), and Bodnar et al. (2023).

[7] In Section 2.C.7 of the Supplementary Material, we also study the expected out-of-sample Sharpe ratio.

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$ is the vector of *combination coefficients*. We denote $\mu_{ew} = \boldsymbol{w}_{ew}'\boldsymbol{\mu}$ and $\sigma_{ew}^2 = \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\boldsymbol{w}_{ew}$. We make the mild assumption that $\mu_{ew} \geq 0$.[8] Moreover, we define γ_{ew} as[9]

$$\gamma_{ew} = \mu_{ew}/\sigma_{ew}^2, \tag{2.8}$$

and ψ_{ew}^2 as the difference between the maximum squared Sharpe ratio and that of EW:

$$\psi_{ew}^2 = \theta^2 - \theta_{ew}^2 \geq 0. \tag{2.9}$$

Our objective is to determine $\boldsymbol{\kappa}$ that maximizes the EU. If the investor wants to outperform the SMV and EW portfolios, it is sufficient to impose a convexity constraint,

$$\kappa_1 + \kappa_2 = 1, \tag{2.10}$$

and optimize a single combination coefficient. This is what Tu and Zhou (2011) consider, and they show that the resulting portfolio combination outperforms both the SMV and EW portfolios, due to the concavity of the EU. In the next section, we discuss why imposing the convexity constraint (2.10) is actually unnecessary and study the *unconstrained* combination.

In Section 2.C.3 of the Supplementary Material we also consider the combination of the SMV and SGMV portfolios in Kan and Zhou (2007). Moreover, in Section C.5, we consider the combination of the SMV, EW, and SGMV portfolios, but find that adding the SGMV portfolio to the mix only delivers a small EU gain in theory when the optimal combination coefficients are known, and thus often underperforms empirically when they are estimated.

2.3 The convexity constraint and oracle estimators

The convexity constraint (2.10) is necessary when combining fully invested portfolios as in Bodnar et al. (2018), Kan et al. (2021), and Bodnar et al. (2023), as it ensures that the combination is fully invested too. However, the SMV portfolio is *not* fully invested: $\mathbf{1}_N'\hat{\boldsymbol{w}}^\star = \mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}/\gamma \neq 1$, in general. Therefore, combining the SMV and EW portfolios amounts to investing in three funds: the sample tangent portfolio, the EW portfolio, *and* the risk-free asset. But, under constraint (2.10), a *single* coefficient κ_1 controls the weights on the three funds, the second one being $\kappa_2 = 1 - \kappa_1$. Reducing the weight κ_1 on the SMV portfolio to reduce exposure to estimation risk can only be done by increasing the weight $\kappa_2 = 1 - \kappa_1$ on the EW portfolio, which increases the exposure to risky assets because EW is fully invested. Thus, the investor cannot disentangle her exposure to financial risk and estimation risk. To optimally balance the three funds, we propose to relax the convexity constraint (2.10).

[8] We make this assumption to simplify the exposition of our theory. The results that need this assumption can also be extended to the case in which $\mu_{ew} < 0$.

[9] The parameter γ_{ew} has the following interpretation: an investor with risk aversion γ investing in the EW portfolio and the risk-free asset fully allocates her wealth to the EW portfolio if and only if $\gamma = \gamma_{ew}$.

2.3.1 Combination coefficients and underlying funds

In the next proposition, we derive the oracle constrained combination coefficients in Tu and Zhou (2011). Our presentation has two benefits relative to Tu and Zhou: we provide a closed-form expression for the coefficients, which makes their implementation easier and more efficient, and we also provide an explicit expression for the EU they deliver, which is useful for future results. We then derive similar results for the unconstrained combination.[10]

Proposition 2.1. *Concerning the combination of the SMV and EW portfolios in* (2.7)*:*

1. *The expected out-of-sample utility is*

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = \frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew} - \frac{\gamma}{2}\left(\frac{\kappa_1^2}{\gamma^2}\left(\theta^2+d\right) + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew}\right), \tag{2.11}$$

   where

$$d = cN/T + (c-1)\theta^2 \quad \textit{with} \quad c = \frac{(T-N-2)(T-2)}{(T-N-1)(T-N-4)} \geq 1. \tag{2.12}$$

2. *The oracle combination coefficients* $\boldsymbol{\kappa}^c$ *maximizing* (2.11) *under the convexity constraint* (2.10) *are*

$$\kappa_1^c = \frac{\psi_{ew}^2 + \ell^2}{\psi_{ew}^2 + \ell^2 + d} \in [0,1] \quad \textit{and} \quad \kappa_2^c = 1 - \kappa_1^c \in [0,1], \tag{2.13}$$

   where $\ell = \sigma_{ew}(\gamma - \gamma_{ew})$*. Moreover, the oracle unconstrained combination coefficients* $\boldsymbol{\kappa}^u$ *maximizing* (2.11) *are*[11]

$$\kappa_1^u = \frac{\psi_{ew}^2}{\psi_{ew}^2 + d} \in [0,1] \quad \textit{and} \quad \kappa_2^u = \frac{\gamma_{ew}}{\gamma}(1-\kappa_1^u) \in \mathbb{R}, \tag{2.14}$$

   which are equal to $\boldsymbol{\kappa}^c$ *in* (2.13) *if and only if* $\gamma = \gamma_{ew}$ *in* (2.8)*.*

3. *The expected out-of-sample utility of the constrained and unconstrained portfolio combinations, denoted* $\hat{\boldsymbol{w}}^c = \hat{\boldsymbol{w}}(\boldsymbol{\kappa}^c)$ *and* $\hat{\boldsymbol{w}}^u = \hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)$*, are*

$$EU(\hat{\boldsymbol{w}}^c) = U^\star - \frac{d}{2\gamma}\kappa_1^c, \tag{2.15}$$

$$EU(\hat{\boldsymbol{w}}^u) = U^\star - \frac{d}{2\gamma}\kappa_1^u. \tag{2.16}$$

   Both combinations outperform the SMV and EW portfolios, the unconstrained combination always outperforms the risk-free asset, but the constrained combination underperforms the risk-free asset if and only if $d > \theta^2$ *and*

$$\gamma > \gamma_{ew}\left(1 + \sqrt{1 + \frac{\theta^4/\theta_{ew}^2}{d-\theta^2}}\right) = \gamma_{neg}. \tag{2.17}$$

   Moreover, γ_{neg} *decreases with* d *defined in* (2.12)*.*

[10] All the proofs are available in Section 2.D of the Supplementary Material.

[11] Kan and Wang (2023, Section 2.3.2) use similar coefficients in a different context, that of combining a benchmark factor model with a set of test assets.

Several comments are in order. First, the parameter d measures estimation risk: it increases with N, decreases with T, and is equal to zero as $T \to \infty$, in which case both the constrained and unconstrained strategies correspond to the SMV portfolio.

Second, the convexity constraint impacts the combination coefficients on the SMV and EW portfolios: $\boldsymbol{\kappa}^c \neq \boldsymbol{\kappa}^u$ except if $\gamma = \gamma_{ew}$. We compare these coefficients in detail in Section 2.C.1 of the Supplementary Material. In summary, the constrained strategy *always overinvests in the SMV portfolio* in the sense that $\kappa_1^c \geq \kappa_1^u$. Moreover, it also overinvests in the EW portfolio, $\kappa_2^c \geq \kappa_2^u$, when $\gamma \in [\gamma_{ew}, (\theta^2 + d)/\mu_{ew}]$, which is typically a large interval. Finally, the constrained strategy underinvests in the risk-free asset relative to the unconstrained strategy when $\gamma \geq \gamma_{ew}$, and leverages less otherwise. These differences can be substantial. For the 25SBTM dataset, $\gamma = 5 > \gamma_{ew} = 1.86$, and $T = 120$, the constrained strategy allocates 22% to the tangent portfolio, 71% to the EW portfolio, and 7% to the risk-free asset. In contrast, the unconstrained strategy invests much more in the risk-free asset (55%) and much less in the tangent and EW portfolios (15% and 30%).

Third, because the convexity constraint severely constrains the allocation to the risk-free asset, the constrained strategy may underperform the risk-free asset if estimation risk is sufficient ($d > \theta^2$) and risk aversion too ($\gamma > \gamma_{neg}$). This problem does not arise in the unconstrained strategy that can fully invest in the risk-free asset if needed by setting $\boldsymbol{\kappa} = \mathbf{0}_2$, where $\mathbf{0}_2$ is a 2×1 vector of zeros. This deficiency of the constrained strategy is particularly notable because the two conditions under which it underperforms the risk-free asset can easily be met.[12]

Fourth, Proposition 2.1 suggests two cases in which outperforming the EW portfolio is likely difficult in practice for the unconstrained strategy. First, when ψ_{ew}^2 is small, which happens when EW provides a Sharpe ratio close to the maximum, e.g., because there is not much cross-sectional variation in mean returns.[13] Second, when estimation risk d is large. In these two cases, the weight κ_1^u on the SMV portfolio is close to zero, and the unconstrained strategy is close to the optimal combination of EW and the risk-free asset given by $\boldsymbol{\kappa} = (0, \gamma_{ew}/\gamma)$. Therefore, the oracle unconstrained strategy delivers a small improvement:

$$EU(\hat{\boldsymbol{w}}^u) - EU(\gamma_{ew}/\gamma \boldsymbol{w}_{ew}) = \frac{\psi_{ew}^2}{2\gamma} \kappa_1^u \tag{2.18}$$

decreases when d increases and ψ_{ew}^2 decreases. Thus, it may even underperform when the combination coefficients are estimated.

2.3.2 Outperformance in expected out-of-sample utility

By relaxing the convexity constraint, the unconstrained strategy delivers a larger EU than the constrained strategy. In the next proposition, we quantify and decompose this gain.

[12] For the 25SBTM dataset and $T = 120$, $d > \theta^2$ holds as soon as $N > 8$, and $\gamma_{neg} = 5.27$.

[13] Provided $\boldsymbol{\Sigma}$ is a constant diagonal matrix, $\psi_{ew}^2 = 0$ if and only if $\boldsymbol{\mu} = c\mathbf{1}_N$ for some $c \in \mathbb{R}$.

Proposition 2.2. *The unconstrained strategy always delivers a larger expected out-of sample utility (EU) than the constrained strategy:*

$$EU(\hat{\boldsymbol{w}}^u) - EU(\hat{\boldsymbol{w}}^c) = \frac{d}{2\gamma}\left(\kappa_1^c - \kappa_1^u\right) \geq 0, \tag{2.19}$$

with equality if and only if $\gamma = \gamma_{ew}$ or $\gamma = \infty$. Moreover, the gain in (2.19) increases with estimation risk d and can be decomposed as $\Delta_{smv} + \Delta_{ew} + \Delta_{cov}$, where:

1. *Δ_{smv} is the difference in EU coming from the exposure to the SMV portfolio:*[14]

$$\Delta_{smv} = EU(\kappa_1^u \hat{\boldsymbol{w}}^\star) - EU(\kappa_1^c \hat{\boldsymbol{w}}^\star) = \frac{d^2\ell^2\left[\ell^2(\psi_{ew}^2 + d - \theta_{ew}^2) - 2\theta_{ew}^2(d + \psi_{ew}^2)\right]}{2\gamma(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2 + d)^2}. \tag{2.20}$$

2. *Δ_{ew} is the difference in EU coming from the exposure to the EW portfolio:*[15]

$$\begin{aligned}\Delta_{ew} &= EU(\kappa_2^u \boldsymbol{w}_{ew}) - EU(\kappa_2^c \boldsymbol{w}_{ew}) \\ &= \frac{d\ell(\theta^2 + d - \mu_{ew}\gamma)}{2\gamma(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2 + d)^2}\Big[d\ell(\theta^2 + d - \mu_{ew}\gamma) - 2\theta_{ew}\psi_{ew}^2(\psi_{ew}^2 + \ell^2 + d)\Big].\end{aligned} \tag{2.21}$$

3. *Δ_{cov} is the difference in EU coming from the covariance between the return of the SMV and EW portfolios:*

$$\begin{aligned}\Delta_{cov} &= \mathbb{E}\left[(\kappa_1^c \hat{\boldsymbol{w}}^\star)'\boldsymbol{\Sigma}(\kappa_2^c \boldsymbol{w}_{ew})\right] - \mathbb{E}\left[(\kappa_1^u \hat{\boldsymbol{w}}^\star)'\boldsymbol{\Sigma}(\kappa_2^u \boldsymbol{w}_{ew})\right] \\ &= \frac{d\ell\theta_{ew}}{\gamma(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2 + d)^2}\Big[(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2) - \ell\theta_{ew}(\psi_{ew}^2(\psi_{ew}^2 + \ell^2) - d^2)\Big].\end{aligned} \tag{2.22}$$

Proposition 2.2 shows that the EU gain delivered by the unconstrained strategy increases with estimation risk d. This is because the constrained strategy overinvests in the SMV portfolio, which is the only portfolio in the combination that is exposed to estimation risk.

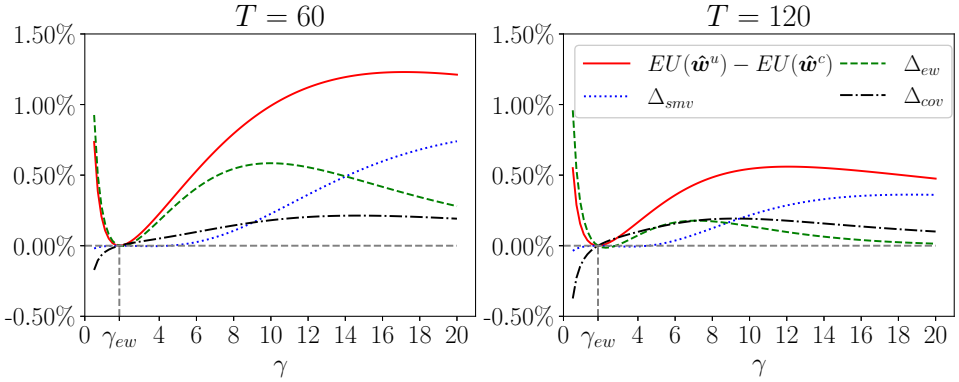
We illustrate Proposition 2.2 in Figure 2.1 using the 25SBTM dataset. The total EU gain delivered by the unconstrained strategy can be substantial when γ departs from $\gamma_{ew} = 1.86$. When $T = 120$ and γ varies between 0.5 and 20, the monthly EU gain

[14] $\Delta_{smv} \leq 0$ if $\psi_{ew}^2 + d - \theta_{ew}^2 \leq 0$, and else $\Delta_{smv} \geq 0$ if $\gamma \geq \gamma_{ew}\left(1 + \sqrt{\frac{2(\psi_{ew}^2+d)}{\psi_{ew}^2+d-\theta_{ew}^2}}\right)$, and $\Delta_{smv} < 0$ otherwise.

[15] If d is small enough such that $d^2(\psi_{ew}^2 + d) - 8\theta_{ew}^2\psi_{ew}^2(2\psi_{ew}^2 + d) \leq 0$, then $\Delta_{ew} \leq 0$ if $\gamma \in [\gamma_{ew}, (\theta^2 + d)/\mu_{ew}]$, and else $\Delta_{ew} > 0$. Otherwise, $\Delta_{ew} \geq 0$ if either $\gamma \leq \gamma_{ew}$, $\gamma \geq (\theta^2 + d)/\mu_{ew}$, or

$$\gamma \in \left[\frac{d^2 + d(\theta^2 + \theta_{ew}^2) + 4\theta_{ew}^2\psi_{ew}^2 \pm \sqrt{(\psi_{ew}^2 + d)(d^2(\psi_{ew}^2 + d) - 8\theta_{ew}^2\psi_{ew}^2(2\psi_{ew}^2 + d))}}{2\mu_{ew}(2\psi_{ew}^2 + d)}\right].$$

Figure 2.1: Expected out-of-sample utility of constrained versus unconstrained strategy



Notes. This figure depicts the gain in expected out-of-sample utility (red solid) delivered by the unconstrained combination strategy $\hat{\boldsymbol{w}}^u$ relative to the constrained strategy $\hat{\boldsymbol{w}}^c$, and its decomposition in the three components given in Proposition 2.2: Δ_{smv} (dotted blue), Δ_{ew} (dashed green) and Δ_{cov} (dash-dotted black). The figure is constructed by calibrating the population vector of means and covariance matrix of excess asset returns from monthly returns on the 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to December 2022. We consider sample sizes $T = 60$ (left panel) and $T = 120$ (right panel) and depict the utility gain as a function of the risk-aversion coefficient γ between 0.5 and 20.

in (2.19) goes up to around 0.5-0.6%, i.e., 6-7% annually, and it is even larger when $T = 60$. Moreover, when γ is around γ_{ew} and is of small or medium magnitude, most of the EU gain comes from taking an optimal exposure to the EW portfolio. This is because while the unconstrained coefficient κ_2^u in (2.14) is proportional to $1/\gamma$ and thus is large when γ is small, the constrained coefficient κ_2^c in (2.13) is bounded above by $\max_\gamma \kappa_2^c = d/(\psi_{ew}^2 + d)$ and thus is too small. In contrast, when γ gets large, most of the EU gain comes from taking an optimal exposure to the SMV portfolio. This is because as γ increases, both (κ_2^u, κ_2^c) go to zero but $\kappa_1^u = \psi_{ew}^2/(\psi_{ew}^2 + d)$ is independent of γ while κ_1^c in (2.14) goes to one and thus is too large. The third component, Δ_{cov}, is such that the three components sum to (2.19). It is negative for $\gamma \leq \gamma_{ew}$ due to the larger exposure to the EW portfolio in the unconstrained strategy, and is positive otherwise.

2.3.3 Outperformance relative to the optimal two-fund strategy

We show that the constrained strategy does not fully benefit from investing in the EW portfolio. Indeed, it can underperform the unconstrained two-fund strategy of Kan and Zhou (2007) that drops the EW portfolio from the combination in (2.7), which is defined as

$$\hat{\boldsymbol{w}}^{2f} = \kappa^{2f}\hat{\boldsymbol{w}}^\star \quad \text{with} \quad \kappa^{2f} = \frac{\theta^2}{\theta^2 + d}. \tag{2.23}$$

Specifically, we show that while the unconstrained three-fund strategy $\hat{\boldsymbol{w}}^u$ always outperforms the two-fund strategy $\hat{\boldsymbol{w}}^{2f}$, the constrained three-fund strategy $\hat{\boldsymbol{w}}^c$ does so only if $\gamma \leq 2\gamma_{ew}$.

Proposition 2.3. *The unconstrained three-fund strategy $\hat{\boldsymbol{w}}^u$ always delivers a larger expected out-of-sample utility than the unconstrained two-fund strategy $\hat{\boldsymbol{w}}^{2f}$. In contrast, the constrained three-fund strategy $\hat{\boldsymbol{w}}^c$ does so if and only if $\gamma \leq 2\gamma_{ew}$.*

2.3.4 Less extreme portfolio weights

Finally, another benefit of the unconstrained strategy is that it delivers on average less extreme weights on risky assets, i.e., with smaller norm. Large weights are undesirable because they often result in larger turnover and transaction costs, which we confirm in Section 2.6.

Proposition 2.4. *Let $\gamma \in [\gamma_{ew}, (\theta^2 + d)/\mu_{ew}]$. Then, the expected portfolio weights of the unconstrained strategy have a smaller L_2-norm than those of the constrained strategy: $||\boldsymbol{w}^u||_2 \leq ||\boldsymbol{w}^c||_2$, where $\boldsymbol{w}^u = \mathbb{E}[\hat{\boldsymbol{w}}^u]$ and $\boldsymbol{w}^c = \mathbb{E}[\hat{\boldsymbol{w}}^c]$.*

This result can be explained because, as discussed in Section 2.3.1, the constrained strategy always allocates more weight to SMV relative to the unconstrained strategy, and also to EW when $\gamma \in [\gamma_{ew}, (\theta^2 + d)/\mu_{ew}]$. This interval is wide in practice; it is equal to $[1.86, 43.1]$ for the 25SBTM dataset and $T = 120$. Moreover, this is only a sufficient condition: the unconstrained strategy may deliver a smaller L_2-norm when γ is outside this interval, too.

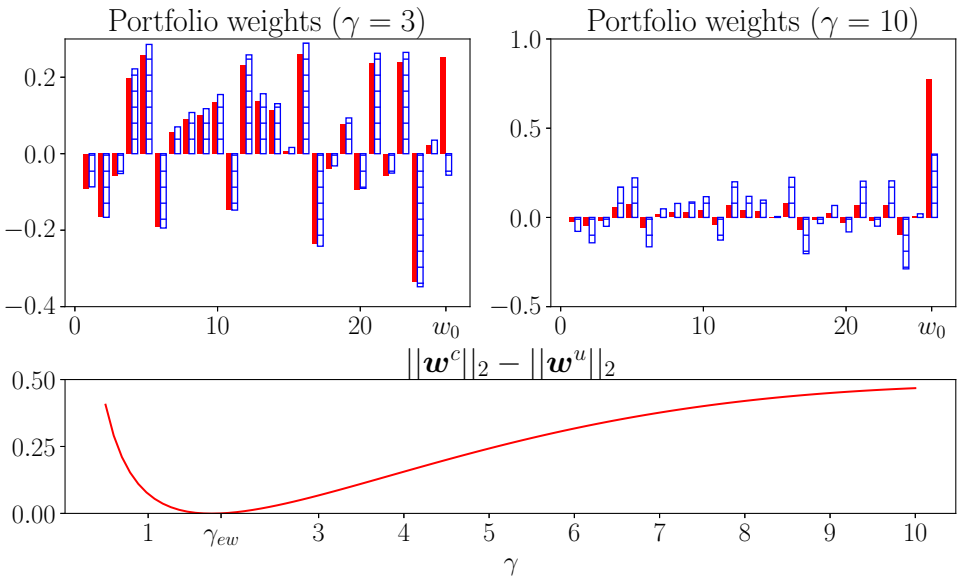
We illustrate Proposition 2.4 in Figure 2.2 for the 25SBTM dataset and $T = 120$. When $\gamma = 3$, the unconstrained strategy allocates a positive weight to the risk-free asset, whereas the constrained strategy allocates a negative weight. When $\gamma = 10$, the L_2-norm is much smaller for the unconstrained strategy, $0.25 = ||\boldsymbol{w}^u||_2 < ||\boldsymbol{w}^c||_2 = 0.72$, and consequently it allocates much more weight to the risk-free asset, $0.77 = w_0^u > w_0^c = 0.35$. Finally, the constrained strategy delivers a larger L_2-norm for any γ in the considered range $\gamma \leq 10$.

2.4 Shrinkage estimator of unconstrained combination

The results in Section 2.3 hold for the *oracle* combination coefficients. To assess the impact of the convexity constraint in practice, we need to compare the EU of the constrained and unconstrained strategies when $\boldsymbol{\kappa}^c$ and $\boldsymbol{\kappa}^u$ are fully estimated. However, they are complex functions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and thus, this is not feasible analytically. Still, what we want to capture is the *difference* in the way estimation errors impact the constrained and unconstrained coefficients, and we show that the main part of this difference is treatable analytically.

We proceed as follows. First, we show that the estimation errors are essentially driven by the scaling coefficient μ_{ew} in κ_2^u.Second, we study analytically the impact of estimation errors resulting from μ_{ew} in κ_2^u on the EU of the unconstrained strategy, and propose a

Figure 2.2: Portfolio weights and L_2-norms of constrained versus unconstrained strategy



Notes. The two upper panels depict the expected portfolio weights allocated to the 25 risky assets and the risk-free asset (w_0), for risk-aversion coefficients of $\gamma = 3$ and $\gamma = 10$. The red solid bars depict the expected weights in the unconstrained strategy, and the blue striped bars those in the constrained strategy. The bottom panel depicts the difference between the L_2-norm of the unconstrained and constrained portfolio weights on the risky assets as a function of γ. The figure is constructed by calibrating the population vector of means and covariance matrix of excess asset returns from monthly returns on the 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to December 2022, and using a sample size $T = 120$.

shrinkage estimator of the latter. Third, we compare via simulations the EU of the fully estimated constrained, unconstrained, and shrinkage unconstrained strategies.

2.4.1 Main impact of estimation errors in combination coefficients

In this section, we provide simulation and theoretical evidence that the main impact of dropping the convexity constraint on estimation risk is that κ_2^u becomes much more sensitive to the errors in combination coefficients. This is because the constrained coefficients in (2.13) are both between zero and one whereas, for the unconstrained coefficients in (2.14), it is only the case for κ_1^u while κ_2^u is unbounded.

In Section 2.B of the Supplementary Material, we assess this difference using simulations. We simulate normal asset returns with parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and repeatedly estimate the combination coefficients. We find first that errors in $\boldsymbol{\mu}$ have a much larger impact on the $\boldsymbol{\kappa}$'s than errors in $\boldsymbol{\Sigma}$. And second, while κ_1^c, κ_2^c, and κ_1^u are not very sensitive to estimation errors in $\boldsymbol{\mu}$, κ_2^u is highly impacted by these errors because it is directly proportional to $\boldsymbol{\mu}$ via μ_{ew}.

In Section 2.B of the Supplementary Material, we also give theoretical complement to these simulation results. Specifically, we derive in closed form the estimation errors in combination coefficients as a function of the estimation errors in mean returns, $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$. These expressions allow us to confirm that $\boldsymbol{\varepsilon}$ has a much larger impact on $\hat{\kappa}_2^u$ than on the

other coefficients, and that the impact of $\boldsymbol{\varepsilon}$ on $\hat{\kappa}_2^u$ comes mostly from the proportionality to $\hat{\mu}_{ew}$.

2.4.2 Alleviating errors in unconstrained combination coefficients

Following Section 2.4.1, we assess the impact of estimation risk in the combination coefficients by computing the EU loss of the unconstrained strategy when only μ_{ew} is estimated in κ_2^u:

$$\tilde{\kappa}_2^u = \frac{\hat{\mu}_{ew}}{\gamma\sigma_{ew}^2}(1-\kappa_1^u) \quad \text{with} \quad \hat{\mu}_{ew} = \boldsymbol{w}_{ew}'\hat{\boldsymbol{\mu}}. \tag{2.24}$$

Proposition 2.5. *Let the combination coefficient κ_2^u be estimated by $\tilde{\kappa}_2^u$ in (2.24). Then,*

1. *The expected out-of-sample utility of the estimated unconstrained strategy satisfies*

$$EU\left(\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u)\right) - EU(\hat{\boldsymbol{w}}^u) = -\frac{1-(\kappa_1^u)^2}{2\gamma T} \leq 0, \tag{2.25}$$

   where $EU(\hat{\boldsymbol{w}}^u) = EU\left(\hat{\boldsymbol{w}}(\kappa_1^u, \kappa_2^u)\right)$ is the utility of the oracle strategy in (2.16).

2. *Let $N \geq 2$. Then, the oracle constrained strategy delivers a larger expected out-of-sample utility than the estimated unconstrained strategy, $EU(\hat{\boldsymbol{w}}^c) \geq EU(\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u))$, if and only if γ belongs to the following interval:*

$$\gamma \in \left[\gamma_{ew} \pm \sigma_{ew}^{-1}\sqrt{\frac{(\psi_{ew}^2+d)(2\psi_{ew}^2+d)}{dT(\psi_{ew}^2+d)-(2\psi_{ew}^2+d)}}\right]. \tag{2.26}$$

   Moreover, the length of this interval decreases with estimation risk d.

Proposition 2.5 shows that when γ is sufficiently close to γ_{ew} and thus the EU gain provided by the oracle unconstrained strategy in (2.19) is not large, the constrained strategy can deliver a larger EU due to the impact of estimation errors in μ_{ew} on κ_2^u. However, as d increases and the EU loss of the oracle constrained strategy in (2.19) is larger, the interval (2.26) is narrower.

The length of the interval in (2.26) is not negligible; for the 25SBTM dataset, the interval is $\gamma \in [1.86 \pm 1.64]$ when $T = 120$. Therefore, the constrained strategy is more desirable than the unconstrained one for most small values of γ, but the unconstrained strategy remains preferable when γ is medium or large. This result is expected because when γ increases, the constrained strategy overinvests in risky assets, as we discuss in Section 2.3.1.

To alleviate estimation errors, we propose to shrink $\tilde{\kappa}_2^u$ in (2.24) toward the target $1-\kappa_1^u$,

$$\tilde{\kappa}_2^u(\delta) = (1-\delta)\tilde{\kappa}_2^u + \delta(1-\kappa_1^u). \tag{2.27}$$

This target is justified because $1-\kappa_1^u$ is less sensitive to estimation errors since it is bounded between zero and one, and because it is equal to the true κ_2^u when $\gamma=\gamma_{ew}$, which is the γ for which estimation errors hurt most relative to $\boldsymbol{\kappa}^c$ as shown in Proposition 2.5.

Proposition 2.6. *The following holds about $\tilde{\kappa}_2^u(\delta)$ in (2.27):*

1. *The shrinkage intensity δ minimizing the mean squared error of $\tilde{\kappa}_2^u(\delta)$ is*

$$\delta^\star=\frac{1}{1+T\ell^2}\in[0,1]. \tag{2.28}$$

Moreover, $\delta^\star=1$ for $\gamma=\gamma_{ew}$ and $\delta^\star\to 0$ as $T\to\infty$ and $\gamma\to\infty$.

2. *The optimal shrinkage estimator of κ_2^u is*

$$\tilde{\kappa}_2^u(\delta^\star)=\frac{1+\frac{\hat{\mu}_{ew}}{\gamma\sigma_{ew}^2}T\ell^2}{1+T\ell^2}(1-\kappa_1^u), \tag{2.29}$$

and is less sensitive to $\hat{\mu}_{ew}$ than $\tilde{\kappa}_2^u$:

$$\frac{\frac{\partial}{\partial\hat{\mu}_{ew}}\tilde{\kappa}_2^u(\delta^\star)}{\frac{\partial}{\partial\hat{\mu}_{ew}}\tilde{\kappa}_2^u}=1-\delta^\star\leq 1. \tag{2.30}$$

3. *$\tilde{\kappa}_2^u(\delta^\star)$ delivers a larger expected out-of-sample utility than $\tilde{\kappa}_2^u$:*

$$EU(\hat{\boldsymbol{w}}(\kappa_1^u,\tilde{\kappa}_2^u(\delta^\star)))-EU(\hat{\boldsymbol{w}}(\kappa_1^u,\tilde{\kappa}_2^u))=\frac{\delta^\star(1-(\kappa_1^u)^2)}{2\gamma T}\geq 0. \tag{2.31}$$

The optimal shrinkage intensity $\delta^\star$ behaves as desired, i.e., it shrinks $\tilde{\kappa}_2^u$ more intensively when γ is closer to γ_{ew} and when T decreases. As a result, $\tilde{\kappa}_2^u(\delta^\star)$ is less sensitive to $\hat{\mu}_{ew}$ than $\tilde{\kappa}_2^u$. Moreover, when $\delta^\star$ is known, $\tilde{\kappa}_2^u(\delta^\star)$ delivers a larger EU than $\tilde{\kappa}_2^u$.

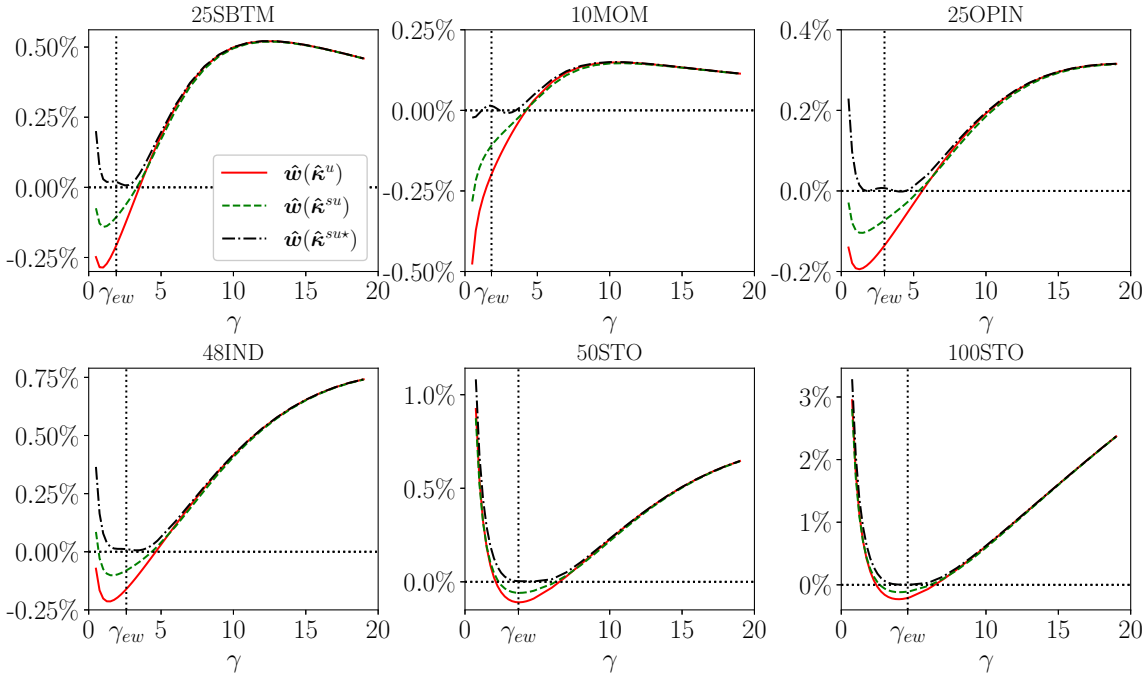
In Section 2.C.3 of the Supplementary Material, we derive a similar shrinkage estimator for the unconstrained combination of the SMV and SGMV portfolios in Kan and Zhou (2007). We compare the performance of this shrinkage unconstrained strategy to that of the plain unconstrained strategy and find that it delivers an improved EU.

The theoretical results in this section only account for the impact of estimation error in μ_{ew} on κ_2^u for analytical tractability. Nonetheless, in order to *evaluate* the EU of the novel shrinkage unconstrained strategy and compare it to that of the constrained and unconstrained strategies, it is necessary to consider the *feasible* estimators of these strategies, i.e., with fully estimated combination coefficients. This is what we do in the next section.

2.4.3 Out-of-sample performance of feasible portfolio combinations

We now compare the EU of the three different combinations of the SMV and EW portfolios: constrained, unconstrained, and shrinkage unconstrained. The corresponding oracle combination coefficients are defined in part 2 of Proposition 2.1 and part 2 of Proposition 2.6, and they depend on the following functions of the population $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$: μ_{ew}, σ_{ew}^2, θ^2, ψ_{ew}^2.

Figure 2.3: Expected out-of-sample utility relative to the feasible constrained combination with simulated i.i.d. normal data



Notes. This figure depicts the difference between the EU of three different strategies relative to the EU of the feasible constrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^c)$ in (2.33): (i) the feasible unconstrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^u)$ in (2.34) (solid red), (ii) the feasible shrinkage unconstrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su})$ in (2.35) (dashed green), and (iii) the unfeasible shrinkage unconstrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su\star})$ in (2.36) that uses the oracle shrinkage intensity $\delta^\star$ (dash-dotted black). We consider a sample size $T = 120$ and report the EU difference as a function of the risk-aversion coefficient γ for all six datasets described in Section 2.6.1. 50STO and 100STO are 500 randomly drawn investment sets of 50 and 100 U.S. individual stocks from a total pool of 235 stocks. For these two datasets, we report the average EU across the 500 random selections. We obtain the EU using 100,000 Monte Carlo simulations assuming i.i.d. normal returns whose population moments are calibrated from the whole sample of each dataset.

As in Tu and Zhou (2011), we estimate μ_{ew} and σ^2_{ew} by plugging the maximum-likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, i.e.,

$$\hat{\mu}_{ew} = \boldsymbol{w}'_{ew}\hat{\boldsymbol{\mu}} \quad \text{and} \quad \hat{\sigma}^2_{ew} = \boldsymbol{w}'_{ew}\widehat{\boldsymbol{\Sigma}}_{ml}\boldsymbol{w}_{ew}, \quad \text{where} \quad \widehat{\boldsymbol{\Sigma}}_{ml} = \frac{T-N-2}{T}\widehat{\boldsymbol{\Sigma}}. \tag{2.32}$$

Moreover, we estimate θ^2 and ψ^2_{ew} using the adjusted estimators in Kan and Zhou (2007) and Kan and Wang (2023) because the plug-in estimators are highly biased, and we denote them as $\hat{\theta}^2$ and $\hat{\psi}^2_{ew}$. The feasible constrained, unconstrained, and shrinkage unconstrained combination strategies are obtained by plugging the estimates $(\hat{\mu}_{ew}, \hat{\sigma}^2_{ew}, \hat{\theta}^2, \hat{\psi}^2_{ew})$ in (2.13), (2.14), (2.27) and (2.28) to compute $\hat{\kappa}^c_1, \hat{\kappa}^u_1, \hat{\kappa}^u_2$ and $\hat{\kappa}^u_2(\hat{\delta}^\star)$:

$$\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^c) = \hat{\kappa}^c_1\hat{\boldsymbol{w}}^\star + (1-\hat{\kappa}^c_1)\boldsymbol{w}_{ew}, \tag{2.33}$$

$$\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^u) = \hat{\kappa}^u_1\hat{\boldsymbol{w}}^\star + \hat{\kappa}^u_2\boldsymbol{w}_{ew}, \tag{2.34}$$

$$\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su}) = \hat{\kappa}^u_1\hat{\boldsymbol{w}}^\star + \hat{\kappa}^u_2(\hat{\delta}^\star)\boldsymbol{w}_{ew}. \tag{2.35}$$

We evaluate the EU of the three feasible portfolio strategies in (2.33)–(2.35) via Monte Carlo simulations. Specifically, we calibrate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to the six datasets of monthly excess

returns we consider in the empirical analysis; see Section 2.6.1. Then, for each dataset, we simulate $M = 100{,}000$ times T monthly return observations for each of the N assets, from which we compute the feasible strategies $\hat{\boldsymbol{w}}_m(\hat{\boldsymbol{\kappa}}_m^c)$, $\hat{\boldsymbol{w}}_m(\hat{\boldsymbol{\kappa}}_m^u)$, and $\hat{\boldsymbol{w}}_m(\hat{\boldsymbol{\kappa}}_m^{su})$, with $m = 1, \ldots, M$. Finally, the EU of the three feasible strategies is approximated by

$$EU(\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}})) \approx \frac{1}{M}\sum_{m=1}^{M} \hat{\boldsymbol{w}}_m(\hat{\boldsymbol{\kappa}}_m)'\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}_m(\hat{\boldsymbol{\kappa}}_m)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_m(\hat{\boldsymbol{\kappa}}_m). \tag{2.36}$$

In Figure 2.3, we set $T = 120$ and depict as a function of γ the difference between the EU of $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^u)$ and $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su})$ and the EU of $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^c)$. We also report the difference for the *unfeasible* shrinkage unconstrained strategy that uses the oracle $\delta^\star$ in (2.28), i.e.,

$$\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su\star}) = \hat{\kappa}_1^u\hat{\boldsymbol{w}}^\star + \hat{\kappa}_2^u(\delta^\star)\boldsymbol{w}_{ew}. \tag{2.37}$$

We make three main observations. First, $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su\star})$ that uses the oracle $\delta^\star$ essentially outperforms the constrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^c)$ for all γ, because $\delta^\star$ shrinks intensively when estimation errors hurt performance the most, i.e., when γ is close to γ_{ew}. Second, $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su})$ that uses the plug-in estimator $\hat{\delta}^\star$ has smaller EU than $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su\star})$, and underperforms $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^c)$ for a range of γ's around γ_{ew}. However, this EU loss is quite small, while the EU gain outside of this range can be substantial. Third, $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^u)$ significantly underperforms the shrinkage unconstrained strategy $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su})$ for γ close to γ_{ew}, while the two strategies are close to one another otherwise. Therefore, it is largely preferable to rely on $\hat{\boldsymbol{w}}(\hat{\boldsymbol{\kappa}}^{su})$ over the plain unconstrained strategy.

2.5 Robustness to normality

So far, our theory assumes that asset returns are i.i.d. normally distributed. This assumption being typically rejected in empirical data, an important question is whether the optimality of the portfolio combination strategies we study is robust to dropping this assumption. The literature addresses this question empirically or via simulations, and finds that assuming normality delivers accurate results even if the true distribution is not normal; see, e.g., Kan and Smith (2008, Section 3.5) and Tu and Zhou (2004). Comparing the results under t distributions with various degrees of freedom, the latter conclude that "[...] the certainty–equivalent losses associated with ignoring fat tails are small. This suggests that the normality assumption works well in evaluating portfolio performance for a mean-variance investor."

In this section, we give *theoretical* foundation to the findings of Tu and Zhou (2004). Specifically, we derive the constrained and unconstrained combinations of the SMV and EW portfolios when asset returns are *elliptically* (not just t) distributed. We derive the EU loss incurred by the investor who relies on the combination coefficients calibrated assuming normality while returns are elliptical, and we show that this EU loss is typically very small.

2.5.1 Elliptical model for asset returns

We consider the elliptical model for asset returns handled in El Karoui (2010, 2013), i.e.,

$$\boldsymbol{r}_t \overset{d}{=} \boldsymbol{\mu} + (\tau_t \boldsymbol{\Sigma})^{1/2} \boldsymbol{z}_t, \tag{2.38}$$

where $\{\tau_t\}_{t=1}^T$ are identically distributed, $\boldsymbol{z}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}_N, \boldsymbol{I}_N)$ is a standard multivariate normal, and $\{\tau_t\}_{t=1}^T$ is independent of $\{\boldsymbol{z}_t\}_{t=1}^T$. We impose $\mathbb{E}[\tau_t] = 1$ so that $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{r}_t$. For example, we obtain the t distribution with $\nu > 2$ degrees of freedom when

$$\tau_t \sim \frac{\nu - 2}{\chi^2_\nu}. \tag{2.39}$$

We only require that the second moment of $\boldsymbol{r}_t$ exists; moments of higher order may not exist.

As in the i.i.d. normal model, which we recover when $\nu \to \infty$ in (2.39), the distribution of $\boldsymbol{r}_t$ is identical for all t. However, this elliptical model is less restrictive than the i.i.d. normal model in two main ways. First, the distribution of returns is elliptical, which is a more general family of distributions than the normal. Second, we do not require $\{\tau_t\}_{t=1}^T$ to be independent, in which case there can be serial dependence in the asset returns.

2.5.2 Optimal portfolio combinations under elliptical returns

Recent literature shows that in the high-dimensional regime in which $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, one can rely on less restrictive distributional assumptions than the normal; see, e.g., Bodnar et al. (2018) and Bodnar et al. (2023) who only require the existence of the fourth moment. In the next proposition, we consider this asymptotic regime and exploit results in El Karoui (2010, 2013) to derive the out-of-sample utility of the portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$, and the oracle constrained and unconstrained combination coefficients, $\boldsymbol{\kappa}^c$ and $\boldsymbol{\kappa}^u$.

For comparison purposes, note that as $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, the constrained and unconstrained combination coefficients under normally distributed returns in (2.13) and (2.14) converge to

$$\bar{\boldsymbol{\kappa}}_n^c = (\bar{\kappa}_{1,n}^c, \bar{\kappa}_{2,n}^c) \quad \text{where} \quad \bar{\kappa}_{1,n}^c = \frac{\psi_{ew}^2 + \ell^2}{\psi_{ew}^2 + \ell^2 + \frac{\phi}{1-\phi}(1+\theta^2)} \quad \text{and} \quad \bar{\kappa}_{2,n}^c = 1 - \bar{\kappa}_{1,n}^c, \tag{2.40}$$

$$\bar{\boldsymbol{\kappa}}_n^u = (\bar{\kappa}_{1,n}^u, \bar{\kappa}_{2,n}^u) \quad \text{where} \quad \bar{\kappa}_{1,n}^u = \frac{\psi_{ew}^2}{\psi_{ew}^2 + \frac{\phi}{1-\phi}(1+\theta^2)} \quad \text{and} \quad \bar{\kappa}_{2,n}^u = \frac{\gamma_{ew}}{\gamma}(1 - \bar{\kappa}_{1,n}^u). \tag{2.41}$$

Proposition 2.7. *Let asset returns $\boldsymbol{r}_t$ follow the elliptical model in* (2.38)*. Let $\eta \geq 1$ be the unique positive solution to*

$$\mathbb{E}\left[\frac{1}{1 - \phi + \eta\phi\tau_t}\right] = 1, \tag{2.42}$$

and $\varphi \geq \eta^2$ *be defined as*

$$\varphi = \frac{1-\phi}{\eta^{-2} - \phi \mathbb{E}\left[\frac{\tau_t^2}{(1-\phi+\eta\phi\tau_t)^2}\right]}. \tag{2.43}$$

Let $N, T \to \infty$ *and* $N/T \to \phi \in (0,1)$ *while* $(\mu_{ew}, \sigma_{ew}^2, \theta^2)$ *stay bounded, and (Assumption-BB) in El Karoui (2013) holds.*[16] *Then, the following holds:*

1. *The out-of-sample utility of the portfolio combination* $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ *in* (2.7) *converges to*

$$U(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) \xrightarrow{p} \bar{U}(\boldsymbol{\kappa}) = \frac{\kappa_1}{\gamma}\eta\theta^2 + \kappa_2\mu_{ew} - \frac{\gamma}{2}\left(\frac{\kappa_1^2}{\gamma^2}\frac{\varphi\theta^2+\eta\phi}{1-\phi} + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\eta\mu_{ew}\right). \tag{2.44}$$

2. *The oracle constrained combination coefficients maximizing* (2.44) *under the convexity constraint* (2.10), $\bar{\boldsymbol{\kappa}}^c = (\bar{\kappa}_1^c, \bar{\kappa}_2^c)$, *are*

$$\bar{\kappa}_1^c = \frac{\eta\psi_{ew}^2 + \sigma_{ew}\ell(\gamma - \eta\gamma_{ew})}{\eta\psi_{ew}^2 + \sigma_{ew}\ell(\gamma - \eta\gamma_{ew}) + (1-\eta)\gamma\mu_{ew} + \frac{(\varphi-\eta)\theta^2+\eta\phi(1+\theta^2)}{1-\phi}} \tag{2.45}$$

and $\bar{\kappa}_2^c = 1 - \bar{\kappa}_1^c$. *Moreover, the asymptotic loss in out-of-sample utility when using the normal coefficients* $\bar{\boldsymbol{\kappa}}_n^c$ *instead of the elliptical ones* $\bar{\boldsymbol{\kappa}}^c$ *is*

$$\bar{U}(\bar{\boldsymbol{\kappa}}_n^c) - \bar{U}(\bar{\boldsymbol{\kappa}}^c) = -\frac{1}{2\gamma}\left(\eta\psi_{ew}^2 + \sigma_{ew}\ell(\gamma - \eta\gamma_{ew})\right)\frac{(\bar{\kappa}_1^c - \bar{\kappa}_{1,n}^c)^2}{\bar{\kappa}_1^c}. \tag{2.46}$$

3. *The oracle unconstrained combination coefficients maximizing* (2.44), $\bar{\boldsymbol{\kappa}}^u = (\bar{\kappa}_1^u, \bar{\kappa}_2^u)$, *are*

$$\bar{\kappa}_1^u = \frac{\eta\psi_{ew}^2}{\eta^2\psi_{ew}^2 + \frac{(\varphi-\eta^2)\theta^2+\eta\phi(1+\eta\theta^2)}{1-\phi}} \quad \textit{and} \quad \bar{\kappa}_2^u = \frac{\gamma_{ew}}{\gamma}(1 - \eta\bar{\kappa}_1^u). \tag{2.47}$$

Moreover, the asymptotic loss in out-of-sample utility when using the normal coefficients $\bar{\boldsymbol{\kappa}}_n^u$ *instead of the elliptical ones* $\bar{\boldsymbol{\kappa}}^u$ *is*

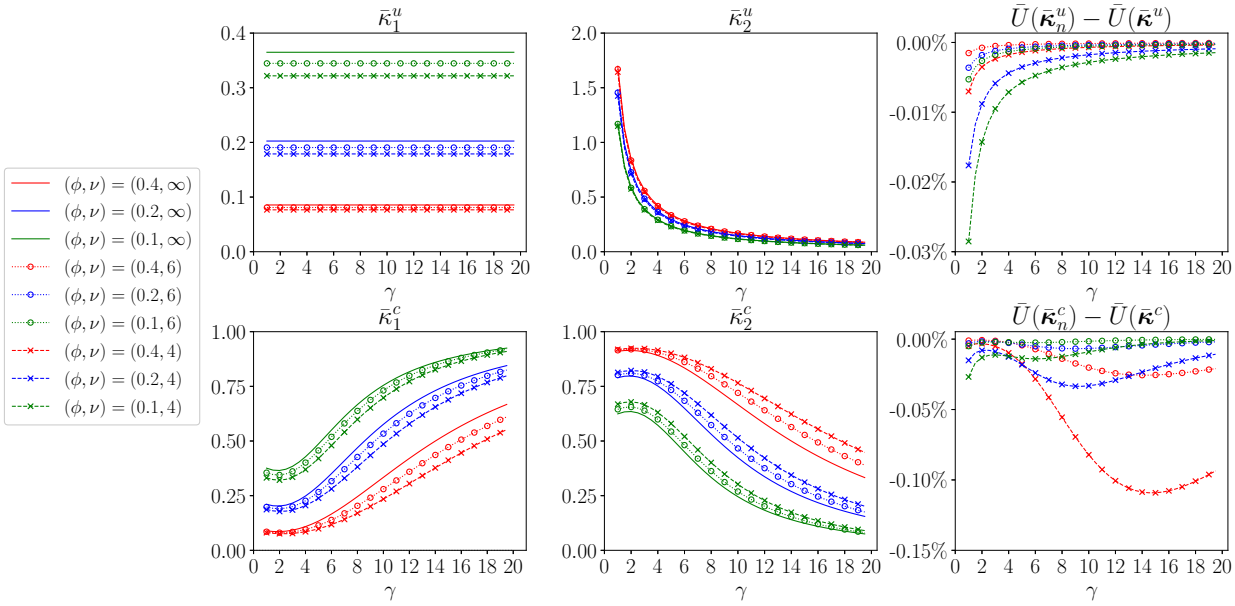
$$\bar{U}(\bar{\boldsymbol{\kappa}}_n^u) - \bar{U}(\bar{\boldsymbol{\kappa}}^u) = -\frac{1}{2\gamma}\left(\eta\psi_{ew}^2\frac{(\bar{\kappa}_1^u - \bar{\kappa}_{1,n}^u)^2}{\bar{\kappa}_1^u} + (\eta-1)^2\theta_{ew}^2(\bar{\kappa}_{1,n}^u)^2\right). \tag{2.48}$$

In Figure 2.4, we calibrate η and φ to the t distribution with ν degrees of freedom in (2.39).[17] First, we compare the normal combination coefficients in (2.40)–(2.41) with the elliptical ones in (2.45) and (2.47) for $\phi = (0.1, 0.2, 0.4)$ and $\nu = (4, 6)$, and find they are close to each other. For example, when $\gamma = 5$, $\phi = 0.2$, and $\nu = 6$, which corresponds to a realistic excess kurtosis of three for monthly returns, the normal coefficients are $(\bar{\kappa}_{1,n}^u, \bar{\kappa}_{2,n}^u, \bar{\kappa}_{1,n}^c, \bar{\kappa}_{2,n}^c) = (0.20, 0.29, 0.30, 0.70)$, while the elliptical ones are $(\bar{\kappa}_1^u, \bar{\kappa}_2^u, \bar{\kappa}_1^c, \bar{\kappa}_2^c) =$

[16] (Assumption-BB) ensures that the distribution of the τ_t's does not put too much mass near zero.

[17] We recover the case of i.i.d. normal asset returns when $\nu \to \infty$, i.e., $\tau_t \to 1$. This situation leads to $\eta = \varphi = 1$, and thus, $\bar{\boldsymbol{\kappa}}_n^c = \bar{\boldsymbol{\kappa}}^c$ and $\bar{\boldsymbol{\kappa}}_u^c = \bar{\boldsymbol{\kappa}}^u$.

Figure 2.4: Combination coefficients and utility loss under normal versus elliptical returns



Notes. The left and middle panels depict the constrained and unconstrained combination coefficients when returns are normally distributed (solid, no markers), as in (2.40)–(2.41), and elliptically distributed (dashed with cross markers and dotted with circle markers), as in (2.45) and (2.47). The right panels depict the expected out-of-sample utility loss when using the normal coefficients instead of the elliptical ones, as in (2.46) and (2.48). We consider the high-dimensional asymptotic regime in which $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$. The elliptical distribution is a t distribution with ν degrees of freedom. We depict the combination coefficients and utility losses as a function of the risk-aversion coefficient γ for different (ϕ, ν). The figure is constructed by calibrating the population vector of means and covariance matrix of asset excess returns from monthly returns on the 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to December 2022.

$(0.19, 0.29, 0.27, 0.73)$. Second, we depict the EU loss in (2.46) and (2.48) and we observe that it is typically very small. For example, when $\gamma = 5$, $\phi = 0.2$, and $\nu = 6$, the EU loss is a mere 0.0007% and 0.0038% per month for the unconstrained and constrained strategies, respectively.

The theoretical results above suggest that we can safely use the normality assumption to derive combination strategies even if returns are non-normal. Moreover, in Section 2.C.8 of the Supplementary Material we compare the *empirical* net out-of-sample utility of the constrained and unconstrained strategies calibrated under the normal and t distributions, and we also find that the loss caused by the normality assumption is small or even negative.

2.6 Empirical analysis

2.6.1 Datasets and portfolio strategies

Table 2.1 lists the six datasets of monthly excess returns we consider in this empirical analysis, which are three datasets of characteristic-sorted portfolios, one of industry-sorted portfolios, and two composed of 50 and 100 individual stocks. To construct the datasets of individual stocks, we collect adjusted returns from CRSP for the 235 stocks traded on the three major U.S. stock exchanges that have an history of returns between 1998 and 2022. To ensure the robustness of our results to stock selection, we report the average

Table 2.1: List of datasets considered in the empirical analysis

Dataset	N	Time period	T_{tot}	Abbreviation
25 portfolios sorted on size and book-to-market	25	07/1926 - 12/2022	1158	25SBTM
10 portfolios sorted on momentum	10	01/1927 - 12/2022	1152	10MOM
25 portfolios sorted on operating profitability and investment	25	07/1963 - 12/2022	714	25OPIN
48 portfolios sorted on industry	48	07/1969 - 12/2022	642	48IND
500 random portfolios of 50 U.S. stocks	50	01/1998 - 03/2022	291	50STO
500 random portfolios of 100 U.S. stocks	100	01/1998 - 03/2022	291	100STO

Notes. This table lists the datasets of monthly excess returns we use in the empirical analysis of Section 2.6. T_{tot} is the total number of months in each dataset. The first four datasets are downloaded from Kenneth French's website, and the last two from CRSP.

performance across 500 datasets of size $N = 50$ and $N = 100$, randomly drawn from the total pool of 235 stocks.

We evaluate the out-of-sample performance of 10 portfolio strategies that we list and describe in Table 2.2. The three most relevant ones are those we study in this chapter that combine the SMV and EW portfolios. First and second, the unconstrained and shrinkage unconstrained strategies (UNC3F and SUNC3F). Third, the constrained strategy (CON3F). We also report the performance of the SMV portfolio, $(\kappa_1, \kappa_2) = (1, 0)$, to which CON3F, UNC3F, and SUNC3F tend when estimation risk $d \to 0$. Conversely, we also report the performance of the combination of the EW portfolio with the risk-free asset (EWRF), $(\kappa_1, \kappa_2) = (0, \gamma_{ew}/\gamma)$, to which UNC3F tends when estimation risk $d \to \infty$.

The remaining five strategies are other combination rules. First and second, the two-fund and three-fund rules in Kan and Zhou (2007) (KZ2F and KZ3F). Third and fourth, the combination rules in Kan et al. (2021) and Bodnar et al. (2023) that do not invest in the risk-free asset (KWZ and BOP). Fifth, the combination of the SGMV portfolio with the risk-free asset (GMVRF): $(\mu_g/(c\gamma))\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N$.

We implement the portfolio strategies both with the sample covariance matrix estimate in (2.4) and the linear shrinkage estimate of Ledoit and Wolf (2004).[18] This is because Kan et al. (2021) and Lassance et al. (2024a) find that using both a portfolio combination and a shrinkage covariance matrix delivers better performance than using each in isolation.

We need to estimate several parameters to obtain the combination coefficients in the 10 strategies. We estimate the parameters μ_{ew}, σ^2_{ew}, θ^2, and ψ^2_{ew} as in the simulation analysis of Section 2.4.3. We estimate μ_g and σ^2_g using the maximum-likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and estimate $\psi^2_g = \theta^2 - \mu^2_g/\sigma^2_g$ via the adjusted estimator of Kan and Zhou (2007).

Finally, consistent with DeMiguel et al. (2009b) and Martellini and Ziemann (2010), we consider risk-aversion coefficients $\gamma = 1, 5, 10$, and 15.

[18]We also consider the nonlinear shrinkage estimate of Ledoit and Wolf (2020) in Section 2.C.9 of the Supplementary Material and find the conclusions are similar to those obtained with linear shrinkage.

Table 2.2: List of portfolio strategies considered in the empirical analysis

Name	Funds	Description
UNC3F	tan-ew-rf	Unconstrained combination of the sample tangent portfolio, EW portfolio, and risk-free asset. See Proposition 2.1.
SUNC3F	tan-ew-rf	Shrinkage estimator of the unconstrained combination UNC3F. See Proposition 2.6.
CON3F	tan-ew-rf	Constrained combination of the sample tangent portfolio, EW portfolio, and risk-free asset in Tu and Zhou (2011).
KZ2F	tan-rf	Combination of the sample tangent portfolio and risk-free asset in Kan and Zhou (2007).
KZ3F	tan-gmv-rf	Combination of the sample tangent portfolio, SGMV portfolio, and risk-free asset in Kan and Zhou (2007).
KWZ	tan-gmv	Combination of the sample tangent portfolio and SGMV portfolio with no-risk free asset in Kan et al. (2021).
BOP	tan-gmv-ew	Combination of the sample tangent portfolio, SGMV portfolio, and EW portfolio with no-risk free asset in Bodnar et al. (2023).
EWRF	ew-rf	Combination of the EW portfolio and the risk-free asset.
GMVRF	gmv-rf	Combination of the SGMV portfolio and the risk-free asset.
SMV	tan-rf	Sample mean-variance portfolio.

Notes. This table lists the portfolio strategies we use in the empirical analysis of Section 2.6, which are estimated with the sample covariance matrix in (2.4) or the linear shrinkage estimate of Ledoit and Wolf (2004).

2.6.2 Performance measure and statistical significance

We measure portfolio performance using rolling windows. At the end of month t, we estimate portfolio k using the T previous months, and we compute its out-of-sample return in month $t+1$. We consider $T = 120$ and 240 months. We repeat this iteratively, giving a time series of $T_{tot} - T$ out-of-sample gross returns $r_{gross,k,t}$, where T_{tot} is the total number of months in the dataset. We then compute the net out-of-sample returns, $r_{net,k,t} = r_{gross,k,t}$ if $t = T+1$ and

$$r_{net,k,t} = (1 + r_{gross,k,t})\,(1 - p \times \text{turnover}_{k,t-1}) - 1 \quad \text{if} \quad t = T+2, \ldots, T_{tot}, \tag{2.49}$$

where p is the proportional cost required to rebalance the portfolio and

$$\text{turnover}_{k,t} = \sum_{i=1}^{N} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, T_{tot}, \tag{2.50}$$

with $w_{i,k,t}$ the weight of asset i in month t and $w_{i,k,(t-1)^+}$ the prior-month weight before rebalancing in month t. We set $p = 10$ basis points as in Ao et al. (2019). Finally, we compare the portfolio strategies in terms of *annualized out-of-sample utility net of transaction costs*,

$$U_k = 12 \times \left(\hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2\right), \tag{2.51}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the sample mean and variance of $r_{net,k,t}$. In Section 2.C.7 of the Supplementary Material, we also report the net out-of-sample Sharpe ratio.

We test the statistical significance of the difference in net utilities of UNC3F and SUNC3F versus CON3F by using the stationary block bootstrap approach of Politis and Romano (1994) and Ledoit and Wolf (2008, Remark 3.2) to produce the resulting p-values, using 10,000 bootstrap samples and an average block size of five. For the 50STO and 100STO datasets, we compute the p-values by merging the out-of-sample returns across the 500 randomly drawn datasets. Moreover, we identify the superior set of models at a confidence level of 5% using the model confidence set (MCS) of Hansen et al. (2011).[19]

2.6.3 Discussion of results

The out-of-sample performance of the 10 portfolio strategies listed in Table 2.2, and their statistical significance, is reported in Table 2.3 for the sample covariance matrix and Table 2.4 for the linear shrinkage covariance matrix. In Section 2.C.10 of the Supplementary Material, we also report the average monthly turnover. The sets of superior models obtained with the MCS test typically include many strategies, and thus in Tables 2.3 and 2.4 we underline those that are rejected. For the 50STO and 100STO datasets we obtain 500 sets of models, and we underline in the tables the three portfolio strategies that are most often rejected.

In the vast majority of cases, we find that exploiting the shrinkage covariance matrix of Ledoit and Wolf (2004) to estimate the portfolio strategies in Table 2.4 delivers better performance than exploiting the sample covariance matrix in Table 2.3. Therefore, we recommend investors to rely on a shrinkage covariance matrix, even if the theory builds on the sample one. However, the ranking of the different strategies is similar in both tables, and therefore the main conclusions we discuss below are applicable to both estimators.

In essentially all cases, the SMV portfolio delivers by far the worst out-of-sample performance. In particular, it is outperformed by UNC3F, SUNC3F, and CON3F that combine SMV with the EW portfolio, as we predict in part 3 of Proposition 2.1. As a result, the SMV portfolio is nearly consistently rejected by the MCS test.

The main comparison concerns the three-fund rules that combine tangent, EW, and the risk-free asset. When γ is close enough to the estimated values of γ_{ew}, the constrained strategy CON3F can outperform the unconstrained strategy UNC3F, in line with Proposition 2.5 and Figure 2.3. For the 25SBTM, 10MOM, 25OPIN, and 48IND datasets where the median estimated γ_{ew} is around three, this happens when $\gamma = 1$ and 5. For the 50STO and 100STO where the median estimated γ_{ew} is larger and around four, this happens when $\gamma = 5$. This result shows that the convexity constraint can help performance by alleviating the impact of estimation errors on combination coefficients.

[19]The MCS test needs as an input a performance difference for each pair of models and for all time t. However, our performance measure, the net out-of-sample utility, can only be computed over several out-of-sample monthly returns. Therefore, to provide a performance difference for all time t, we collect daily returns for all our datasets and evaluate the net out-of-sample utility of a given portfolio at time t from the next-month daily returns, using daily rebalancing.

Table 2.3: Annualized net out-of-sample utility (sample covariance matrix)

	$T = 120$						$T = 240$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
$\hat{\gamma}_{ew}^{med}$	3.20	2.78	3.15	3.16	3.95	3.95	2.88	2.67	3.16	3.29	4.43	4.43
Panel (a): $\gamma = 1$												
UNC3F	29.4	22.5	6.63	2.92	15.4***	10.7***	20.9	25.3	24.6	6.25	31.2***	30.0***
SUNC3F	30.1	22.5	7.67	3.43	13.9***	9.38***	20.8	25.2	23.9	5.11	30.4***	29.1***
CON3F	31.2	23.4	8.31	5.36	8.59	4.95	21.8	26.6	21.3	2.82	23.2	18.7
KZ2F	27.1	21.6	4.07	-0.98	-3.24	-8.13	19.1	25.0	17.0	-3.91	10.2	3.18
KZ3F	24.4	26.6	6.20	-7.33	-14.3	-21.4	22.2	30.0	30.8	2.57	18.7	5.85
KWZ	17.1	16.5	4.33	2.73	3.16	-2.45	16.0	24.0	15.4	0.66	10.3	9.07
BOP	17.0	7.12	-1.24	-18.4	-31.8	-25.7	15.6	20.9	13.6	-5.70	-51.1	-102
EWRF	8.32	5.72	-0.59	2.85	17.5	17.5	5.99	2.47	8.93	9.87	27.4	27.7
GMVRF	15.9	11.5	7.67	-7.49	-12.5	-18.1	16.1	8.33	28.1	4.11	18.6	5.77
SMV	-140	-21.1	-158	-492	-583	-3140	-47.5	8.99	-58.0	-217	-147	-505
Panel (b): $\gamma = 5$												
UNC3F	5.71	4.45	1.28	0.57	3.07***	2.12***	3.96	5.02*	4.90	1.25	6.24***	5.99***
SUNC3F	5.95	4.75	1.93	1.39	4.30***	3.61***	4.02	5.14**	5.02	1.30	6.79***	6.71***
CON3F	5.08	4.41	2.43	1.67	5.67	5.02	4.01	5.88	5.56	1.90	7.40	7.33
KZ2F	5.26	4.27	0.77	-0.21	-0.66	-1.65	3.59	4.96	3.37	-0.79	2.04	0.64
KZ3F	4.69	5.27	1.16	-1.51	-2.95	-4.39	4.24	5.95	6.13	0.50	3.71	1.12
KWZ	6.19	5.50	3.47	0.23	1.57	-10.5	7.22	8.15	8.24	0.61	6.78	4.91
BOP	4.35	3.05	1.73	-3.05	-4.00	-2.96	5.31	7.21	6.27	-0.62	-5.95	-20.3
EWRF	1.66	1.14	-0.12	0.57	3.49	3.49	1.20	0.49	1.79	1.97	5.48	5.55
GMVRF	3.07	2.29	1.49	-1.53	-2.59	-3.71	3.17	1.66	5.62	0.82	3.69	1.10
SMV	-30.7	-4.48	-32.8	-104	-120	-784	-10.4	1.69	-12.0	-44.2	-29.9	-104
Panel (c): $\gamma = 10$												
UNC3F	2.84***	2.22*	0.64*	0.28**	1.53***	1.06***	1.97**	2.51	2.45	0.62*	3.12***	3.00***
SUNC3F	2.76***	2.21*	0.52*	0.06**	1.30***	0.72***	1.92**	2.48	2.46	0.59**	3.05***	2.92***
CON3F	-0.74	1.06	-1.53	-4.01	-1.59	-2.80	0.29	2.24	1.13	-2.83	-0.16	-0.94
KZ2F	2.62	2.13	0.38	-0.11	-0.33	-0.83	1.78	2.48	1.68	-0.39	1.02	0.32
KZ3F	2.33	2.63	0.58	-0.76	-1.48	-2.20	2.11	2.98	3.07	0.25	1.85	0.56
KWZ	0.32	-1.06	-1.43	-4.60	-3.58	-24.1	2.30	1.63	2.84	-3.06	1.71	-0.69
BOP	-3.03	-2.74	-2.84	-6.31	-5.79	-7.84	0.27	1.04	1.05	-4.21	-2.65	-12.1
EWRF	0.83	0.57	-0.06	0.28	1.75	1.75	0.60	0.25	0.89	0.99	2.74	2.77
GMVRF	1.53	1.14	0.74	-0.77	-1.30	-1.86	1.58	0.83	2.81	0.41	1.84	0.55
SMV	-15.5	-2.26	-16.5	-52.3	-60.5	-403	-5.28	0.84	-6.01	-22.2	-15.0	-52.3
Panel (d): $\gamma = 15$												
UNC3F	1.89***	1.48***	0.42***	0.19***	1.02***	0.70***	1.31***	1.67	1.63**	0.42***	2.08***	2.00***
SUNC3F	1.84***	1.46***	0.38***	0.12***	0.94***	0.59***	1.29***	1.66	1.64**	0.41***	2.06***	1.98***
CON3F	-2.93	-0.11	-3.55	-7.85	-6.76	-10.5	-1.04	1.07	-0.63	-5.18	-3.39	-6.20
KZ2F	1.75	1.42	0.25	-0.07	-0.22	-0.55	1.18	1.65	1.12	-0.26	0.68	0.21
KZ3F	1.55	1.75	0.38	-0.51	-0.99	-1.47	1.40	1.98	2.04	0.17	1.23	0.37
KWZ	-4.96	-7.09	-6.56	-9.51	-8.94	-37.9	-2.18	-3.90	-2.29	-6.96	-3.39	-6.32
BOP	-8.55	-8.35	-7.59	-11.2	-10.1	-15.1	-4.05	-4.22	-3.73	-8.23	-6.36	-12.2
EWRF	0.55	0.38	-0.04	0.19	1.16	1.16	0.40	0.16	0.60	0.66	1.83	1.85
GMVRF	1.02	0.76	0.50	-0.51	-0.87	-1.24	1.05	0.55	1.87	0.27	1.23	0.36
SMV	-10.4	-1.51	-11.0	-35.0	-40.4	-271	-3.54	0.56	-4.02	-14.8	-10.0	-34.9

Notes. This table reports the annualized net out-of-sample utility in percentage points for the portfolio strategies in Table 2.2 and the datasets in Table 2.1 when using the sample covariance matrix. The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. For 50 and 100STO, we report the average performance across 500 randomly drawn datasets from a total pool of 235 stocks. We evaluate the statistical significance of the difference in net utilities of UNC3F vs CON3F and SUNC3F vs CON3F using the stationary block bootstrap approach of Politis and Romano (1994) and Ledoit and Wolf (2008) to produce p-values. The stars *, **, and *** indicate that the p-value is less than 10%, 5%, and 1%, respectively. The underlined numbers indicate the strategies that do not belong to the superior set of models of Hansen et al. (2011) at a confidence level of 5%. For 50STO and 100STO, we underline the three strategies that are most often rejected, or more in case of equalities.

Table 2.4: Annualized net out-of-sample utility (linear shrinkage covariance matrix)

	$T = 120$						$T = 240$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
$\hat{\gamma}_{ew}^{med}$	3.20	2.78	3.15	3.16	3.95	3.95	2.88	2.67	3.16	3.29	4.43	4.43
Panel (a): $\gamma = 1$												
UNC3F	37.3	31.6	14.3	3.70	15.9***	16.7***	34.1	30.5	26.4	4.37	30.5***	30.5***
SUNC3F	38.4	31.7	15.8	4.60	14.7***	15.5***	34.1	30.5	25.8	3.27	29.9***	29.8***
CON3F	42.2	34.5	18.8	6.93	11.5	13.2	36.1	32.4	23.7	0.91	24.3	21.8
KZ2F	38.6	32.6	15.1	1.28	-0.07	-0.51	34.0	31.0	19.7	-5.86	11.7	5.78
KZ3F	33.3	34.7	14.9	-3.33	-4.31	8.87	31.9	34.5	31.2	3.01	12.0	0.55
KWZ	23.9	25.7	10.3	4.67	4.14	4.82	24.4	28.6	16.6	-1.36	10.1	9.88
BOP	16.8	7.23	-1.24	-18.5	-31.9	-25.7	15.7	20.9	13.6	-5.67	-51.0	-102
EWRF	8.32	5.72	-0.59	2.85	17.5	17.5	5.99	2.47	8.93	9.87	27.4	27.7
GMVRF	14.9	10.0	10.7	-4.64	-3.09	9.55	11.8	7.77	27.1	8.41	11.9	0.21
SMV	-26.2	13.6	-112	-508	-539	-1838	1.39	21.4	-53.7	-236	-165	-626
Panel (b): $\gamma = 5$												
UNC3F	7.44	6.31	2.83	0.73	3.18***	3.34***	6.76	6.08**	5.26	0.87	6.10***	6.11***
SUNC3F	7.68	6.64	3.49	1.52	4.43***	4.85***	6.82	6.20**	5.34	0.92	6.61***	6.80***
CON3F	8.05	6.96	4.15	1.85	5.88	6.34	7.24	7.02	5.86	1.50	7.19	7.40
KZ2F	7.72	6.50	2.98	0.25	-0.02	-0.10	6.74	6.18	3.91	-1.18	2.34	1.16
KZ3F	6.59	6.93	2.92	-0.71	-0.93	1.76	6.31	6.88	6.21	0.59	2.37	0.04
KWZ	8.17	7.67	5.57	3.07	4.05	3.21	8.97	9.13	8.53	1.36	7.05	6.63
BOP	4.09	3.14	1.72	-3.02	-3.86	-2.94	5.28	7.12	6.22	-0.51	-5.69	-20.1
EWRF	1.66	1.14	-0.12	0.57	3.49	3.49	1.20	0.49	1.79	1.97	5.48	5.55
GMVRF	2.91	1.99	2.11	-0.97	-0.69	1.90	2.33	1.55	5.41	1.68	2.35	-0.03
SMV	-5.85	2.61	-23.3	-106	-110	-388	-0.07	4.22	-11.1	-48.1	-33.4	-128
Panel (c): $\gamma = 10$												
UNC3F	3.72**	3.15	1.41	0.36**	1.59***	1.67***	3.37	3.04	2.63	0.44**	3.05***	3.05***
SUNC3F	3.60**	3.13	1.26	0.11**	1.32***	1.32***	3.32	3.02	2.63	0.40**	2.96***	2.96***
CON3F	1.98	2.76	-0.46	-3.96	-1.44	-1.94	2.53	3.04	1.21	-3.37	-0.95	-1.36
KZ2F	3.86	3.25	1.49	0.12	-0.01	-0.05	3.37	3.09	1.95	-0.59	1.17	0.58
KZ3F	3.29	3.46	1.46	-0.36	-0.47	0.88	3.15	3.44	3.10	0.29	1.18	0.01
KWZ	2.28	0.56	0.72	-0.57	0.39	-0.59	3.37	2.31	3.18	-1.55	2.29	1.98
BOP	-3.26	-2.59	-2.82	-6.27	-5.67	-7.80	0.18	1.02	1.04	-4.10	-2.66	-12.0
EWRF	0.83	0.57	-0.06	0.28	1.75	1.75	0.60	0.25	0.89	0.99	2.74	2.77
GMVRF	1.45	0.99	1.05	-0.49	-0.35	0.95	1.16	0.77	2.71	0.84	1.17	-0.02
SMV	-2.97	1.30	-11.7	-53.1	-55.3	-195	-0.06	2.10	-5.56	-24.1	-16.7	-64.4
Panel (d): $\gamma = 15$												
UNC3F	2.48***	2.10	0.94***	0.24***	1.06***	1.11***	2.25*	2.03	1.75**	0.29***	2.03***	2.04***
SUNC3F	2.41***	2.08	0.89***	0.16***	0.97***	1.00***	2.23**	2.02	1.75**	0.28***	2.01***	2.01***
CON3F	-0.14	1.34	-2.57	-8.04	-6.62	-9.49	0.93	1.72	-0.59	-5.90	-4.68	-7.76
KZ2F	2.57	2.17	0.99	0.08	-0.01	-0.03	2.24	2.06	1.30	-0.39	0.78	0.39
KZ3F	2.19	2.31	0.97	-0.24	-0.31	0.59	2.10	2.29	2.07	0.20	0.79	0.01
KWZ	-2.58	-5.41	-4.03	-4.30	-3.51	-4.52	-1.21	-3.20	-1.80	-4.92	-2.52	-2.72
BOP	-8.59	-8.02	-7.46	-11.2	-10.2	-15.1	-4.02	-4.14	-3.66	-8.14	-6.41	-12.2
EWRF	0.55	0.38	-0.04	0.19	1.16	1.16	0.40	0.16	0.60	0.66	1.83	1.85
GMVRF	0.97	0.66	0.70	-0.33	-0.23	0.63	0.77	0.52	1.80	0.56	0.78	-0.01
SMV	-1.99	0.86	-7.80	-35.4	-36.9	-131	-0.04	1.40	-3.71	-16.1	-11.2	-43.0

Notes. This table reports the annualized net out-of-sample utility in percentage points for the portfolio strategies in Table 2.2 and the datasets in Table 2.1 when using the shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. For 50 and 100STO, we report the average performance across 500 randomly drawn datasets from a total pool of 235 stocks. We evaluate the statistical significance of the difference in net utilities of UNC3F vs CON3F and SUNC3F vs CON3F using the stationary block bootstrap approach of Politis and Romano (1994) and Ledoit and Wolf (2008) to produce p-values. The stars *, **, and *** indicate that the p-value is less than 10%, 5%, and 1%, respectively. The underlined numbers indicate the strategies that do not belong to the superior set of models of Hansen et al. (2011) at a confidence level of 5%. For 50STO and 100STO, we underline the three strategies that are most often rejected, or more in case of equalities.

However, when $\gamma = 1$, the CON3F strategy can sometimes largely underperform the UNC3F strategy. A first case in which this happens is for the 48IND dataset when $T = 240$. For example, when using the sample covariance matrix, UNC3F and CON3F deliver a net utility of 6.25 and 2.82, respectively. We can explain this theoretically because, as discussed in Section 2.3.1, CON3F overinvests in the SMV portfolio relative to UNC3F, and this exposure to the SMV portfolio is larger for $T = 240$ than 120. This overinvestment is problematic for the 48IND dataset because SMV performs notoriously poorly for industry portfolios that present little cross-sectional variation in sample mean returns (Kirby and Ostdiek, 2012).

A second case in which UNC3F largely outperforms CON3F for $\gamma = 1$ is for the 50STO and 100STO datasets, because the median estimated γ_{ew} is further from one relative to the other datasets, and thus there is more room for UNC3F to outperform CON3F for a small γ. For example, when $T = 120$ in Table 2.3, UNC3F delivers a net utility of 15.4 and 10.7 for the 50STO and 100STO datasets, respectively, versus 8.59 and 4.95 delivered by CON3F.

Although CON3F can perform well when $\gamma = 1$ or 5, it quickly deteriorates as γ increases, is outperformed by UNC3F in a statistically significant way, and is often out of the MCS set of models. Take the sample covariance matrix, $T = 120$, and $\gamma = 10$. Then, CON3F delivers net utilities of -0.74, 1.06, -1.53, -4.01, -1.59, -2.80 for the six datasets, versus 2.84, 2.22, 0.64, 0.28, 1.53, 1.06 for UNC3F. These negative utilities are consistent with Proposition 2.1, in which we show that CON3F underperforms the risk-free asset once γ is too large. Similarly, Proposition 2.3 predicts that CON3F underperforms the two-fund rule of Kan and Zhou (2007) (KZ2F) once γ is too large, and this is what we observe empirically too. The out-of-sample utility of UNC3F is on the contrary *always* positive, and in most cases larger than that of KZ2F. UNC3F is also *never* rejected by the MCS test.

Moreover, we find in Section 2.C.10 of the Supplementary Material that UNC3F systematically delivers a smaller turnover than that of CON3F, which is in line with Proposition 2.4.

We turn next to the *shrinkage* unconstrained strategy (SUNC3F), which alleviates the sensitivity to estimation errors of UNC3F, particularly for γ's close to γ_{ew}. Our results show that this objective is generally accomplished: SUNC3F outperforms UNC3F for all datasets when $\gamma = 5$, and for all datasets except 50STO and 100STO when $\gamma = 1$. For a larger $\gamma = 10$ and 15, and when $\gamma = 1$ for 50STO and 100STO, SUNC3F remains close to UNC3F, and thus largely outperforms CON3F. In Section 2.C.11 of the Supplementary Material, we replicate Figure 2.3 empirically and find that the results closely mimic the simulation findings. Finally, just like UNC3F, the SUNC3F strategy is never rejected by the MCS test.

Next, we compare our strategies with the EWRF and GMVRF strategies that do

not invest in the SMV. In the majority of cases, the two novel strategies we introduce, UNC3F and SUNC3F, largely outperform. Let us concentrate on the results using the linear shrinkage covariance matrix in Table 2.4. Then, UNC3F and SUNC3F outperform GMVRF and EWRF in all cases except for the 48IND dataset when $T = 240$, where they underperform both, and the 50STO and 100STO datasets when $T = 120$, where they underperform EWRF. These two cases in which EWRF is difficult to beat are exactly those we predict theoretically in Section 2.3.1. The first case is when ψ^2_{ew} is small, which happens for the 48IND dataset because there is only little cross-sectional variation in sample mean returns for industry-sorted portfolios (Kirby and Ostdiek, 2012). The second case is when estimation risk d is large, which happens in the 50STO and 100STO datasets when $T = 120$.

It is notable that even for the 100STO dataset and $T = 120$, a case of high estimation risk, UNC3F and SUNC3F lose little relative to EWRF in Table 2.4, while they outperform when $T = 240$. This is reassuring because it is not known a priori for which sample size EWRF will outperform, and thus investors can safely rely on UNC3F or SUNC3F without fearing to lose much. In contrast, CON3F can deliver large performance losses relative to EWRF.

We now compare the three-fund rule of Kan and Zhou (2007) (KZ3F) and its fully invested counterpart in Kan et al. (2021) (KWZ). When γ is small ($\gamma = 1$) and large ($\gamma = 10$ and 15), KZ3F generally outperforms KWZ because KWZ cannot invest in the risk-free asset. However, the contrary nearly systematically happens when $\gamma = 5$. We demonstrate in Section 2.C.4 of the Supplementary Material that this result can be explained with a similar reasoning to that in Section 2.4: Kan et al. (2021) impose the convexity constraint so that KWZ is fully invested in risky assets, which alleviates the impact of estimation errors on combination coefficients relative to KZ3F and can help improve performance for some γ's. As a result, our SUNC3F strategy largely outperforms KWZ when $\gamma = 1$, 10, and 15, but not $\gamma = 5$. Note also that KWZ and our two proposed strategies, UNC3F and SUNC3F, are superior to the fully invested combination of Bodnar et al. (2023) (BOP).

Finally, the comparison of the unconstrained combination of SMV and EW (UNC3F) with the combination of SMV and SGMV in Kan and Zhou (2007) (KZ3F) shows that it is more preferable to combine with EW than SGMV, for three main reasons. First, EWRF has a much lower turnover than GMVRF, and thus UNC3F has a smaller turnover than KZ3F too. Second, while EW has no estimation error, SGMV depends on the inverse covariance matrix. Therefore, when N is rather large in the 48IND, 50STO, and 100STO datasets, EWRF largely outperforms GMVRF, and UNC3F outperforms KZ3F too. The third reason concerns diversification. Consider the 25SBTM, 10MOM, and 25OPIN datasets. GMVRF largely outperforms EWRF, and thus, one would expect KZ3F to also outperform UNC3F. However, this is not always the case: they perform similarly for the 25OPIN

dataset, and UNC3F outperforms for 25SBTM. This result can be explained because of diversification. Specifically, in Section 2.C.6 of the Supplementary Material we derive a closed-form expression for the correlation between the out-of-sample return of SMV and that of either SGMV or EW. By calibrating this correlation to our datasets, we find that the correlation with the EW portfolio is systematically much smaller, which is in line with the empirical correlations.[20]

2.7 Conclusion: How to combine?

We conclude with a summary of insights on the portfolio combination problem. First, which portfolios to combine? The SMV portfolio is optimal without estimation risk, and thus is a natural choice. Our results favor combining it with the EW portfolio over the SGMV portfolio because EW has lower turnover, suffers no estimation risk, offers larger diversification gains, and leads to portfolio combinations that perform largely better when N/T is sizable. Moreover, investing in three funds — the sample tangent portfolio, the EW portfolio, and the risk-free asset — hits the sweet spot as it outperforms two-fund and four-fund strategies.

Second, how to estimate the combination coefficients? The constrained coefficients are less sensitive to estimation errors than the unconstrained ones, and thus outperform for a range of risk-aversion coefficients γ. However, the constrained combination can severely underperform outside of this range. The shrinkage unconstrained combination offers an appealing tradeoff between the two approaches by closely matching the one that performs best for the chosen γ.

Third, what is the impact of estimation risk? When N/T is sizable it becomes even more necessary to combine SMV with a more robust portfolio, and combining with EW largely outperforms combining with SGMV. The range of γ for which the constrained combination outperforms the unconstrained one decreases with estimation risk, and indeed, the unconstrained combination is particularly superior for the two datasets of 50 and 100 individual stocks. It can also be difficult to surpass the combination of the EW portfolio and the risk-free asset when N/T is large, but this combination is largely suboptimal otherwise.

Fourth, when can we expect to outperform the combined portfolios? Both the constrained and unconstrained combinations easily outperform the SMV portfolio. While the unconstrained combination consistently adds value to the risk-free asset, the constrained combination is inferior to the risk-free asset once γ is large enough. Compared to the combination of the EW portfolio and the risk-free asset, the unconstrained and shrinkage

[20] In Section 2.C.5 of the Supplementary Material, we also derive a four-fund strategy that combines SMV, EW, and SGMV. However, we find that it is generally outperformed by UNC3F that does not invest in SGMV. This can be explained because the theoretical gain from adding SGMV is small and thus is offset by the estimation risk coming from the additional combination coefficient to estimate.

unconstrained combinations deliver large gains in most settings, but can underperform in two cases: when there is little cross-sectional variation in mean returns and when estimation risk is large. However, the shrinkage unconstrained strategy estimated with a shrinkage covariance matrix only loses little performance in such cases, and thus can be safely relied upon.

Fifth, how does the distributional assumption matter? Given that empirical data are typically not i.i.d. normal, one can wonder if we can safely exploit our portfolio combinations calibrated under this assumption. Our answer is positive because in a rather general elliptical model for asset returns, the expected out-of-sample utility loss incurred by an investor who wrongly assumes that the data is i.i.d. normal is very small. The satisfying performance of our portfolio strategies in the six empirical datasets we consider reinforces this conclusion.

Overall, the shrinkage unconstrained combination of the SMV and EW portfolios, calibrated under the i.i.d. normal assumption, offers the best all-around out-of-sample performance across different datasets, levels of estimation risk, and degrees of risk aversion.

Supplementary Material

This Supplementary Material is divided in four sections. In Section 2.A, we provide a detailed literature review. In Section 2.B, we give simulation and theoretical evidence regarding the impact of estimation errors on constrained and unconstrained combination coefficients. In Section 2.C, we provide additional theoretical results and corresponding empirical results. In Section 2.D, we detail the proofs of all theoretical results in the main body of the chapter and in this Supplementary Material.

2.A Literature review

Our work is related to three strands of the portfolio selection literature. First, we contribute to the literature on portfolio combination, which was pioneered by Kan and Zhou (2007) and extended in, e.g., Frahm and Memmel (2010), Tu and Zhou (2011), DeMiguel et al. (2013, 2015), Branger et al. (2019), Kan et al. (2021), Kazak and Pohlmeier (2022), Yuan and Zhou (2024), Kan and Wang (2023), Lassance et al. (2024a), and Bodnar et al. (2023). We consider the setting of Tu and Zhou (2011) who combine the SMV and EW portfolios to optimize expected out-of-sample utility under a convexity constraint on the combination coefficients. We derive the *unconstrained* combination and demonstrate several of its benefits relative to the constrained strategy. We also introduce a shrinkage methodology to alleviate the impact of estimation errors on the unconstrained combination coefficients, which can be applied to other types of portfolio combination considered in the above literature.

Second, our work is related to other papers that study the value of the naive EW portfolio. DeMiguel et al. (2009b) find that the SMV portfolio and 13 robust extensions cannot outperform EW consistently. Pflug et al. (2012), Yan and Zhang (2017), and Yuan and Zhou (2024) explain that EW is not so naive because it is nearly optimal under high model ambiguity, in the absence of mispricing, and under a one-factor model, respectively. Tu and Zhou (2011) and Yuan and Zhou (2024) combine the EW and SMV portfolios to optimize out-of-sample utility and Sharpe ratio, respectively. While they allow for a risk-free asset, Frahm and Memmel (2010), Simaan and Simaan (2019), and Bodnar et al. (2023) combine EW with SMV or SGMV portfolios in the setting with no risk-free asset. In this chapter, we develop two novel estimators to optimally combine the SMV and EW portfolios. Our results affirm the value of the EW portfolio because we show theoretically and empirically that our unconstrained strategy consistently outperforms the two-fund rule of Kan and Zhou (2007) that does not invest in EW. We also demonstrate that combining the SMV portfolio with EW rather than SGMV delivers larger diversification gains, lower turnover, and largely superior performance when the number of assets is sizable relative to the sample size.

Third, we contribute to the literature on the role of constraints in portfolio selection. Frost and Savarino (1988), Jagannathan and Ma (2003), and Behr et al. (2013) show the performance benefits of imposing lower and upper bound constraints on portfolio weights. Levy and Levy (2014) introduce a class of differential variance-based constraints that significantly outperforms homogeneous bound constraints. DeMiguel et al. (2009a), Yen (2016), Callot et al. (2021), and Zhao et al. (2021) show the large benefits of norm constraints. Li (2015) and Ao et al. (2019) introduce constrained linear regression approaches to estimate mean-variance portfolios. A key differentiating feature of our work is that we show the benefits of constraints for combining *sample portfolios*, instead of assets. In particular, we demonstrate in several settings that imposing a convexity constraint on combination coefficients, although unnecessary in theory when a risk-free asset is available, acts as a bound constraint that decreases their sensitivity to estimation errors. We exploit this result to construct an improved shrinkage estimator of unconstrained portfolio combination strategies.

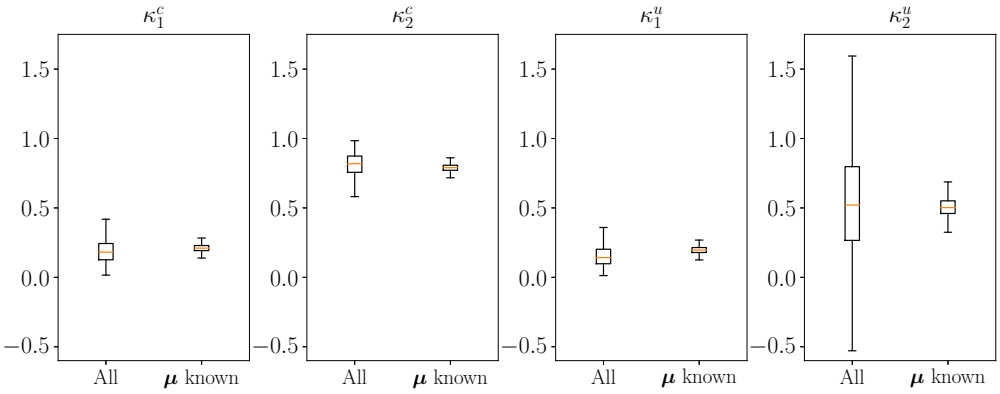
2.B Estimation errors in combination coefficients

To establish more precisely the effect of estimation errors in mean returns $\boldsymbol{\mu}$ on combination coefficients discussed in Section 2.4, consider the experiment in Figure 2.5. We calibrate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to the 25SBTM dataset and simulate 10,000 times $T = 120$ normal returns. For each simulation, we estimate the constrained and unconstrained combination coefficients with a risk-aversion coefficient $\gamma = 3$, in two different cases. First, all parameters in the coefficients are estimated. We estimate the parameters as in the simulation analysis of Section 2.4.3. Second, mean returns $\boldsymbol{\mu}$ are known, and we only plug in the sample estimate of $\boldsymbol{\Sigma}$.

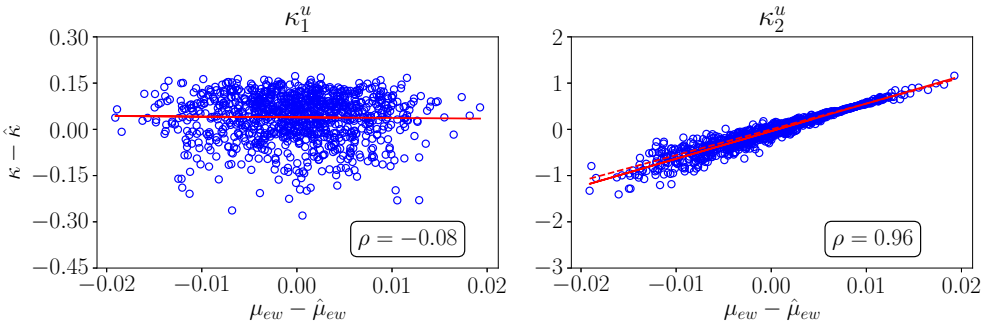
We make two main observations from Figure 2.5. First, mostly errors in $\boldsymbol{\mu}$ matter, as the boxplots are much thinner when only $\boldsymbol{\Sigma}$ is estimated. Second, the boxplots are similar and quite thin for κ_1^c, κ_2^c, and κ_1^u, because they are bounded between zero and one and estimation errors compensate in their numerator and denominator. In contrast, when $\boldsymbol{\mu}$ is estimated, the boxplot for κ_2^u is substantially wider. Given that $\kappa_2^u = \frac{1}{\gamma}\frac{\mu_{ew}}{\sigma_{ew}^2}(1-\kappa_1^u)$ while κ_1^u is moderately sensitive to $\boldsymbol{\mu}$ and errors in $\boldsymbol{\mu}$ are more impactful than errors in $\boldsymbol{\Sigma}$, this difference is mostly explained by the proportionality to $\mu_{ew} = \boldsymbol{w}_{ew}'\boldsymbol{\mu}$. Indeed, in Figure 2.5 we depict scatterplots of estimation errors in κ_1^u and κ_2^u as a function of estimation errors in μ_{ew} and we find a near-zero correlation for κ_1^u versus a near-perfect one for κ_2^u. Moreover, the best linear fit for estimation errors in κ_2^u is nearly identical to the derivative of κ_2^u with respect to μ_{ew} assuming that κ_1^u is independent of μ_{ew}.

In the next proposition, we give theoretical complement to these simulation results by deriving in closed form the estimation error in combination coefficients as a function of estimation errors in mean returns, $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$.

Figure 2.5: Estimation errors in constrained and unconstrained combination coefficients



(a) Boxplots of estimation errors in all combination coefficients



(b) Scatterplots of estimation errors in unconstrained combination coefficients

Notes. This figure depicts estimation errors in constrained combination coefficients, κ_1^c and κ_2^c in (2.13), and unconstrained combination coefficients, κ_1^u and κ_2^u in (2.14), when the risk-aversion coefficient $\gamma = 3$. The results are obtained by simulating 10,000 times $T = 120$ asset returns from a multivariate normal distribution whose moments are calibrated from a dataset of 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to December 2022. In the boxplots of Panel (a), we consider two cases to estimate the coefficients: 1) all parameters are estimated, 2) mean returns $\boldsymbol{\mu}$ are known and we only plug in the sample estimate of $\boldsymbol{\Sigma}$. In Panel (b), we depict how estimation errors in unconstrained combination coefficients depend on estimation errors in the parameter μ_{ew}. The best linear fit in each scatterplot is depicted in solid red. The dashed red line in the right plot of Panel (b) is the derivative of κ_2^u with respect to μ_{ew} assuming that κ_1^u is independent of μ_{ew}.

Proposition 2.8. *Let* $\boldsymbol{\Sigma}$ *be known and denote* $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ *and* $\bar{\varepsilon} = \boldsymbol{w}_{ew}'\boldsymbol{\varepsilon}$*. Then, the estimation errors in combination coefficients are given by*

$$\hat{\kappa}_1^u - \kappa_1^u = \frac{\psi_{ew}^2 + 2\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma_{ew} - \bar{\varepsilon}^2/\sigma_{ew}^2}{\psi_{ew}^2 + d + 2c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma_{ew} - \bar{\varepsilon}^2/\sigma_{ew}^2} - \frac{\psi_{ew}^2}{\psi_{ew}^2 + d}, \tag{2.52}$$

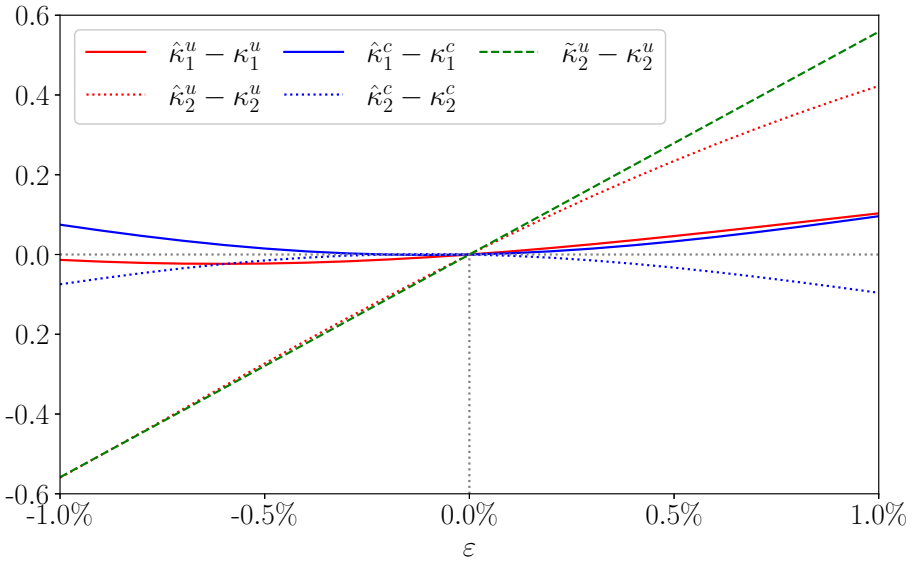
$$\hat{\kappa}_2^u - \kappa_2^u = \frac{\gamma_{ew}}{\gamma}(\kappa_1^u - \hat{\kappa}_1^u) + \frac{\bar{\varepsilon}}{\gamma\sigma_{ew}^2}(1 - \hat{\kappa}_1^u), \tag{2.53}$$

$$\hat{\kappa}_1^c - \kappa_1^c = -(\hat{\kappa}_2^c - \kappa_2^c) = \frac{\psi_{ew}^2 + \ell^2 + 2\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma}{\psi_{ew}^2 + \ell^2 + d + 2c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma} - \frac{\psi_{ew}^2 + \ell^2}{\psi_{ew}^2 + \ell^2 + d}. \tag{2.54}$$

Moreover, the estimation error for $\tilde{\kappa}_2^u = \frac{\hat{\mu}_{ew}}{\gamma\sigma_{ew}^2}(1 - \kappa_1^u)$ *is*

$$\tilde{\kappa}_2^u - \kappa_2^u = \frac{\bar{\varepsilon}d}{\gamma\sigma_{ew}^2(\psi_{ew}^2 + d)}. \tag{2.55}$$

Figure 2.6: Estimation errors in combination coefficients as a function of estimation errors in mean returns



Notes. This figure depicts the estimation errors in combination coefficients given by Proposition 2.8 when errors in sample mean returns are of the form $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} = \varepsilon \mathbf{1}_N$, as a function of ε. We use a sample size $T = 120$ and a risk-aversion coefficient $\gamma = 3$. The figure is constructed by calibrating the population vector of means and covariance matrix of asset excess returns from monthly returns on the 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to December 2022.

In Figure 2.6, we consider the 25SBTM dataset, $T = 120$ and $\gamma = 3$, as in Figure 2.5. We depict the estimation errors in combination coefficients derived in Proposition 2.8 in the specific case where $\boldsymbol{\varepsilon} = \varepsilon \mathbf{1}_N$, as a function of $\varepsilon \in [-1\%, 1\%]$. Consistent with the simulation results in Figure 2.5, it is the estimation error in $\hat{\kappa}_2^u$ that is most sensitive to ε. Moreover, the estimation errors for $\hat{\kappa}_2^u$ and $\tilde{\kappa}_2^u$ are similar, which confirms that it is mostly the proportionality to $\hat{\mu}_{ew}$ that matters.

In conclusion, the results presented in this section confirm that we can focus on the impact of estimation errors in μ_{ew} on κ_2^u to capture most of the difference in the way estimation errors impact the constrained and unconstrained combination coefficients.

2.C Additional theoretical and empirical results

In this section, we provide theoretical and empirical results that complement those in the main body of the chapter. First, we compare the exposure to three underlying funds in the constrained and unconstrained strategies. Second, we provide theoretical results for the high-dimensional asymptotic regime. Third, we provide a shrinkage estimator for the unconstrained combination of the SMV and SGMV portfolios. Fourth, we show that the convexity constraint helps explain why combining portfolios with the risk-free asset can hurt performance relative to fully invested combinations. Fifth, we derive an unconstrained four-fund strategy that combines the SMV portfolio with both the SGMV

and EW portfolios. Sixth, we compare the correlation between the out-of-sample return of the SMV portfolio and that of the SGMV and EW portfolios. Seventh, we compare the expected out-of-sample Sharpe ratio of the constrained and unconstrained strategies. Eighth, we compare the empirical performance under the normal versus the elliptical distribution. Ninth, we replicate our empirical results with a nonlinear shrinkage estimator of the covariance matrix. Tenth, we report the average monthly turnover of the considered portfolio strategies. Eleventh, we depict the empirical outperformance relative to the constrained strategy.

2.C.1 Comparison of exposure to the three funds

The portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ invests in three funds: the EW portfolio, the sample tangent portfolio, and the risk-free asset. In this section, we compare the allocation to these three funds implied by the constrained and unconstrained combination coefficients, $\boldsymbol{\kappa}^c$ in (2.13) and $\boldsymbol{\kappa}^u$ in (2.14). Note that these two combination strategies are not always different: (i) they fully invest in the SMV portfolio in the absence of estimation risk ($d \to 0$), (ii) they fully invest in the risk-free asset as $\gamma \to \infty$, and (iii) they coincide when $\gamma = \gamma_{ew}$.

Specifically, in the next proposition, we compare the expected value of the weight allocated to the three funds:[21]

$$\pi_{tan} = \frac{\gamma_{tan}}{\gamma}\kappa_1, \tag{2.56}$$

$$\pi_{ew} = \kappa_2, \tag{2.57}$$

$$\pi_{rf} = 1 - \pi_{tan} - \pi_{ew}, \tag{2.58}$$

Proposition 2.9. *The following holds regarding the allocation to the three funds in the constrained and unconstrained strategies:*

1. *The constrained strategy overinvests in the SMV portfolio:* $0 \leq \kappa_1^u \leq \kappa_1^c \leq 1$. *Similarly, for the weight on the sample tangent portfolio,* $0 \leq \pi_{tan}^u \leq \pi_{tan}^c$ *if* $\gamma_{tan} \geq 0$.[22]

2. *The constrained strategy overinvests in the EW portfolio,* $0 \leq \kappa_2^u \leq \kappa_2^c$, *if* $\gamma \in \left[\gamma_{ew}, \frac{\theta^2+d}{\mu_{ew}}\right]$, *and* $\kappa_2^u \geq \kappa_2^c$ *otherwise.*[23]

3. *Let* $\gamma_{ew} < \gamma_{tan} < (\theta^2 + d)/\mu_{ew}$.[24] *Then, the constrained strategy underexploits the risk-free asset:* $\pi_{rf}^u \leq \pi_{rf}^c \leq 0$ *if* $\gamma \leq \gamma_{ew}$ *and* $\pi_{rf}^u \geq \pi_{rf}^c$ *otherwise.*

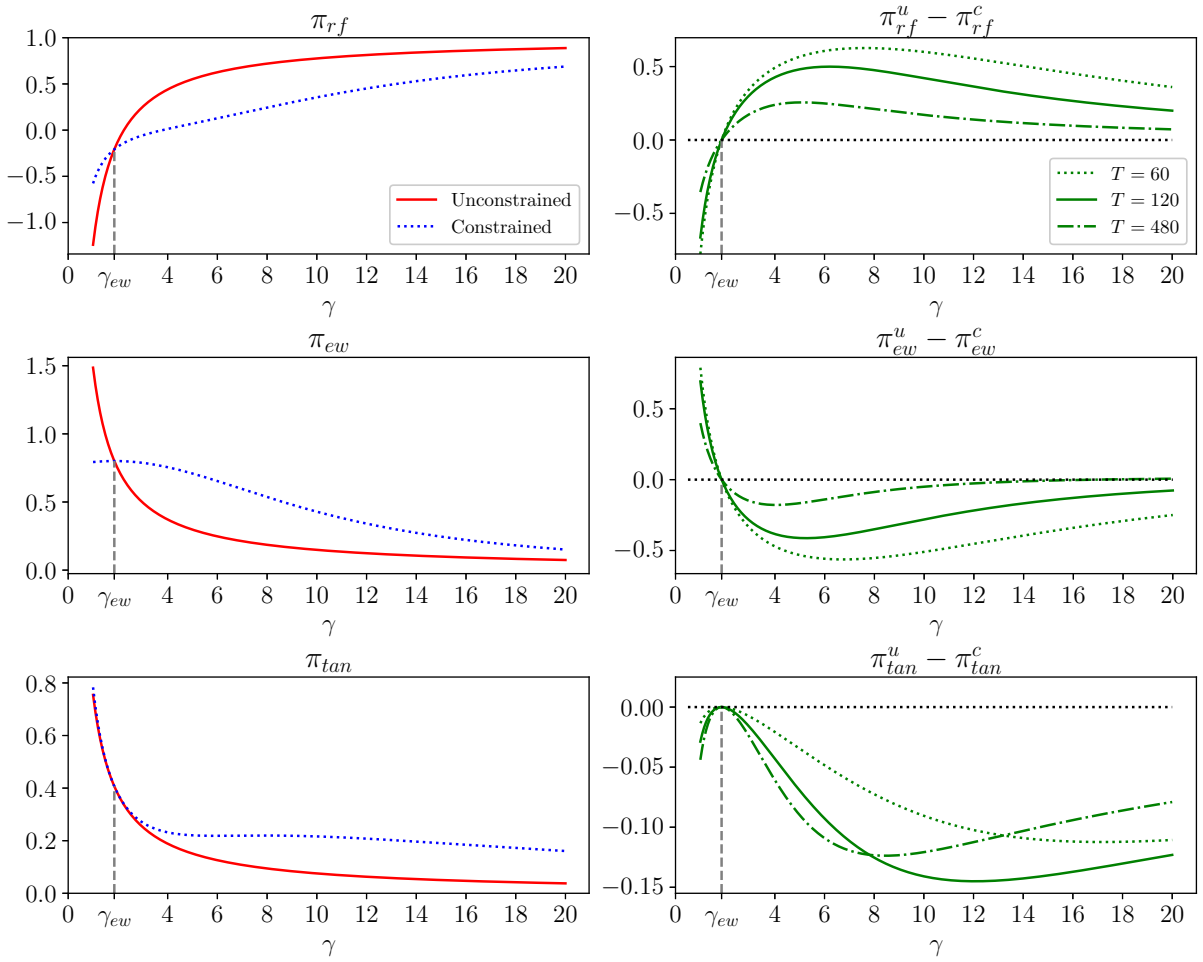
[21] We study the expected value of because the actual weights on the sample tangent portfolio and the risk-free asset depend on $\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$, which is sample-dependent.

[22] $\gamma_{tan} \geq 0$ if the GMV portfolio has a positive mean return, which is a mild assumption.

[23] This interval is typically wide. For the data used in Figure 2.7, it is equal to $[1.86, 43.06]$ when $T = 120$.

[24] This is a mild assumption fulfilled in all datasets we use in the empirical analysis of Section 2.6.

Figure 2.7: Weights on the three funds in the constrained and unconstrained strategies



Notes. This figure compares the expected weight allocated to the three funds by the constrained and unconstrained combination strategies defined in (2.13) and (2.14), respectively. The three funds (weights) are the risk-free asset (π_{rf}), the equally weighted portfolio (π_{ew}), and the sample tangent portfolio (π_{tan}). The formulas for the expected weights are given by (2.56)–(2.58). The figure is constructed by calibrating the population vector of means and covariance matrix of asset excess returns from monthly returns on the 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to December 2022. The left panels depict the weight allocated to the three funds as a function of γ when the sample size $T = 120$. The right panels depict the difference in weights between the unconstrained and constrained strategies as a function of γ when $T = 60, 120$, and 480.

We illustrate Proposition 2.9 in Figure 2.7 by calibrating the moments of asset excess returns, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, to the 25SBTM dataset. The left panels use $T = 120$, and the right panels use $T = 60, 120$, and 480. The figure shows that the allocation to the three funds in the constrained strategy is substantially different from that in the unconstrained strategy. Consider an investor with $\gamma = 5$, which is larger than $\gamma_{ew} = 1.86$. When $T = 120$, the constrained strategy commands a weight of 22% on the tangent portfolio, 71% on the EW portfolio, and 7% on the risk-free asset. In contrast, the unconstrained strategy invests much more in the risk-free asset (55%) and much less in the tangent and EW portfolios (15% and 30%, respectively). When γ increases to 10, the difference is even more striking: $(\pi^u_{rf}, \pi^u_{tan}, \pi^u_{ew}) = (77\%, 8\%, 15\%)$ while $(\pi^c_{rf}, \pi^c_{tan}, \pi^c_{ew}) = (35\%, 22\%, 43\%)$.

Finally, the right panels show that the difference between the constrained and unconstrained strategies is larger when T is smaller. That is, it is when combining portfolios is

most needed because estimation errors are large that the convexity constraint hurts most.

2.C.2 High-dimensional asymptotic regime

A recent literature has extensively studied the behavior of sample portfolios in the high-dimensional asymptotic regime, i.e., when $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, where ϕ is called the concentration ratio; see Bodnar et al. (2018), Ao et al. (2019), Ledoit and Wolf (2020), Bodnar et al. (2022), Kan et al. (2024b), and Bodnar et al. (2023), among others. This asymptotic regime is useful for evaluating portfolio performance in high dimension, and also because it is often analytically more tractable than the finite-sample setting.

In the next proposition, we derive the asymptotic joint distribution of the out-of-sample mean return and variance of the portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ in (2.7), and also the asymptotic oracle constrained and unconstrained combination coefficients.

Proposition 2.10. *Let $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$ while $(\mu_{ew}, \sigma^2_{ew}, \theta^2)$ stay bounded. Then, the following holds:*

1. *The joint distribution of the out-of-sample mean return and variance of the portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ in (2.7) is*

$$\sqrt{T-N}\begin{bmatrix} \hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu} - \left(\frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew}\right) \\ \hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) - \left(\frac{\kappa_1^2}{\gamma^2}\frac{\theta^2+\phi}{1-\phi} + \kappa_2^2\sigma^2_{ew} + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew}\right) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_2, \boldsymbol{V}\right), \quad (2.59)$$

   where $\mathbf{0}_2$ is a 2×1 vector of zeros and $\boldsymbol{V}$ is a 2×2 matrix with entries

$$\boldsymbol{V}_{11} = \frac{\kappa_1^2}{\gamma^2}\theta^2(1 + 2\theta^2), \quad (2.60)$$

$$\boldsymbol{V}_{12} = \boldsymbol{V}_{21} = \frac{2\kappa_1^2}{\gamma^2}(1 + 2\theta^2)\left(\frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew}\right), \quad (2.61)$$

$$\boldsymbol{V}_{22} = \frac{2\kappa_1^2}{\gamma^2}\Bigg(\frac{\kappa_1^2}{\gamma^2}\left(\theta^4(7+\phi) + 2\theta^2(2+\phi) + \phi\right) + 2\kappa_2^2\sigma^2_{ew}(1 + \theta^2_{ew} + \theta^2) + \frac{4\kappa_1\kappa_2}{\gamma}\mu_{ew}(1 + 2\theta^2)\Bigg). \quad (2.62)$$

2. *The asymptotic oracle constrained combination coefficients $\boldsymbol{\kappa}^c$ in (2.13) are*

$$\kappa_1^c \xrightarrow{p} \bar{\kappa}_1^c = \frac{\psi^2_{ew} + \ell^2}{\psi^2_{ew} + \ell^2 + \frac{\phi}{1-\phi}(1+\theta^2)} \quad \text{and} \quad \kappa_2^c \xrightarrow{p} 1 - \bar{\kappa}_1^c. \quad (2.63)$$

3. *The asymptotic oracle unconstrained combination coefficients $\boldsymbol{\kappa}^u$ in (2.14) are*

$$\kappa_1^u \xrightarrow{p} \bar{\kappa}_1^u = \frac{\psi^2_{ew}}{\psi^2_{ew} + \frac{\phi}{1-\phi}(1+\theta^2)} \quad \text{and} \quad \kappa_2^u \xrightarrow{p} \frac{\gamma_{ew}}{\gamma}(1 - \bar{\kappa}_1^u). \quad (2.64)$$

Using the joint asymptotic distribution of the out-of-sample mean return and variance in (2.59), we can also obtain the asymptotic distribution of out-of-sample utility as

$$\sqrt{T-N}\left[U(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}))-\left(\frac{\kappa_1}{\gamma}\theta^2+\kappa_2\mu_{ew}-\frac{\gamma}{2}\left(\frac{\kappa_1^2}{\gamma^2}\frac{\theta^2+\phi}{1-\phi}+\kappa_2^2\sigma_{ew}^2+\frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew}\right)\right)\right]$$
$$\xrightarrow{d}\mathcal{N}\left(0,\boldsymbol{V}_{11}+\frac{\gamma^2}{4}\boldsymbol{V}_{22}-\gamma\boldsymbol{V}_{12}\right), \tag{2.65}$$

where

$$\boldsymbol{V}_{11}+\frac{\gamma^2}{4}\boldsymbol{V}_{22}-\gamma\boldsymbol{V}_{12}=\frac{\kappa_1^2}{2\gamma^2}\left(\kappa_1^2\left(\theta^4(7+\phi)+2\theta^2(2+\phi)+\phi\right)+2(1-2\kappa_1)\theta^2(1+2\theta^2)\right)$$
$$-\frac{2}{\gamma}\kappa_1^2(1-\kappa_1)\kappa_2\mu_{ew}(1+2\theta^2)+\kappa_1^2\kappa_2^2\sigma_{ew}^2(1+\theta_{ew}^2+\theta^2). \tag{2.66}$$

2.C.3 Shrinkage estimator of Kan and Zhou three-fund rule

In Section 2.4, we introduce a shrinkage estimator of the unconstrained combination of the SMV and EW portfolios. In this section, we derive a similar shrinkage estimator for the three-fund rule of Kan and Zhou (2007) that combines the SMV and SGMV portfolios. Specifically, we consider the three-fund portfolio combination

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa})=\kappa_1\hat{\boldsymbol{w}}^\star+\kappa_2\hat{\boldsymbol{w}}_g, \tag{2.67}$$

where $\hat{\boldsymbol{w}}_g=\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N/(\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N)$ is the SGMV portfolio and $\hat{\boldsymbol{w}}^\star$ is the SMV portfolio in (2.5). This combination is comparable to that in Tu and Zhou (2011) because, just like $\boldsymbol{w}_{ew}$, $\hat{\boldsymbol{w}}_g$ is fully invested in risky assets. Kan and Zhou (2007) consider a similar portfolio combination but with unnormalized weights for the SGMV portfolio. We use the notation

$$\mu_g=\boldsymbol{w}_g'\boldsymbol{\mu},\quad \sigma_g^2=\boldsymbol{w}_g'\boldsymbol{\Sigma}\boldsymbol{w}_g,\quad \theta_g=\mu_g/\sigma_g,\quad \text{and}\quad \psi_g^2=\theta^2-\theta_g^2\geq 0. \tag{2.68}$$

In the following proposition, we derive the EU of the portfolio combination in (2.67) as well as the unconstrained combination coefficients.

Proposition 2.11. *The following holds concerning the three-fund portfolio in* (2.67)*:*

1. *The expected out-of-sample utility is*

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}))=\frac{\kappa_1}{\gamma}\theta^2+\kappa_2\mu_g-\frac{\gamma}{2}\left(\frac{\kappa_1^2}{\gamma^2}(\theta^2+d)+c_1\kappa_2^2\sigma_g^2+\frac{2c_1\kappa_1\kappa_2}{\gamma}\mu_g\right), \tag{2.69}$$

where

$$c_1=\frac{T-2}{T-N-1}\geq 1. \tag{2.70}$$

2. *The unconstrained combination coefficients maximizing* (2.69) *are*

$$\kappa_1^u=\frac{1}{c}\frac{\psi_g^2}{\psi_g^2+N/T+\frac{2\theta_g^2}{T-N-2}}\in[0,1]\quad \textit{and}\quad \kappa_2^u=\frac{\gamma_{tan}}{\gamma}(1/c_1-\kappa_1^u). \tag{2.71}$$

3. The expected out-of-sample utility delivered by the unconstrained combinations coefficients is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)) = U^\star - \frac{d}{2\gamma}\kappa_1^u - \frac{1}{2}(c_1 - 1)\mu_g\kappa_2^u. \tag{2.72}$$

Similar to the combination with EW in Section 2.4, κ_2^u is unbounded and proportional to γ_{tan}, and thus is sensitive to estimation errors in mean returns $\boldsymbol{\mu}$ via μ_g. Therefore, in the next proposition we derive the EU of the unconstrained strategy when κ_2^u is estimated as

$$\tilde{\kappa}_2^u = \frac{\hat{\mu}_g}{\gamma\sigma_g^2}(1/c_1 - \kappa_1^u) = \frac{\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\gamma}(1/c_1 - \kappa_1^u). \tag{2.73}$$

Proposition 2.12. *Let the combination coefficient κ_2^u be estimated by $\tilde{\kappa}_2^u$ in* (2.73)*. Then, the expected out-of-sample utility of the estimated unconstrained strategy is*

$$EU(\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u)) = EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)) - \frac{c_1}{2\gamma T}(1/c_1^2 - (\kappa_1^u)^2), \tag{2.74}$$

where $EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u))$ in (2.72) *is the utility of the oracle unconstrained strategy.*

We find numerically that the EU loss in (2.74) can be substantial. To alleviate the impact of estimation errors on the performance of unconstrained strategy, we proceed as in Section 2.4 and shrink $\tilde{\kappa}_2^u$ in (2.73) toward the target $1/c_1 - \kappa_1^u$,

$$\tilde{\kappa}_2^u(\delta) = (1 - \delta)\tilde{\kappa}_2^u + \delta(1/c_1 - \kappa_1^u), \tag{2.75}$$

where δ is the shrinkage intensity. We now calibrate δ to minimize the mean squared error of $\tilde{\kappa}_2^u(\delta)$.

Proposition 2.13. *The shrinkage intensity δ minimizing the mean squared error of $\tilde{\kappa}_2^u(\delta)$ in* (2.75) *is*

$$\delta^\star = \frac{1}{1 + \sigma_g^2 T(\gamma - \gamma_{tan})^2}. \tag{2.76}$$

In Table 2.5, we compare the out-of-sample performance of the shrinkage estimator of the unconstrained three-fund rule (SUNC3F), which we compare with its non-shrunk version (UNC3F). The table shows that SUNC3F greatly improves upon UNC3F for small values of $\gamma = 1$ and 5, while they perform similarly when γ increases to 10 and 15.

2.C.4 Investing in the risk-free asset can hurt performance

We observe in our empirical results in Tables 2.3 and 2.4 that the combination of the SMV and SGMV portfolios without a risk-free asset in Kan et al. (2021) often outperforms the combination with a risk-free asset in Kan and Zhou (2007) when the risk-aversion

Table 2.5: Annualized net out-of-sample utility of shrinkage estimator of Kan and Zhou three-fund rule

	$T = 120$						$T = 240$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
Panel (a): $\gamma = 1$												
SUNC3F	14.1	25.2	-2.77	-31.6	-104	-939	19.8	29.3	28.2	0.45	-1.80	-79.7
UNC3F	11.9	25.3	-4.83	-37.3	-113	-949	19.7	29.8	28.4	-0.77	-5.61	-85.0
Panel (b): $\gamma = 5$												
SUNC3F	2.54	5.31	-0.28	-5.87	-20.8	-219	4.04	6.17	6.07	0.38	-0.18	-16.1
UNC3F	2.05	5.00	-1.10	-7.68	-23.2	-221	3.71	5.93	5.66	-0.17	-1.20	-17.4
Panel (c): $\gamma = 10$												
SUNC3F	1.41	2.68	-0.14	-3.44	-10.2	-111	2.01	3.00	2.91	-0.21	0.21	-7.91
UNC3F	1.00	2.50	-0.56	-3.85	-11.6	-113	1.84	2.96	2.83	-0.09	-0.60	-8.74
Panel (d): $\gamma = 15$												
SUNC3F	0.87	1.63	-0.24	-2.49	-6.76	-74.7	1.23	1.95	1.78	-0.22	-0.22	-5.06
UNC3F	0.66	1.66	-0.38	-2.57	-7.75	-75.6	1.22	1.97	1.88	-0.06	-0.40	-5.83

Notes. This table reports the annualized net out-of-sample utility in percentage points for two portfolio strategies derived in Section 2.C.3 when using the sample covariance matrix. The first strategy (UNC3F) is the unconstrained combination of the sample tangent portfolio, the sample global-minimum-variance portfolio, and the risk-free asset. The second strategy (SUNC3F) is the shrinkage estimator of the unconstrained combination. The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. We consider sample sizes $T = 120$ and 240, and risk-aversion coefficients $\gamma = 1, 5, 10$, and 15.

coefficient is $\gamma = 5$. In this section, we show that this can be explained due to the convexity constraint, similar to the results in Section 2.4.

Kan et al. (2021) combine the fully invested SMV and SGMV portfolios,

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) = \kappa_1 \hat{\boldsymbol{w}}_g + \kappa_2 \left(\hat{\boldsymbol{w}}_g + \hat{\boldsymbol{w}}_z\right), \tag{2.77}$$

where $\hat{\boldsymbol{w}}_g = \widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N/(\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N)$ is the SGMV portfolio and

$$\hat{\boldsymbol{w}}_z = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\mu}} - \hat{\mu}_g \mathbf{1}_N\right) \tag{2.78}$$

is a zero-cost portfolio (i.e., $\mathbf{1}_N'\hat{\boldsymbol{w}}_z = 0$). To ensure that $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ is fully invested in risky assets, the authors impose the convexity constraint $\kappa_1 + \kappa_2 = 1$, so that the combination depends on a single combination coefficient κ:

$$\hat{\boldsymbol{w}}(\kappa) = \hat{\boldsymbol{w}}_g + \kappa \hat{\boldsymbol{w}}_z. \tag{2.79}$$

Kan et al. (2021) show that the combination coefficient κ maximizing the EU is

$$\kappa^{kwz} = \frac{(T-N)(T-N-3)}{(T-2)(T-N-2)} \frac{\psi_g^2}{\psi_g^2 + (N-1)/T}. \tag{2.80}$$

Relaxing the convexity constraint would amount to combine the SMV and SGMV portfolios with the risk-free asset, which is what the three-fund rule of Kan and Zhou (2007) is designed to do. Their optimal unconstrained portfolio combination is

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz}) = \frac{\kappa_1^{kz}}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} + \frac{\kappa_2^{kz}}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N, \tag{2.81}$$

where

$$\kappa_1^{kz} = \frac{1}{c}\frac{\psi_g^2}{\psi_g^2 + N/T} \quad \text{and} \quad \kappa_2^{kz} = \mu_g(1/c - \kappa_1^{kz}). \tag{2.82}$$

In the next proposition, we derive the EU of the portfolio combinations of Kan and Zhou (2007) and Kan et al. (2021).

Proposition 2.14. *The expected out-of-sample utility of the optimal portfolio combinations of Kan and Zhou (2007) and Kan et al. (2021) are, respectively,*

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})) = U^\star - \frac{d}{2\gamma}\kappa_1^{kz} - \frac{c-1}{2\gamma}\gamma_{tan}\kappa_2^{kz}, \tag{2.83}$$

$$EU(\hat{\boldsymbol{w}}(\kappa^{kwz})) = \mu_g - \frac{\gamma}{2}c_1\sigma_g^2 + \frac{T-N-2}{T-N-1}\frac{\psi_g^2}{2\gamma}\kappa^{kwz}, \tag{2.84}$$

where c_1 is defined in (2.70).

Similar to the combination with the EW portfolio in Section 2.4, κ_2^{kz} in (2.82) is unbounded and proportional to μ_g, and therefore is more sensitive to estimation errors in mean returns $\boldsymbol{\mu}$ relative to the other coefficients. Therefore, in the next proposition we derive the EU of the three-fund rule of Kan and Zhou (2007) when κ_2^{kz} is estimated as

$$\tilde{\kappa}_2^{kz} = \hat{\mu}_g(1/c - \kappa_1^{kz}) = \frac{\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N}(1/c - \kappa_1^{kz}). \tag{2.85}$$

Moreover, we identify the range of risk aversion γ for which the fully invested rule of Kan et al. (2021) delivers a larger EU than that of the estimated Kan-Zhou three-fund rule that also invests in the risk-free asset.

Proposition 2.15. *Let the combination coefficient κ_2^{kz} be estimated by $\tilde{\kappa}_2^{kz}$ in* (2.85). *Then,*

1. *The expected out-of-sample utility of the estimated unconstrained three-fund strategy of Kan and Zhou (2007) is*

$$EU(\hat{\boldsymbol{w}}(\kappa_1^{kz}, \tilde{\kappa}_2^{kz})) = EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})) - \frac{c}{2\gamma T}(1/c^2 - (\kappa_1^{kz})^2), \tag{2.86}$$

   where $EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz}))$ in (2.83) *is the utility of the oracle unconstrained strategy.*

2. *The fully invested strategy of Kan et al. (2021) delivers a larger expected out-of-sample utility than the estimated unconstrained three-fund strategy of Kan and Zhou (2007), $EU(\hat{\boldsymbol{w}}(\kappa^{kwz})) \geq EU(\hat{\boldsymbol{w}}(\kappa_1^{kz}, \tilde{\kappa}_2^{kz}))$, if and only if $b \geq 0$ and γ belongs to the following interval:*

$$\gamma \in \left[\frac{\gamma_{tan}}{c_1} \pm \frac{\sqrt{b}}{c_1\sigma_g}\right], \tag{2.87}$$

where

$$b = \theta_g^2 + c_1\bigg(-\theta^2 + \psi_g^2\frac{T-N-2}{T-N-1}\kappa^{kwz} + d\kappa_1^{kz} + (c-1)\gamma_{tan}\kappa_2^{kz} + \frac{c}{T}(1/c^2 - (\kappa_1^{kz})^2)\bigg). \tag{2.88}$$

For the 25SBTM dataset and $T = 120$, we find that the interval is $\gamma \in [3.18 \pm 1.73]$. This explains why for $\gamma = 5$ in Tables 2.3 and 2.4 the fully invested rule of Kan et al. (2021) outperforms the three-fund rule of Kan and Zhou (2007), while the opposite happens when $\gamma = 1$, 10, and 15.

2.C.5 The optimal four-fund combination strategy

Kan and Zhou (2007) combine the SMV and SGMV portfolios, while as in Tu and Zhou (2011) we combine the SMV and EW portfolios. In this section, we derive the unconstrained four-fund portfolio that combines SMV with both SGMV and EW, and we assess empirically whether it is beneficial to include SGMV in addition to EW.

The four-fund portfolio combination is

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) = \frac{\kappa_1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} + \kappa_2\boldsymbol{w}_{ew} + \frac{\kappa_3}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N. \tag{2.89}$$

In the next proposition, we derive the optimal unconstrained combination coefficients and the resulting EU for this four-fund combination, which are novel results in the literature.

Proposition 2.16. *The following holds concerning the four-fund portfolio combination in* (2.89)*:*

1. *The expected out-of-sample utility is*

$$\begin{aligned} EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) &= \frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew} + \frac{\kappa_3}{\gamma}\gamma_{tan} \\ &- \frac{\gamma}{2}\left(c\left(\frac{\kappa_1^2}{\gamma^2}\left(\theta^2 + \frac{N}{T}\right) + \frac{\kappa_3^2}{\gamma^2}\frac{1}{\sigma_g^2} + \frac{2\kappa_1\kappa_3}{\gamma^2}\gamma_{tan}\right) + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew} + \frac{2\kappa_2\kappa_3}{\gamma}\right). \end{aligned} \tag{2.90}$$

2. *The unconstrained combination coefficients maximizing* (2.90) *are*

$$\kappa_1^u = \frac{1}{f}\left[\psi_{ew}^2 - \theta_g^2 - \frac{\theta^2}{c}\frac{\sigma_g^2}{\sigma_{ew}^2} + \left(1 + \frac{1}{c}\right)\mu_g\gamma_{ew}\right], \tag{2.91}$$

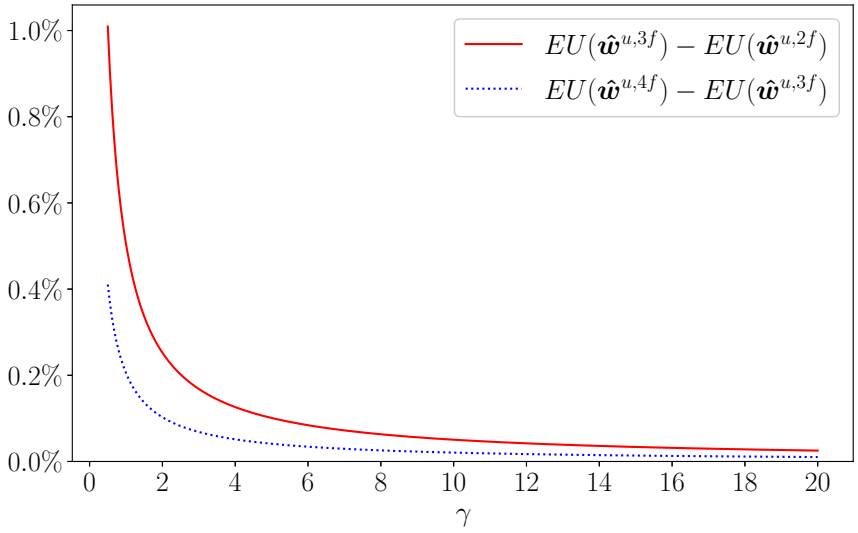
$$\kappa_2^u = \frac{\gamma_{ew}}{\gamma}\frac{1}{f}\left[\frac{N}{T}\left(c - \frac{\mu_g}{\mu_{ew}}\right) + (c-1)\psi_g^2\right], \tag{2.92}$$

$$\kappa_3^u = \frac{\mu_g}{f}\left[\frac{d}{c}\left(1 - \frac{\gamma_{ew}}{\gamma_{tan}}\right) - \left(1 - \frac{1}{c}\right)\psi_{ew}^2\right], \tag{2.93}$$

where

$$f = \psi_{ew}^2 - c\theta_g^2 + d - \frac{\sigma_g^2}{\sigma_{ew}^2}\left(\theta^2 + \frac{N}{T}\right) + 2\mu_g\gamma_{ew}. \tag{2.94}$$

Figure 2.8: Theoretical gain in expected out-of-sample utility with four-fund strategy



Notes. This figure depicts the difference between the expected out-of-sample utility of: (i) the unconstrained three-fund and two-fund rules (solid red), and (ii) the unconstrained four-fund and three-fund rules (dotted blue). The two-fund strategy is that in Kan and Zhou (2007) that combines the sample tangent portfolio with the risk-free asset. The three-fund strategy adds the equally weighted portfolio, and the four-fund strategy adds the sample global-minimum-variance portfolio. The figure is constructed by calibrating the population vector of means and covariance matrix of stock returns from monthly returns on the 25 portfolios of stocks sorted on size and book-to-market spanning July 1926 to December 2022, and using a sample size $T = 120$. The differences are depicted as a function of the risk-aversion coefficient γ between 0.5 and 20.

3. The expected out-of-sample utility of the unconstrained four-fund portfolio is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)) = U^\star - \frac{d}{2\gamma}\kappa_1^u - \frac{c-1}{2\gamma}\gamma_{tan}\kappa_3^u. \tag{2.95}$$

In Figure 2.8, we evaluate how much EU can be gained in theory by opting for the unconstrained four-fund strategy rather than the unconstrained three-fund strategy we consider in the main body of the chapter. That is, how much can be gained by adding the SGMV portfolio in the portfolio mix besides the SMV and EW portfolios. We depict this theoretical gain for the 25SBTM dataset and $T = 120$, and we compare it to the theoretical gain when going from the unconstrained two-fund strategy of Kan and Zhou (2007) to the unconstrained three-fund strategy that also invests in the EW portfolio. The figure shows that the incremental gain from adding the SGMV portfolio and relying on the four-fund portfolio is quite small relative to the gain obtained when going from two-fund to three-fund. Combined with the additional coefficient to estimate in the four-fund strategy relative to the three-fund strategy, this means that the unconstrained four-fund strategy may not outperform in practical settings.

We now evaluate the empirical performance of the unconstrained four-fund combination strategy. We compare its out-of-sample performance with that of the unconstrained three-fund strategy in the main body of the chapter that does not invest in the SGMV portfolio. We report the net out-of-sample utility of the two strategies in Table 2.6.

Table 2.6: Annualized net out-of-sample utility of three-fund and four-fund strategies

	$T = 120$						$T = 240$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
Panel (a): $\gamma = 1$												
UNC3F	29.4	22.5	6.63	2.92	15.4	10.7	20.9	25.3	24.6	6.25	31.2	30.0
UNC4F	23.2	21.1	-0.74	-10.7	-20.6	-13.8	22.4	27.4	25.1	2.83	16.2	2.89
Panel (b): $\gamma = 5$												
UNC3F	5.71	4.45	1.28	0.57	3.07	2.12	3.96	5.02	4.90	1.25	6.24	5.99
UNC4F	4.39	4.14	-0.25	-2.21	-4.24	-2.92	4.26	5.45	4.97	0.55	3.20	0.51
Panel (c): $\gamma = 10$												
UNC3F	2.84	2.22	0.64	0.28	1.53	1.06	1.97	2.51	2.45	0.62	3.12	3.00
UNC4F	2.18	2.07	-0.13	-1.11	-2.13	-1.47	2.11	2.72	2.48	0.28	1.60	0.25
Panel (d): $\gamma = 15$												
UNC3F	1.89	1.48	0.42	0.19	1.02	0.70	1.31	1.67	1.63	0.42	2.08	2.00
UNC4F	1.45	1.38	-0.09	-0.74	-1.42	-0.98	1.41	1.81	1.66	0.18	1.06	0.16

Notes. This table reports the annualized net out-of-sample utility in percentage points for two portfolio strategies when using the sample covariance matrix. The first strategy is the unconstrained three-fund strategy in the main body of the chapter, which combines the sample tangent portfolio, the equally weighted portfolio, and the risk-free asset. The second strategy is the unconstrained four-fund strategy in Section 2.C.5, which adds the sample global-minimum-variance portfolio. The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. We consider sample sizes $T = 120$ and 240, and risk-aversion coefficients $\gamma = 1, 5, 10$, and 15.

The table shows that the three-fund strategy outperforms the four-fund strategy, with the exception of the 25SBTM and 10MOM datasets when $T = 240$, in which case the four-fund portfolio is slightly superior. This result is consistent with the previous discussion that the theoretical gain from adding the SGMV portfolio is small and thus may disappear in practical settings due to estimation errors in the additional combination coefficient to estimate.

2.C.6 Out-of-sample return correlations

Tables 2.3 and 2.4 introduce the puzzling result that even in those cases where the GMVRF portfolio largely outperforms the EWRF portfolio, combining the SMV portfolio with the EW portfolio can deliver better performance than combining with the SGMV portfolio in Kan and Zhou (2007). Specifically, consider the 25SBTM, 10MOM, and 25OPIN datasets. GMVRF largely outperforms EWRF for these datasets, and thus, one would expect KZ3F to also outperform UNC3F. However, this is not always the case. Indeed, they perform similarly for the 25OPIN dataset, and UNC3F outperforms for 25SBTM. For instance, consider the shrinkage covariance matrix, $T = 120$, $\gamma = 1$, and the 25SBTM dataset. Then, GMVRF delivers a net utility of 14.9 and EWRF only 8.32, but still, UNC3F delivers a net utility of 37.3 while KZ3F delivers 33.3. This suggests that more favorable diversification properties arise when combining with the EW portfolio. To see this, in the next proposition we derive the correlation between the out-of-sample return of the SMV portfolio and that of either the SGMV portfolio or the EW portfolio.

Proposition 2.17. *Let* $\boldsymbol{r}_{T+1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *be the out-of-sample asset excess returns,* $\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ *the sample mean-variance portfolio,* $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$ *the equally weighted portfolio, and* $\hat{\boldsymbol{w}}_g = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N$ *the sample global-minimum-variance portfolio. Then,*

$$\mathbb{C}\text{orr}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}] = \frac{\theta_{ew}}{\sqrt{\frac{c}{T}(\theta^2(T+1)+N) + \frac{2\theta^4}{T-N-4}}} \overset{T\to\infty}{\longrightarrow} \frac{\theta_{ew}}{\theta}, \tag{2.96}$$

$$\mathbb{C}\text{orr}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}] = \frac{\theta_g(c+(c-1)\theta^2)}{\sqrt{\left(c+(c-1)\theta_g^2\right)\left(\frac{c}{T}(\theta^2(T+1)+N) + \frac{2\theta^4}{T-N-4}\right)}} \overset{T\to\infty}{\longrightarrow} \frac{\theta_g}{\theta}. \tag{2.97}$$

Assuming that θ_{ew} and θ_g are positive, it holds that $\mathbb{C}\text{orr}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}]$ is smaller than $\mathbb{C}\text{orr}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}]$ if

$$\theta_{ew} \leq \theta_g \frac{c+(c-1)\theta^2}{\sqrt{c+(c-1)\theta_g^2}} \approx \theta_g\sqrt{c}. \tag{2.98}$$

In Figure 2.9, we calibrate Equations (2.96)–(2.97) to the four datasets of characteristic and industry-sorted portfolios we use in our empirical analysis and we depict the two correlations as a function of the sample size. Moreover, we also depict the empirical value of the correlations obtained from our empirical analysis in Section 2.6 for $T = 120$ months, and we find that these empirical correlations are overall in line with the theoretical ones.[25] The figure shows indeed that, for all datasets and sample size, the correlation is much smaller between the SMV portfolio and the EW portfolio than between the SMV portfolio and the SGMV portfolio. For example, for the 25SBTM dataset, the empirical correlation between the out-of-sample returns of the SMV and EW portfolios is 0.18 when $T = 120$. In contrast, the empirical correlation between SMV and SGMV is twice larger: 0.37 for $T = 120$. Therefore, there are larger out-of-sample diversification gains from combining the SMV and EW portfolios, which is beneficial to out-of-sample performance.

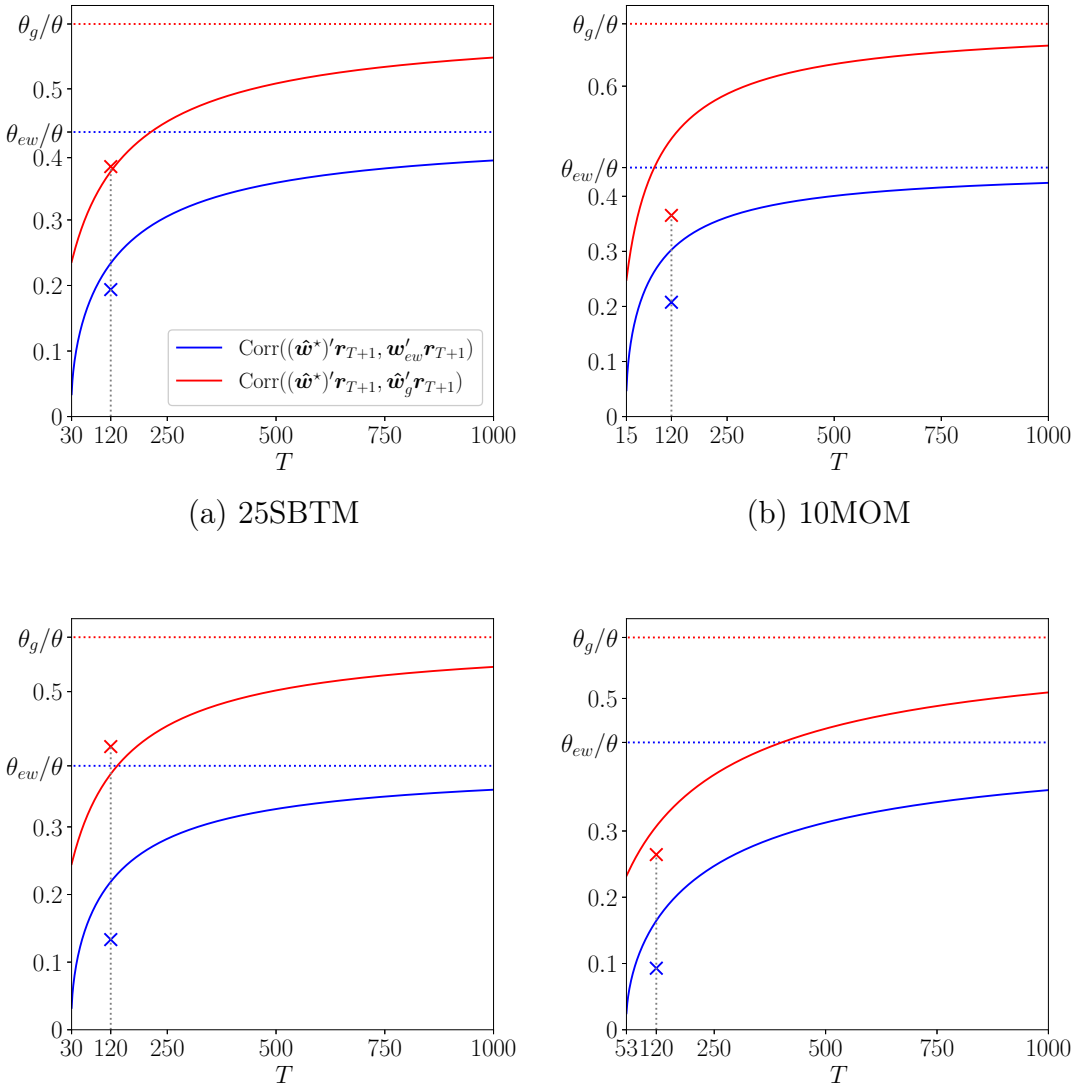
2.C.7 Expected out-of-sample Sharpe ratio

In Proposition 2.2, we show that the oracle unconstrained strategy delivers a larger EU than the oracle constrained strategy. In this section, we show moreover that the oracle unconstrained strategy delivers the largest expected out-of-sample Sharpe ratio (ESR), and thus, a larger one than the constrained strategy. However, this theoretical gain is relatively small and mostly disappears when accounting for the fact that unconstrained combination coefficients are more sensitive to estimation errors in mean returns, as discussed in Section 2.4.

Just like the maximum utility in (2.2) is unattainable in the presence of parameter uncertainty, the maximum Sharpe ratio $\theta = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$ is also unattainable. To quantify

[25] We depict the empirical correlations using net out-of-sample returns, which we find are essentially the same as those obtained using gross returns.

Figure 2.9: Out-of-sample return correlations



(a) 25SBTM (b) 10MOM

(c) 25OPIN (d) 48IND

Notes. This figure depicts the correlation between the out-of-sample return of the sample mean-variance portfolio and that of the equally weighted portfolio (in blue) and sample global-minimum-variance portfolio (in red). The correlation is depicted for four datasets listed in Table 2.1 as a function of the sample size T from $N+5$ up to 1,000 months. The solid lines depict the theoretical value of the correlations obtained from Proposition 2.17. The dotted horizontal lines depict the value as the sample size T goes to infinity. The crosses depict the empirical value of the correlations, which we obtain from the time series of net out-of-sample portfolio returns in the empirical analysis of Section 2.6 for $T = 120$.

the impact of parameter uncertainty on the out-of-sample Sharpe ratio of an estimated portfolio $\hat{\boldsymbol{w}}$, we define the expected out-of-sample Sharpe ratio (ESR) as in DeMiguel et al. (2013):

$$ESR(\hat{\boldsymbol{w}}) = \frac{\mathbb{E}[\hat{\boldsymbol{w}}'\boldsymbol{\mu}]}{\sqrt{\mathbb{E}[\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}]}}. \tag{2.99}$$

From the proof of Proposition 2.1, we have that the ESR of the combination between the SMV and EW portfolios in (2.7) is

$$ESR(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = \frac{\frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew}}{\sqrt{\frac{\kappa_1^2}{\gamma^2}(\theta^2+d) + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew}}}. \tag{2.100}$$

In the next proposition, we show that the oracle unconstrained combination strategy $\hat{\boldsymbol{w}}^u$ delivers the maximum ESR, which is independent of the risk-aversion coefficient γ. In contrast, the oracle constrained strategy $\hat{\boldsymbol{w}}^c$ does not deliver the maximum ESR, except for two specific values of γ, and generally performs worst in ESR as γ tends to zero and infinity.

Proposition 2.18. *The following holds concerning the expected out-of-sample Sharpe ratio of combination strategies:*

1. *The oracle unconstrained strategy $\hat{\boldsymbol{w}}^u$ achieves the maximum ESR of all combination strategies, which is bounded by the Sharpe ratios of the EW and optimal mean-variance portfolios:*

$$\theta_{ew} \leq \max_{\boldsymbol{\kappa}} ESR(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = ESR(\hat{\boldsymbol{w}}^u) = \sqrt{\theta_{ew}^2 + (\theta^2 - \theta_{ew}^2)\kappa_1^u} \leq \theta. \tag{2.101}$$

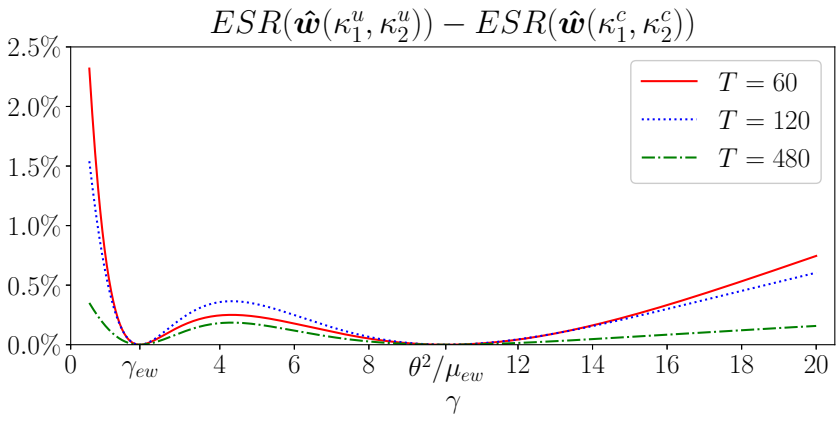
2. *The constrained combination strategy $\hat{\boldsymbol{w}}^c$ delivers the maximum ESR if and only if $\gamma = \gamma_{ew}$ or $\gamma = \theta^2/\mu_{ew}$. Moreover, if and only if $d > \frac{\theta^2}{\theta_{ew}^2}(\theta - 3\theta_{ew})(\theta - \theta_{ew})$, $\hat{\boldsymbol{w}}^c$ achieves its lowest ESR when $\gamma = 0$ or ∞:*[26]

$$\min_{\gamma} ESR(\hat{\boldsymbol{w}}^c) = \lim_{\gamma\to 0} ESR(\hat{\boldsymbol{w}}^c) = \lim_{\gamma\to\infty} ESR(\hat{\boldsymbol{w}}^c) = \frac{\theta^2}{\sqrt{\theta^2+d}}. \tag{2.102}$$

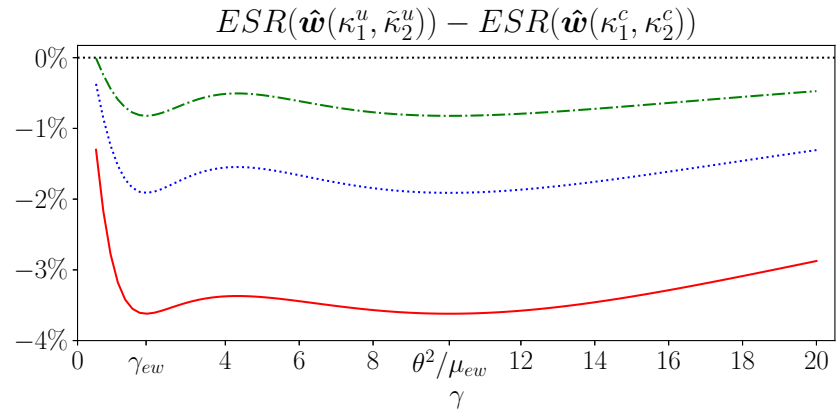
The intuition behind part 1 of Proposition 2.18 is similar to the classical result from portfolio theory that any portfolio maximizing utility in (2.2) also achieves the maximum Sharpe ratio. Similarly, any unconstrained portfolio combination maximizing the EU also achieves the maximum ESR. However, it is no longer the case under the convexity constraint (2.10), apart from two specific values of γ, and the loss in ESR is generally largest for investors with small and large degrees of risk aversion.

We illustrate Proposition 2.18 in Panel (a) of Figure 2.10 for the 25SBTM dataset. We depict the difference between the ESR of the unconstrained strategy $\hat{\boldsymbol{w}}^u$ and that of the constrained strategy $\hat{\boldsymbol{w}}^c$ as a function of γ for a samples size $T = 60$, 120, and 480. The

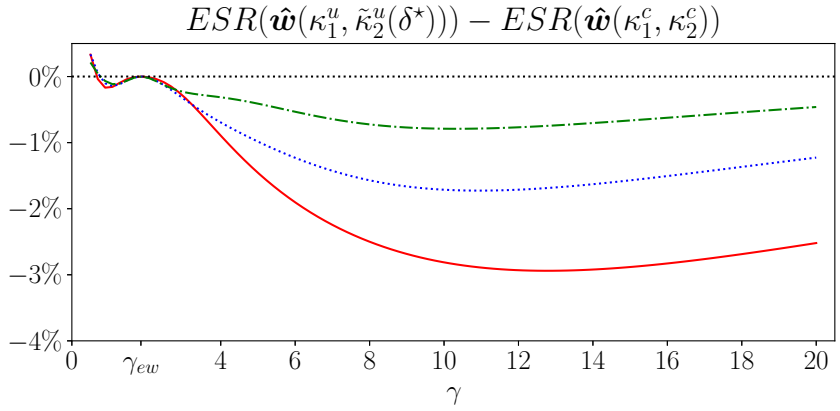
Figure 2.10: Expected out-of-sample Sharpe ratio of constrained and unconstrained strategies



(a) Known combination coefficients



(b) κ_2^u is estimated by $\tilde{\kappa}_2^u$



(c) κ_2^u is estimated by $\tilde{\kappa}_2^u(\delta^\star)$

Notes. This figure depicts the difference between the expected out-of-sample Sharpe ratio (ESR) of the unconstrained combination strategy $\hat{\boldsymbol{w}}^u$ and that of the constrained combination strategy $\hat{\boldsymbol{w}}^c$ in three different settings. In panel (a), all combination coefficients are known without errors. In panel (b), κ_2^u is estimated by $\tilde{\kappa}_2^u$ in (2.24). In panel (c), κ_2^u is estimated by $\tilde{\kappa}_2^u(\delta^\star)$ in (2.29). The figure is constructed by calibrating the population vector of means and covariance matrix of asset excess returns from monthly returns on the 25 portfolios of stocks sorted on size and book-to-market spanning July 1926 to December 2022. The ESR difference is depicted as a function of the risk-aversion coefficient γ between 0.5 and 20 for a sample size $T = 60$ (red solid), $T = 120$ (blue dotted), and $T = 480$ (green dash-dotted).

figure shows that the difference is typically not large, expect for small and large values of γ. Nonetheless, it is always positive and gets larger as estimation risk, N/T, increases.

The consistent outperformance in ESR delivered by the unconstrained strategy in Proposition 2.18 assumes however that combination coefficients are known. As discussed in Section 2.4, the main difference in estimation errors between the constrained and unconstrained coefficients is that while the constrained ones are bounded, κ_2^u is very sensitive to errors in mean returns because it is unbounded and proportional to μ_{ew}. When κ_2^u is estimated by $\tilde{\kappa}_2^u$ in (2.24), we have from the proof of Proposition 2.5 that the ESR of the estimated unconstrained strategy becomes

$$ESR(\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u)) = \frac{\mathbb{E}[(\hat{\boldsymbol{w}}^u)'\boldsymbol{\mu}]}{\sqrt{\mathbb{E}\left[(\hat{\boldsymbol{w}}^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^u\right] + \frac{1-(\kappa_1^u)^2}{\gamma^2 T}}}, \tag{2.103}$$

where

$$\mathbb{E}[(\hat{\boldsymbol{w}}^u)'\boldsymbol{\mu}] = \frac{\kappa_1^u}{\gamma}\theta^2 + \kappa_2^u \mu_{ew}, \tag{2.104}$$

$$\mathbb{E}\left[(\hat{\boldsymbol{w}}^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^u\right] = \frac{(\kappa_1^u)^2}{\gamma^2}(\theta^2 + d) + (\kappa_2^u)^2\sigma_{ew}^2 + \frac{2\kappa_1^u\kappa_2^u}{\gamma}\mu_{ew}, \tag{2.105}$$

are the expected out-of-sample mean return and variance of the unconstrained strategy when κ_2^u is known, respectively. We can also obtain the ESR of the shrinkage estimator of the unconstrained strategy proposed in Section 2.4 by plugging $\delta^\star$ in (2.28) into (2.151)–(2.152).

In panels b) and c) of Figure 2.10, we depict the difference between the ESR of the two unconstrained strategies with that of the constrained strategy. When the unconstrained strategy is not shrunk, panel b) shows that the theoretical gain in ESR completely disappears. When the unconstrained strategy is shrunk, panel c) shows that the ESR difference is reduced and is negligible around $\gamma = \gamma_{ew}$.

Finally, we report the empirical out-of-sample Sharpe ratio of the 10 portfolio strategies we consider in the main body of the chapter, which are listed in Table 2.2. We follow the methodology of Section 2.6 and define the annualized out-of-sample net Sharpe ratio of portfolio strategy k as

$$SR_k = \sqrt{12} \times \frac{\hat{\mu}_k}{\hat{\sigma}_k}, \tag{2.106}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the sample mean and standard deviation of the out-of-sample portfolio returns net of proportional transaction costs of 10 basis points for strategy k.

Table 2.7 reports the annualized net out-of-sample Sharpe ratios. We only consider the sample covariance matrix for conciseness; the conclusions are robust to using the shrinkage covariance matrices of Ledoit and Wolf (2004, 2020). In line with the theoretical predictions above, the constrained strategy CON3F outperforms the unconstrained strategy UNC3F

[26] Because $d > 0$, a sufficient condition for $d > \frac{\theta^2}{\theta_{ew}^2}(\theta - 3\theta_{ew})(\theta - \theta_{ew})$ is $\theta_{ew} > \theta/3$, which often holds.

Table 2.7: Annualized net out-of-sample Sharpe Ratio (sample covariance matrix)

	$T = 120$						$T = 240$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
$\hat{\gamma}_{ew}^{med}$	3.40	2.91	3.28	3.40	4.51	4.51	2.95	2.76	3.23	3.42	4.69	4.69
Panel (a): $\gamma = 1$												
UNC3F	0.88	0.77	0.54	0.39	0.61***	0.56***	0.85	0.80	0.76	0.40	0.80***	0.78***
SUNC3F	0.88	0.77	0.54	0.36	0.57***	0.51***	0.84	0.79	0.75	0.36	0.78***	0.77***
CON3F	0.87	0.77	0.53	0.35	0.43	0.36	0.84	0.80	0.72	0.28	0.72	0.72
KZ2F	0.83	0.75	0.47	0.20	0.15	-0.13	0.82	0.79	0.68	0.11	0.44	0.26
KZ3F	0.87	0.82	0.61	0.28	0.47	0.05	0.86	0.84	0.84	0.37	0.80	0.66
KWZ	0.61	0.65	0.39	0.26	0.25	0.08	0.64	0.74	0.56	0.21	0.72	0.65
BOP	0.61	0.55	0.28	0.01	0.22	0.12	0.63	0.70	0.52	0.13	0.04	-0.02
EWRF	0.46	0.41	0.31	0.35	0.62	0.63	0.40	0.31	0.43	0.45	0.75	0.75
GMVRF	0.70	0.55	0.55	0.23	0.48	0.08	0.68	0.49	0.77	0.36	0.80	0.67
SMV	0.81	0.79	0.53	0.19	0.17	-0.12	0.86	0.83	0.71	0.15	0.56	0.39
Panel (b): $\gamma = 5$												
UNC3F	0.88	0.77	0.54*	0.39**	0.61***	0.56***	0.84	0.80***	0.76**	0.40***	0.80***	0.78***
SUNC3F	0.89	0.78	0.58*	0.46	0.69***	0.65***	0.85	0.80***	0.77**	0.44***	0.83***	0.83***
CON3F	0.90	0.80	0.65	0.54	0.78	0.75	0.88	0.85	0.82	0.53	0.87	0.87
KZ2F	0.83	0.75	0.47	0.20	0.15	-0.13	0.82	0.79	0.68	0.11	0.44	0.26
KZ3F	0.87	0.82	0.61	0.28	0.47	0.05	0.86	0.84	0.84	0.37	0.80	0.67
KWZ	0.82	0.83	0.66	0.40	0.49	0.16	0.89	0.94	0.91	0.40	0.84	0.71
BOP	0.79	0.74	0.58	0.36	0.60	0.55	0.84	0.90	0.81	0.44	0.53	0.38
EWRF	0.46	0.40	0.31	0.35	0.62	0.63	0.40	0.31	0.43	0.45	0.75	0.75
GMVRF	0.70	0.55	0.55	0.23	0.48	0.08	0.68	0.49	0.77	0.36	0.80	0.67
SMV	0.81	0.79	0.53	0.19	0.17	-0.11	0.85	0.83	0.71	0.15	0.56	0.39
Panel (c): $\gamma = 10$												
UNC3F	0.88	0.77	0.54*	0.39*	0.61***	0.56***	0.84	0.80***	0.76*	0.40***	0.80***	0.78***
SUNC3F	0.88	0.78	0.55*	0.39*	0.62***	0.57***	0.84	0.80***	0.76*	0.41**	0.79***	0.78***
CON3F	0.86	0.80	0.66	0.52	0.77	0.77	0.89	0.85	0.82	0.49	0.87	0.87
KZ2F	0.83	0.75	0.47	0.20	0.15	-0.13	0.82	0.79	0.68	0.11	0.44	0.26
KZ3F	0.87	0.82	0.61	0.28	0.47	0.05	0.86	0.84	0.84	0.37	0.80	0.67
KWZ	0.83	0.82	0.71	0.40	0.51	0.17	0.95	0.95	0.95	0.41	0.85	0.71
BOP	0.80	0.77	0.66	0.49	0.74	0.70	0.93	0.93	0.88	0.52	0.80	0.66
EWRF	0.46	0.40	0.31	0.35	0.62	0.63	0.40	0.31	0.43	0.45	0.75	0.75
GMVRF	0.70	0.55	0.55	0.23	0.48	0.08	0.68	0.49	0.77	0.36	0.80	0.67
SMV	0.81	0.79	0.53	0.19	0.17	-0.11	0.85	0.83	0.71	0.15	0.56	0.39
Panel (d): $\gamma = 15$												
UNC3F	0.88	0.77	0.54*	0.39	0.61***	0.56***	0.84	0.80***	0.76	0.40	0.80***	0.78***
SUNC3F	0.88	0.78	0.55*	0.39	0.62***	0.56***	0.84	0.80***	0.76	0.41	0.80***	0.78***
CON3F	0.85	0.80	0.65	0.49	0.71	0.76	0.89	0.85	0.81	0.43	0.84	0.86
KZ2F	0.83	0.75	0.47	0.20	0.15	-0.13	0.82	0.79	0.68	0.11	0.44	0.26
KZ3F	0.87	0.82	0.61	0.28	0.47	0.05	0.86	0.84	0.84	0.37	0.80	0.67
KWZ	0.80	0.78	0.70	0.39	0.51	0.17	0.93	0.91	0.95	0.40	0.85	0.71
BOP	0.81	0.76	0.69	0.53	0.76	0.75	0.96	0.91	0.91	0.52	0.87	0.80
EWRF	0.46	0.40	0.31	0.35	0.62	0.63	0.40	0.31	0.43	0.45	0.75	0.75
GMVRF	0.70	0.55	0.55	0.23	0.48	0.08	0.68	0.49	0.77	0.36	0.80	0.67
SMV	0.81	0.79	0.53	0.19	0.17	-0.11	0.85	0.83	0.71	0.15	0.56	0.39

Notes. This table reports the annualized net out-of-sample Sharpe ratio for the portfolio strategies in Table 2.2 and the datasets in Table 2.1 when using the sample covariance matrix. The net out-of-sample Sharpe ratio is computed using proportional transaction costs of 10 basis points. For 50 and 100STO, we report the average performance across 500 randomly drawn datasets from a total pool of 235 stocks. We evaluate the statistical significance of the difference in net Sharpe ratios of UNC3F vs CON3F and SUNC3F vs CON3F using the stationary block bootstrap approach of Politis and Romano (1994) and Ledoit and Wolf (2008) to produce p-values. The stars *, **, and *** indicate that the p-value is less than 10%, 5%, and 1%, respectively.

in most cases, although the difference is small, and the shrinkage unconstrained strategy SUNC3F approaches the Sharpe ratio of the constrained strategy more closely.

2.C.8 Empirical performance under the elliptical distribution

In Table 2.8, we report the annualized net out-of-sample utility of the unconstrained and constrained strategies when the combination coefficients are calibrated in three different ways. First, assuming normality and a finite sample, as in part 2 of Proposition 2.1 (UNC3F, CON3F). Second, assuming normality and the high-dimensional asymptotic regime, as in Equations (2.40)–(2.41) (AsyUNC3F, AsyCON3F). Third, assuming a t distribution and the high-dimensional asymptotic regime, as in Equation (2.39) and parts 2 and 3 of Proposition 2.7 (EllUNC3F, EllCON3F). In the latter case, we estimate the number of degrees of freedom ν via maximum likelihood. We only report the results for a sample size $T = 120$ for conciseness, and we consider both the sample estimate of the covariance matrix and the shrinkage estimate of Ledoit and Wolf (2004).

We make two main observations from the results in Table 2.8. First, AsyUNC3F and AsyCON3F underperform UNC3F and CON3F. This is expected because the high-dimensional asymptotic regime only provides an approximation to the finite-sample EU. However, the difference in performance is small, which suggests that this approximation is accurate. Second, UNC3F and CON3F generally underperform EllUNC3F and EllCON3F. This is expected because the latter two account for the effect of fat tails and also allocate a smaller weight to the SMV portfolio, which reduces turnover and thus transaction costs. However, the difference in performance is small, which is consistent with the theoretical results presented in Section 2.5.2. For the 100STO dataset, we even observe that UNC3F and CON3F outperform EllUNC3F and EllCON3F, which can be explained because SMV performs very poorly and has a very large turnover when $T = 120$ and $N = 100$, and thus the extra variation of the coefficient κ_1 on the SMV portfolio caused by the estimation of the number of degrees of freedom ν hurts the net out-of-sample utility.

2.C.9 Nonlinear shrinkage of the covariance matrix

In Table 2.9, we replicate Tables 2.3 and 2.4 in the main body of the chapter using the nonlinear shrinkage estimator of the covariance matrix of Ledoit and Wolf (2020) to estimate the SMV portfolio. The results are overall similar to those obtained using the linear shrinkage estimator of Ledoit and Wolf (2004) in Table 2.4, and therefore the conclusions drawn in the main body of the chapter are robust to using a nonlinear shrinkage estimator.

Table 2.8: Annualized net out-of-sample utility (elliptical and asymptotic coefficients)

	Sample $\boldsymbol{\Sigma}$						Shrinkage $\boldsymbol{\Sigma}$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
Panel (a): $\gamma = 1$												
UNC3F	29.4	22.5	6.63	2.92	15.4	10.7	37.3	31.6	14.3	3.70	15.9	16.7
AsyUNC3F	28.7	22.1	6.24	2.69	14.6	8.88	37.3	31.5	14.1	3.49	14.9	16.7
EllUNC3F	31.5	24.7	8.32	3.08	15.4	1.85	37.9	32.8	16.0	3.89	15.6	16.1
CON3F	31.2	23.4	8.31	5.36	8.59	4.95	42.2	34.5	18.8	6.93	11.5	13.2
AsyCON3F	30.5	22.9	7.76	5.03	8.88	2.40	42.1	34.3	18.4	6.62	11.4	13.1
EllCON3F	33.4	25.8	10.0	5.43	9.65	-5.96	42.5	35.7	20.3	7.03	12.0	12.2
Panel (b): $\gamma = 5$												
UNC3F	5.71	4.45	1.28	0.57	3.07	2.12	7.44	6.31	2.83	0.73	3.18	3.34
AsyUNC3F	5.56	4.37	1.20	0.52	2.91	1.74	7.43	6.29	2.79	0.68	2.98	3.34
EllUNC3F	6.16	4.91	1.63	0.60	3.06	0.14	7.55	6.54	3.16	0.77	3.11	3.22
CON3F	5.08	4.41	2.43	1.67	5.67	5.02	8.05	6.96	4.15	1.85	5.88	6.34
AsyCON3F	4.91	4.32	2.33	1.60	5.50	4.61	8.04	6.93	4.09	1.79	5.65	6.35
EllCON3F	5.59	4.81	2.75	1.72	5.69	2.95	8.03	7.10	4.43	1.89	5.79	6.21
Panel (c): $\gamma = 10$												
UNC3F	2.84	2.22	0.64	0.28	1.53	1.06	3.72	3.15	1.41	0.36	1.59	1.67
AsyUNC3F	2.77	2.18	0.60	0.26	1.45	0.87	3.71	3.14	1.39	0.34	1.49	1.67
EllUNC3F	3.07	2.45	0.81	0.30	1.53	0.05	3.77	3.27	1.58	0.38	1.55	1.61
CON3F	-0.74	1.06	-1.53	-4.01	-1.59	-2.80	1.98	2.76	-0.46	-3.96	-1.44	-1.94
AsyCON3F	-0.83	1.02	-1.58	-4.06	-1.85	-3.00	1.99	2.75	-0.48	-4.01	-1.73	-1.83
EllCON3F	-0.48	1.18	-1.38	-3.93	-1.70	-3.92	1.91	2.75	-0.35	-3.87	-1.67	-1.91
Panel (d): $\gamma = 15$												
UNC3F	1.89	1.48	0.42	0.19	1.02	0.70	2.48	2.10	0.94	0.24	1.06	1.11
AsyUNC3F	1.84	1.45	0.40	0.17	0.97	0.58	2.48	2.10	0.93	0.23	0.99	1.11
EllUNC3F	2.05	1.63	0.54	0.20	1.02	0.03	2.51	2.18	1.05	0.25	1.04	1.07
CON3F	-2.93	-0.11	-3.55	-7.85	-6.76	-10.5	-0.14	1.34	-2.57	-8.04	-6.62	-9.49
AsyCON3F	-2.98	-0.12	-3.57	-7.90	-6.96	-10.7	-0.12	1.34	-2.57	-8.09	-6.88	-9.23
EllCON3F	-2.81	-0.08	-3.54	-7.78	-6.84	-11.6	-0.28	1.30	-2.61	-7.92	-6.90	-9.33

Notes. This table reports the annualized net out-of-sample utility in percentage points for the datasets in Table 2.1 and the constrained unconstrained and constrained portfolio strategies when the combination coefficients are estimated in three different ways. First, assuming normality and a finite sample, as in part 2 of Proposition 2.1 (UNC3F, CON3F). Second, assuming normality and the high-dimensional asymptotic regime, as in Equations (2.40)–(2.41) (AsyUNC3F, AsyCON3F). Third, assuming a t distribution and the high-dimensional asymptotic regime, as in Equation (2.39) and parts 2 and 3 of Proposition 2.7 (EllUNC3F, EllCON3F). In the latter case, we estimate the number of degrees of freedom ν via maximum likelihood. The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. We report the results for the sample estimate of the covariance matrix and the shrinkage estimate of Ledoit and Wolf (2004). For 50 and 100STO, we report the average performance across 500 randomly drawn datasets from a total pool of 235 stocks. We consider a sample size $T = 120$ and risk-aversion coefficients $\gamma = 1$, 5, 10, and 15.

Table 2.9: Annualized net out-of-sample utility (nonlinear shrinkage covariance matrix)

	$T = 120$						$T = 240$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
$\hat{\gamma}_{ew}^{med}$	3.40	2.91	3.28	3.40	4.51	4.51	2.95	2.76	3.23	3.42	4.69	4.69
Panel (a): $\gamma = 1$												
UNC3F	36.5	24.9	14.6	3.23	15.9***	16.7***	21.1	25.9	25.1	3.85	30.3***	30.4***
SUNC3F	37.4	24.9	16.0	4.09	14.7***	15.5***	21.0	25.8	24.5	2.73	29.7***	29.8***
CON3F	39.8	26.5	18.4	6.51	11.3	13.0	22.4	27.4	21.9	0.30	23.9	22.0
KZ2F	36.1	24.7	14.6	0.71	-0.31	-0.61	19.9	25.8	17.6	-6.52	11.3	5.98
KZ3F	28.6	28.5	12.8	-2.70	-2.93	11.7	20.6	30.5	30.9	3.19	11.3	8.54
KWZ	23.4	19.0	10.3	4.45	3.60	4.57	17.6	24.8	16.8	-2.08	9.93	9.99
BOP	16.9	7.07	-1.22	-18.5	-31.8	-25.7	15.6	20.9	13.6	-5.68	-51.0	-102
EWRF	8.32	5.72	-0.59	2.85	17.5	17.5	5.99	2.47	8.93	9.87	27.4	27.7
GMVRF	12.5	9.20	9.17	-3.26	-1.07	12.8	12.4	7.75	28.3	9.93	11.5	8.50
SMV	-127	-19.0	-140	-503	-548	-1479	-51.0	8.42	-66.8	-237	-169	-523
Panel (b): $\gamma = 5$												
UNC3F	7.17	4.93	2.88	0.64	3.18***	3.33***	3.97	5.13**	5.00	0.77	6.07***	6.09***
SUNC3F	7.39	5.24	3.54	1.43	4.44***	4.84***	4.02	5.25**	5.09	0.81	6.59***	6.77***
CON3F	6.82	4.94	4.20	1.84	5.89	6.32	4.03	6.00	5.63	1.41	7.17	7.38
KZ2F	7.11	4.90	2.88	0.13	-0.07	-0.12	3.74	5.12	3.48	-1.31	2.26	1.20
KZ3F	5.55	5.64	2.48	-0.58	-0.65	2.34	3.90	6.06	6.16	0.63	2.23	1.66
KWZ	7.78	6.10	5.37	3.31	3.75	3.22	7.52	8.33	8.58	1.41	6.96	6.89
BOP	4.27	3.06	1.76	-3.09	-3.90	-2.94	5.31	7.23	6.25	-0.51	-5.76	-20.2
EWRF	1.66	1.14	-0.12	0.57	3.49	3.49	1.20	0.49	1.79	1.97	5.48	5.55
GMVRF	2.39	1.82	1.79	-0.69	-0.27	2.55	2.43	1.54	5.64	1.98	2.26	1.65
SMV	-27.6	-4.08	-29.2	-104	-112	-312	-11.2	1.57	-13.8	-48.1	-34.2	-107
Panel (c): $\gamma = 10$												
UNC3F	3.58*	2.46	1.44	0.32**	1.59***	1.67***	1.97	2.56	2.50	0.38**	3.03***	3.05***
SUNC3F	3.46**	2.45	1.29	0.07**	1.33***	1.32***	1.92	2.54	2.50	0.35**	2.95***	2.95***
CON3F	0.19	1.29	-0.54	-3.84	-1.40	-1.93	0.21	2.26	0.99	-3.41	-0.98	-1.34
KZ2F	3.55	2.45	1.44	0.07	-0.03	-0.06	1.86	2.56	1.74	-0.65	1.13	0.60
KZ3F	2.76	2.82	1.23	-0.29	-0.33	1.17	1.93	3.03	3.08	0.31	1.11	0.82
KWZ	1.72	-0.60	0.41	-0.22	0.16	-0.36	2.54	1.77	3.15	-1.34	2.27	2.43
BOP	-2.94	-2.66	-2.72	-6.31	-5.70	-7.81	0.30	1.08	1.07	-4.04	-2.65	-12.0
EWRF	0.83	0.57	-0.06	0.28	1.75	1.75	0.60	0.25	0.89	0.99	2.74	2.77
GMVRF	1.19	0.91	0.89	-0.35	-0.14	1.27	1.21	0.77	2.82	0.99	1.13	0.82
SMV	-13.9	-2.05	-14.7	-52.4	-56.2	-157	-5.64	0.78	-6.91	-24.1	-17.1	-53.6
Panel (d): $\gamma = 15$												
UNC3F	2.38***	1.64	0.96***	0.21***	1.06***	1.11***	1.31*	1.71	1.66**	0.26***	2.02***	2.03***
SUNC3F	2.32***	1.62	0.91***	0.13***	0.97***	0.99***	1.30**	1.70	1.67**	0.25***	2.00***	2.00***
CON3F	-2.30	0.03	-2.84	-7.88	-6.60	-9.44	-1.17	1.06	-0.90	-5.93	-4.72	-7.44
KZ2F	2.36	1.63	0.96	0.04	-0.02	-0.04	1.23	1.70	1.16	-0.44	0.75	0.40
KZ3F	1.84	1.88	0.82	-0.20	-0.22	0.78	1.29	2.02	2.05	0.21	0.74	0.55
KWZ	-3.32	-6.59	-4.45	-3.88	-3.70	-4.07	-1.89	-3.77	-1.92	-4.59	-2.48	-2.06
BOP	-8.23	-8.18	-7.30	-11.2	-10.1	-15.1	-3.93	-4.17	-3.67	-8.03	-6.33	-12.2
EWRF	0.55	0.38	-0.04	0.19	1.16	1.16	0.40	0.16	0.60	0.66	1.83	1.85
GMVRF	0.79	0.61	0.60	-0.23	-0.09	0.85	0.81	0.51	1.88	0.66	0.75	0.55
SMV	-9.31	-1.37	-9.79	-35.0	-37.5	-105	-3.77	0.52	-4.61	-16.1	-11.4	-35.8

Notes. This table reports the annualized net out-of-sample utility in percentage points for the portfolio strategies in Table 2.2 and the datasets in Table 2.1 when using the nonlinear shrinkage covariance matrix of Ledoit and Wolf (2020). The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. For 50 and 100STO, we report the average performance across 500 randomly drawn datasets from a total pool of 235 stocks. We evaluate the statistical significance of the difference in net utilities of UNC3F vs CON3F and SUNC3F vs CON3F using the stationary block bootstrap approach of Politis and Romano (1994) and Ledoit and Wolf (2008) to produce p-values. The stars *, **, and *** indicate that the p-value is less than 10%, 5%, and 1%, respectively. The underlined numbers indicate the strategies that do not belong to the superior set of models of Hansen et al. (2011) at a confidence level of 5%. For 50STO and 100STO, we underline the three strategies that are most often rejected, or more in case of equalities.

2.C.10 Portfolio turnover

In Table 2.10, we report the average monthly turnover of the 10 portfolio strategies we consider, when the sample size is $T = 120$. The average turnover is computed as the average of the monthly turnover values in (2.50). We observe that UNC3F systematically delivers a smaller turnover than CON3F, which is in line with Proposition 2.4. Similarly, SUNC3F also delivers a smaller turnover than CON3F.

2.C.11 Empirical outperformance relative to constrained strategy

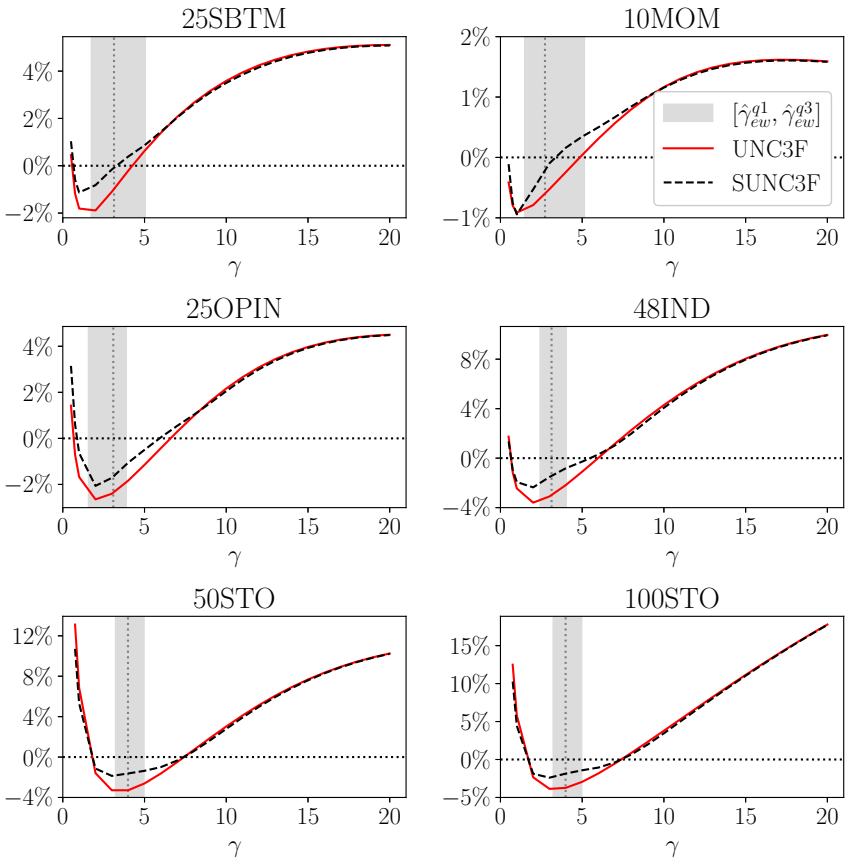
Figure 2.11 replicates the simulation results in Figure 2.3 empirically. Specifically, we depict, for $T = 120$ and the sample covariance matrix, the difference between the net utility of UNC3F and SUNC3F and the net utility of CON3F. The figure shows that SUNC3F trades off between the CON3F and UNC3F strategies. Indeed, for γ around the estimated γ_{ew}, SUNC3F outperforms UNC3F and does not lose much relative to CON3F, and when γ moves away from γ_{ew}, SUNC3F is close to UNC3F and largely outperforms CON3F.

Table 2.10: Average monthly turnover ($T = 120$)

	Sample $\boldsymbol{\Sigma}$						Shrinkage $\boldsymbol{\Sigma}$					
	25SBTM	10MOM	25OPIN	48IND	50STO	100STO	25SBTM	10MOM	25OPIN	48IND	50STO	100STO
Panel (a): $\gamma = 1$												
UNC3F	9.38	4.06	5.48	2.79	1.62	2.92	9.38	4.06	4.29	1.96	1.04	0.76
SUNC3F	9.38	4.06	5.48	2.79	1.62	2.92	9.38	4.06	4.29	1.96	1.05	0.76
CON3F	9.93	4.31	5.79	3.12	2.05	3.29	9.93	4.31	4.54	2.20	1.32	0.81
KZ2F	9.96	4.30	5.76	3.00	1.91	3.06	9.96	4.30	4.52	2.13	1.25	0.77
KZ3F	10.2	4.15	6.07	4.76	4.10	7.98	10.2	4.15	4.68	3.24	2.38	1.95
KWZ	7.90	3.87	4.60	2.69	1.08	3.51	7.90	3.87	3.67	1.84	0.65	0.64
BOP	8.55	5.23	5.24	6.40	5.92	11.0	8.55	5.23	5.24	6.40	5.92	11.0
EWRF	0.17	0.19	0.20	0.21	0.30	0.30	0.17	0.19	0.20	0.21	0.30	0.30
GMVRF	6.12	1.67	3.73	3.63	3.94	7.51	6.12	1.67	2.84	2.50	2.29	1.92
SMV	25.2	7.54	16.9	27.1	16.6	106	25.2	7.54	13.3	19.0	10.5	26.5
Panel (b): $\gamma = 5$												
UNC3F	1.88	0.81	1.10	0.56	0.32	0.58	1.88	0.81	0.86	0.39	0.21	0.15
SUNC3F	1.88	0.81	1.10	0.56	0.32	0.58	1.88	0.81	0.86	0.39	0.21	0.15
CON3F	2.10	0.91	1.20	0.65	0.35	0.61	2.10	0.91	0.94	0.46	0.23	0.16
KZ2F	1.99	0.86	1.15	0.60	0.38	0.61	1.99	0.86	0.90	0.43	0.25	0.15
KZ3F	2.04	0.83	1.21	0.95	0.82	1.60	2.04	0.83	0.94	0.65	0.48	0.39
KWZ	1.81	0.85	1.10	0.98	0.54	2.75	1.81	0.85	0.83	0.54	0.26	0.38
BOP	1.92	0.91	1.14	1.21	1.08	2.20	1.92	0.91	1.13	1.20	1.07	2.20
EWRF	0.03	0.04	0.04	0.04	0.06	0.06	0.03	0.04	0.04	0.04	0.06	0.06
GMVRF	1.22	0.33	0.75	0.73	0.79	1.50	1.22	0.33	0.57	0.50	0.46	0.38
SMV	5.03	1.51	3.37	5.42	3.31	21.2	5.03	1.51	2.67	3.80	2.10	5.31
Panel (c): $\gamma = 10$												
UNC3F	0.94	0.41	0.55	0.28	0.16	0.29	0.94	0.41	0.43	0.20	0.10	0.08
SUNC3F	0.94	0.41	0.55	0.28	0.16	0.29	0.94	0.41	0.43	0.20	0.10	0.08
CON3F	1.32	0.53	0.78	0.59	0.31	0.43	1.32	0.53	0.62	0.41	0.19	0.11
KZ2F	1.00	0.43	0.58	0.30	0.19	0.31	1.00	0.43	0.45	0.21	0.13	0.08
KZ3F	1.02	0.42	0.61	0.48	0.41	0.80	1.02	0.42	0.47	0.32	0.24	0.19
KWZ	1.16	0.51	0.75	0.85	0.50	2.71	1.16	0.51	0.53	0.43	0.23	0.36
BOP	1.24	0.53	0.74	0.51	0.38	1.08	1.24	0.53	0.72	0.51	0.38	1.08
EWRF	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.03	0.03
GMVRF	0.61	0.17	0.37	0.36	0.39	0.75	0.61	0.17	0.28	0.25	0.23	0.19
SMV	2.52	0.75	1.69	2.71	1.66	10.6	2.52	0.75	1.33	1.90	1.05	2.65
Panel (d): $\gamma = 15$												
UNC3F	0.63	0.27	0.37	0.19	0.11	0.19	0.63	0.27	0.29	0.13	0.07	0.05
SUNC3F	0.63	0.27	0.37	0.19	0.11	0.20	0.63	0.27	0.29	0.13	0.07	0.05
CON3F	1.07	0.40	0.68	0.65	0.36	0.47	1.07	0.40	0.53	0.45	0.22	0.12
KZ2F	0.66	0.29	0.38	0.20	0.13	0.20	0.66	0.29	0.30	0.14	0.08	0.05
KZ3F	0.68	0.28	0.40	0.32	0.27	0.53	0.68	0.28	0.31	0.22	0.16	0.13
KWZ	0.98	0.41	0.66	0.82	0.49	2.70	0.98	0.41	0.45	0.41	0.22	0.35
BOP	1.07	0.42	0.65	0.43	0.25	0.68	1.07	0.42	0.62	0.42	0.24	0.69
EWRF	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.02
GMVRF	0.41	0.11	0.25	0.24	0.26	0.50	0.41	0.11	0.19	0.17	0.15	0.13
SMV	1.68	0.50	1.12	1.81	1.10	7.05	1.68	0.50	0.89	1.27	0.70	1.77

Notes. This table reports the average monthly turnover for the portfolio strategies in Table 2.2 and the datasets in Table 2.1 when using either the sample covariance matrix or the shrinkage covariance matrix of Ledoit and Wolf (2004). The average turnover is computed as the average of the monthly turnover values in (2.50). We consider a sample size $T = 120$ and risk-aversion coefficients $\gamma = 1$, 5, 10, and 15.

Figure 2.11: Net annualized out-of-sample utility relative to constrained strategy



Notes. This figure depicts for all six datasets the difference between the annualized net out-of-sample utility of the unconstrained strategy (UNC3F, solid red) and the shrinkage estimator of the unconstrained strategy (SUNC3F, dashed black) relative to the constrained strategy (CON3F). The difference is depicted as a function of the risk-aversion coefficient γ. We use a sample size $T = 120$ and the sample covariance matrix. The net out-of-sample utility is computed using proportional transaction costs of 10 basis points. The gray region depicts the first, second, and third quartiles of the estimated parameter γ_{ew} defined in (2.8).

2.D Proofs

We often use the following three properties throughout the proofs. First, because returns are i.i.d. normal, the sample mean $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/T)$, the sample covariance matrix $(T - N - 2)\widehat{\boldsymbol{\Sigma}} \sim \mathcal{W}_N(T-1, \boldsymbol{\Sigma})$, and they are independent of each other. Second, the inverse sample covariance matrix is unbiased because of the $1/(T-N-2)$ coefficient in (2.4), $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}] = \boldsymbol{\Sigma}^{-1}$. Third, the expectation of a quadratic form in the random variable $\boldsymbol{x}$ is

$$\mathbb{E}[\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}] = \mathbb{E}[\boldsymbol{x}]'\boldsymbol{A}\mathbb{E}[\boldsymbol{x}] + \text{Trace}(\boldsymbol{A}\mathbb{V}[\boldsymbol{x}]), \tag{2.107}$$

where $\boldsymbol{A}$ is a constant matrix (Rencher and Schaalje, 2008).

Proof of Proposition 2.1

Part 1. Because the inverse sample covariance matrix is unbiased, the SMV portfolio $\hat{\boldsymbol{w}}^\star$ is unbiased as well, and the expected out-of-sample mean return of $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ in (2.7) is

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}\right] = \frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew}. \tag{2.108}$$

We can then obtain the expected out-of-sample variance, $\mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\right]$, from Equation (2.107). This requires knowing the covariance matrix of the SMV portfolio $\hat{\boldsymbol{w}}^\star$, which from Javed et al. (2021, Corollary 2.2) is[27]

$$\mathbb{V}[\hat{\boldsymbol{w}}^\star] = \frac{(T-N-2)(\theta^2 + (T-2)/T)\boldsymbol{\Sigma}^{-1} + (T-N)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}}{\gamma^2(T-N-1)(T-N-4)}. \tag{2.109}$$

Using $\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = \boldsymbol{w}(\boldsymbol{\kappa})$ and $\mathbb{V}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = \kappa_1^2\mathbb{V}[\hat{\boldsymbol{w}}^\star]$, we find from (2.107) that the expected out-of-sample variance of $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ is

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\right] = \frac{\kappa_1^2}{\gamma^2}(\theta^2 + d) + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew}. \tag{2.110}$$

Combining (2.108) and (2.110) results in the EU formula in (2.11).

Part 2. We obtain the combination coefficients maximizing the EU under the convexity constraint (2.10) by setting the derivative of (2.11) with respect to κ_1 equal to zero, taking into account that $\kappa_2 = 1 - \kappa_1$. Similarly, the unconstrained combination coefficients are found by setting the derivative of (2.11) with respect to κ_1 and κ_2 equal to zero and solving the resulting linear system of two equations with two unknowns. It is clear by comparing $\boldsymbol{\kappa}^c$ in (2.13) and $\boldsymbol{\kappa}^u$ in (2.14) that they are equal if and only if $\gamma = \gamma_{ew}$.

[27] Javed et al. (2021, Corollary 2.2) provide the covariance matrix of $\hat{\boldsymbol{w}}^\star$ for the case in which the sample covariance matrix is scaled by $T-1$, which we adapt in Equation (2.109) to our case in which it is scaled by $T-N-2$.

Part 3. Plugging the constrained combination coefficients $\boldsymbol{\kappa}^c$ in (2.13) into the EU formula in (2.11), we obtain that the EU of the constrained strategy is

$$\begin{aligned} EU(\hat{\boldsymbol{w}}^c) &= \frac{\psi_{ew}^2+\ell^2}{\psi_{ew}^2+\ell^2+d}\frac{\theta^2}{\gamma} + \frac{d\mu_{ew}}{\psi_{ew}^2+\ell^2+d} - \frac{\gamma}{2}\Bigg[\left(\frac{\psi_{ew}^2+\ell^2}{\psi_{ew}^2+\ell^2+d}\right)^2\frac{\theta^2+d}{\gamma^2} \\ &\quad + \frac{d^2\sigma_{ew}^2}{(\psi_{ew}^2+\ell^2+d)^2} + \frac{2d(\psi_{ew}^2+\ell^2)}{(\psi_{ew}^2+\ell^2+d)^2}\frac{\mu_{ew}}{\gamma}\Bigg] \\ &= U^\star - \frac{1}{2\gamma(\psi_{ew}^2+\ell^2+d)^2}\Big[\theta^2(\psi_{ew}^2+\ell^2+d)^2 + (\theta^2+d)(\psi_{ew}^2+\ell^2)^2 + \gamma^2 d^2\sigma_{ew}^2 \\ &\quad + 2\gamma d\mu_{ew}(\psi_{ew}^2+\ell^2) - 2\theta^2(\psi_{ew}^2+\ell^2)(\psi_{ew}^2+\ell^2+d) - 2\gamma d\mu_{ew}(\psi_{ew}^2+\ell^2+d)\Big]. \end{aligned} \tag{2.111}$$

The last term in brackets is equal to $d(\psi_{ew}^2+\ell^2)(\psi_{ew}^2+\ell^2+d)$, and thus (2.111) simplifies to (2.15). We proceed similarly for the unconstrained strategy: plugging $\boldsymbol{\kappa}^u$ in (2.14) into the EU formula in (2.11), the EU of the unconstrained strategy is

$$\begin{aligned} EU(\hat{\boldsymbol{w}}^u) &= \frac{\psi_{ew}^2}{\psi_{ew}^2+d}\frac{\theta^2}{\gamma} + \frac{d}{\psi_{ew}^2+d}\frac{\theta_{ew}^2}{\gamma} - \frac{\gamma}{2}\left[\frac{\psi_{ew}^4}{(\psi_{ew}^2+d)^2}\frac{\theta^2+d}{\gamma^2} + \frac{d^2}{(\psi_{ew}^2+d)^2}\frac{\theta_{ew}^2}{\gamma^2} + \frac{2d\psi_{ew}^2}{(\psi_{ew}^2+d)^2}\frac{\theta_{ew}^2}{\gamma^2}\right] \\ &= U^\star - \frac{1}{2\gamma(\psi_{ew}^2+d)^2}\Big[\theta^2(\psi_{ew}^2+d)^2 + (\theta^2+d)\psi_{ew}^4 + d^2\theta_{ew}^2 + 2d\psi_{ew}^2\theta_{ew}^2 \\ &\quad - 2\theta^2\psi_{ew}^2(\psi_{ew}^2+d) - 2d\theta_{ew}^2(\psi_{ew}^2+d)\Big]. \end{aligned} \tag{2.112}$$

The last term in brackets is equal to $d\psi_{ew}^2(\psi_{ew}^2+d)$, and thus (2.112) simplifies to (2.16). The unconstrained strategy $\hat{\boldsymbol{w}}^u$ delivers a larger EU than that of the SMV portfolio, the EW portfolio, and the risk-free asset because they are part of the search space by setting $\boldsymbol{\kappa} = (1,0)$, $(0,1)$, and $(0,0)$, respectively. For the same reason, the constrained strategy delivers a larger EU than that of the SMV and EW portfolios. However, the risk-free asset is not part of the search space of the constrained because $\boldsymbol{\kappa} = (0,0)$ does not fulfill the convexity constraint (2.10). Using Equation (2.15), the constrained strategy $\hat{\boldsymbol{w}}^c$ delivers a negative EU, and thus underperforms the risk-free asset, when

$$U^\star - \frac{d}{2\gamma}\kappa_1^c < 0. \tag{2.113}$$

After some developments, we find that inequality (2.113) is equivalent to

$$\gamma^2[\sigma_{ew}^2(d-\theta^2)] + \gamma[2\mu_{ew}(\theta^2-d)] - \theta^4 > 0. \tag{2.114}$$

The discriminant of this second-degree polynomial in γ is

$$\Delta = 4\sigma_{ew}^2(d-\theta^2)(\theta_{ew}^2 d + \theta^2\psi_{ew}^2). \tag{2.115}$$

It follows from (2.115) that $\Delta \geq 0$ and real roots to the polynomial exist if and only if $\theta^2 < d$. In that case, (2.114) holds if and only if γ is not between the two roots:

$$\gamma \notin \left[\gamma_{ew}\left(1 \pm \sqrt{1 + \frac{\theta^4/\theta_{ew}^2}{d-\theta^2}}\right)\right]. \tag{2.116}$$

The negative sign gives a negative root, which we can ignore because any $\gamma > 0$. Therefore, the constrained strategy delivers a negative EU if and only if (2.17) holds. Finally, the threshold γ_{neg} decreases with d because

$$\frac{\partial \gamma_{neg}}{\partial d} = -\frac{\mu_{ew}\theta^2}{2(d-\theta^2)^2\sqrt{1+\frac{\theta^4/\theta^2_{ew}}{d-\theta^2}}} \leq 0. \tag{2.117}$$

Proof of Proposition 2.2

Using Equations (2.15) and (2.16), the EU gain of the unconstrained strategy relative to the constrained strategy is

$$EU(\hat{\boldsymbol{w}}^u) - EU(\hat{\boldsymbol{w}}^c) = \frac{d}{2\gamma}\left(\kappa_1^c - \kappa_1^u\right). \tag{2.118}$$

Since $d \geq 0$, this EU gain is positive if $\kappa_1^c \geq \kappa_1^u$, which is the case because

$$\kappa_1^c - \kappa_1^u = \frac{d\ell^2}{(\psi^2_{ew}+d)(\psi^2_{ew}+\ell^2+d)} \geq 0. \tag{2.119}$$

The EU gain in (2.118) is zero if and only if $\gamma = \infty$ or γ_{ew}; the former case is obvious and the latter case is because $\ell^2 = 0$ when $\gamma = \gamma_{ew}$. Moreover, the EU gain in (2.118) increases with d if $d(\kappa_1^c - \kappa_1^u)$ increases with d, which is the case because

$$\begin{aligned}\frac{\partial}{\partial d} d(\kappa_1^c - \kappa_1^u) &= \frac{\partial}{\partial d}\frac{d^2\ell^2}{(\psi^2_{ew}+d)(\psi^2_{ew}+\ell^2+d)} \\ &= \frac{d\ell^2\left(\ell^2(2\psi^2_{ew}+d) + 2\psi^2_{ew}(\psi^2_{ew}+d)\right)}{(\psi^2_{ew}+d)^2(\psi^2_{ew}+\ell^2+d)^2} \geq 0.\end{aligned} \tag{2.120}$$

Next, we turn to the decomposition of the EU gain in (2.19). Given the EU formula in (2.11), it is clear that the EU of a portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = EU(\kappa_1\hat{\boldsymbol{w}}^\star) + EU(\kappa_2\boldsymbol{w}_{ew}) - \mathbb{E}[(\kappa_1\hat{\boldsymbol{w}}^\star)'\boldsymbol{\Sigma}(\kappa_2\boldsymbol{w}_{ew})], \tag{2.121}$$

where

$$EU(\kappa_1\hat{\boldsymbol{w}}^\star) = \frac{\kappa_1}{\gamma}\theta^2 - \frac{\gamma}{2}\frac{\kappa_1^2}{\gamma^2}(\theta^2+d), \tag{2.122}$$

$$EU(\kappa_2\boldsymbol{w}_{ew}) = \kappa_2\mu_{ew} - \frac{\gamma}{2}\kappa_2^2\sigma^2_{ew}, \tag{2.123}$$

$$\mathbb{E}[(\kappa_1\hat{\boldsymbol{w}}^\star)'\boldsymbol{\Sigma}(\kappa_2\boldsymbol{w}_{ew})] = \kappa_1\kappa_2\mu_{ew}. \tag{2.124}$$

Therefore, the EU gain in (2.19) can be decomposed as $\Delta_{smv} + \Delta_{ew} + \Delta_{cov}$, where Δ_{smv} is defined in (2.20), Δ_{ew} in (2.21), and Δ_{cov} in (2.22). We conclude this proof by deriving a closed-form expression for these three quantities.

Expression for Δ_{smv}. Given (2.122) and the formula for κ_1^c in (2.13) and κ_1^u in (2.14), we have

$$\begin{aligned}
\Delta_{smv} &= \frac{\psi_{ew}^2}{\psi_{ew}^2+d}\frac{\theta^2}{\gamma} - \frac{\psi_{ew}^4}{(\psi_{ew}^2+d)^2}\frac{\theta^2+d}{2\gamma} - \frac{\psi_{ew}^2+\ell^2}{\psi_{ew}^2+\ell^2+d}\frac{\theta^2}{\gamma} + \frac{(\psi_{ew}^2+\ell^2)^2}{(\psi_{ew}^2+\ell^2+d)^2}\frac{\theta^2+d}{2\gamma} \\
&= \frac{1}{2\gamma(\psi_{ew}^2+d)^2(\psi_{ew}^2+\ell^2+d)^2}\Big[2\theta^2\psi_{ew}^2(\psi_{ew}^2+d)(\psi_{ew}^2+\ell^2+d)^2 - \psi_{ew}^4(\theta^2+d)(\psi_{ew}^2+\ell^2+d)^2 \\
&\quad - 2\theta^2(\psi_{ew}^2+\ell^2)(\psi_{ew}^2+\ell^2+d)(\psi_{ew}^2+d)^2 + (\theta^2+d)(\psi_{ew}^2+\ell^2)^2(\psi_{ew}^2+d)^2\Big] \\
&= \frac{d^2\ell^2(\psi_{ew}^2+d-\theta_{ew}^2)}{2\gamma(\psi_{ew}^2+d)^2(\psi_{ew}^2+\ell^2+d)^2}\left[\sigma_{ew}^2\gamma^2 - 2\mu_{ew}\gamma - \theta_{ew}^2\frac{d+\psi_{ew}^2+\theta_{ew}^2}{d+\psi_{ew}^2-\theta_{ew}^2}\right], \qquad (2.125)
\end{aligned}$$

which is equivalent to (2.20). The term in brackets is a second-degree polynomial in γ whose discriminant is $8\mu_{ew}^2(\psi_{ew}^2+d)/(\psi_{ew}^2+d-\theta_{ew}^2)$. When $\psi_{ew}^2+d-\theta_{ew}^2 \leq 0$, the polynomial is always positive, and thus $\Delta_{smv} \leq 0$ because of the term $\psi_{ew}^2+d-\theta_{ew}^2$ in front. Otherwise, the two roots of the polynomial are

$$\gamma_{ew}\left(1 \pm \sqrt{\frac{2(\psi_{ew}^2+d)}{\psi_{ew}^2+d-\theta_{ew}^2}}\right). \qquad (2.126)$$

We have $\Delta_{smv} \geq 0$ when γ is not in between the two roots, and since the smaller root is negative while any $\gamma > 0$, this is the case if $\gamma \geq \gamma_{ew}\left(1+\sqrt{\frac{2(\psi_{ew}^2+d)}{\psi_{ew}^2+d-\theta_{ew}^2}}\right)$. Otherwise, $\Delta_{smv} < 0$.

Expression for Δ_{ew}. Given (2.123) and the formula for κ_2^c in (2.13) and κ_2^u in (2.14), we have

$$\begin{aligned}
\Delta_{ew} &= \frac{d}{\psi_{ew}^2+d}\frac{\theta_{ew}^2}{\gamma} - \frac{d^2}{(\psi_{ew}^2+d)^2}\frac{\theta_{ew}^2}{2\gamma} - \frac{d\mu_{ew}}{\psi_{ew}^2+\ell^2+d} + \frac{\gamma}{2}\frac{d^2\sigma_{ew}^2}{(\psi_{ew}^2+\ell^2+d)^2}, \\
&= \frac{d\sigma_{ew}^2}{2\gamma(\psi_{ew}^2+d)^2(\psi_{ew}^2+\ell^2+d)^2}\Big[2\gamma_{ew}^2(\psi_{ew}^2+d)(\psi_{ew}^2+\ell^2+d)^2 - d\gamma_{ew}^2(\psi_{ew}^2+\ell^2+d)^2 \\
&\quad - 2\gamma_{ew}\gamma(\psi_{ew}^2+\ell^2+d)(\psi_{ew}^2+d)^2 + d\gamma^2(\psi_{ew}^2+d)^2\Big] \\
&= \frac{d\sigma_{ew}^2(\gamma-\gamma_{ew})(\theta^2+d-\mu_{ew}\gamma)}{2\gamma(\psi_{ew}^2+d)^2(\psi_{ew}^2+\ell^2+d)^2}\Big[d(\gamma-\gamma_{ew})(\theta^2+d-\mu_{ew}\gamma) - 2\gamma_{ew}\psi_{ew}^2\left(\psi_{ew}^2+\ell^2+d\right)\Big], \\
& \qquad (2.127)
\end{aligned}$$

which corresponds to (2.21). The last term in brackets is the following second-degree polynomial in γ:

$$\begin{aligned}
&d(\gamma-\gamma_{ew})\theta^2 + d - \mu_{ew}\gamma) - 2\gamma_{ew}\psi_{ew}^2\left(\psi_{ew}^2+\ell^2+d\right) \\
&= -\gamma^2\mu_{ew}[d+2\psi_{ew}^2] + \gamma[d(\theta^2+\theta_{ew}^2+d) + 4\theta_{ew}^2\psi_{ew}^2] - \gamma_{ew}[d(d+\theta_{ew}^2+3\psi_{ew}^2) + 2\psi_{ew}^2\theta^2]. \\
& \qquad (2.128)
\end{aligned}$$

The discriminant of this polynomial is $\Delta = (\psi_{ew}^2+d)(d^2(\psi_{ew}^2+d) - 8\theta_{ew}^2\psi_{ew}^2(2\psi_{ew}^2+d))$. If d is small enough such that $\Delta \leq 0$, then the polynomial in (2.128) is always negative, and thus $\Delta_{ew} \leq 0$ if $\gamma \in [\gamma_{ew}, (\theta^2+d)/\mu_{ew}]$, and else $\Delta_{ew} > 0$. Otherwise, if d is large enough

such that $\Delta > 0$, then the numerator of Δ_{ew}, which is a fourth-degree polynomial in γ, has four roots: $\gamma = \gamma_{ew}$, $\gamma = (\theta^2 + d)/\mu_{ew}$, and the two roots of (2.128), which are

$$\gamma = \frac{d^2 + d(\theta^2 + \theta_{ew}^2) + 4\theta_{ew}^2\psi_{ew}^2 \pm \sqrt{(\psi_{ew}^2 + d)(d^2(\psi_{ew}^2 + d) - 8\theta_{ew}^2\psi_{ew}^2(2\psi_{ew}^2 + d))}}{2\mu_{ew}(2\psi_{ew}^2 + d)}. \tag{2.129}$$

It is easy to check that γ_{ew} is the smallest root, $(\theta^2 + d)/\mu_{ew}$ is the largest root, and the two roots in (2.129) are in between. Therefore, given that the coefficient in front of γ^4 in the numerator of Δ_{ew} is positive, it holds that $\Delta_{ew} \geq 0$ if either $\gamma \leq \gamma_{ew}$, $\gamma \geq (\theta^2 + d)/\mu_{ew}$, or γ is in between the two roots in (2.129).

Expression for Δ_{cov}. Given (2.124) and the formula for (κ_1^c, κ_2^c) in (2.13) and (κ_1^u, κ_2^u) in (2.14), we have

$$\begin{aligned}
\Delta_{cov} &= \mu_{ew}\left(\frac{d(\psi_{ew}^2 + \ell^2)}{(\psi_{ew}^2 + \ell^2 + d)^2} - \frac{d\psi_{ew}^2}{(\psi_{ew}^2 + d)^2}\frac{\gamma_{ew}}{\gamma}\right) \\
&= \frac{d\mu_{ew}}{\gamma(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2 + d)^2}\Big[\gamma(\psi_{ew}^2 + \ell^2)(\psi_{ew}^2 + d)^2 - \gamma_{ew}\psi_{ew}^2(\psi_{ew}^2 + \ell^2 + d)^2\Big] \\
&= \frac{d\mu_{ew}(\gamma - \gamma_{ew})}{\gamma(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2 + d)^2}\Big[\psi_{ew}^2(\psi_{ew}^2 + d)^2 + \sigma_{ew}^2(\gamma - \gamma_{ew})\big(\gamma(\psi_{ew}^2 + d)^2 \\
&\quad - \gamma_{ew}\psi_{ew}^2(\ell^2 + 2(\psi_{ew}^2 + d))\big)\Big] \\
&= \frac{d\mu_{ew}(\gamma - \gamma_{ew})}{\gamma(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2 + d)^2}\Big[(\psi_{ew}^2 + d)^2(\psi_{ew}^2 + \ell^2) + \mu_{ew}(\gamma - \gamma_{ew})(d^2 - \psi_{ew}^2(\psi_{ew}^2 + \ell^2))\Big],
\end{aligned} \tag{2.130}$$

which is equivalent to (2.22).

Proof of Proposition 2.3

The unconstrained three-fund strategy always delivers a larger EU than that of the unconstrained two-fund strategy by construction. To determine when the constrained three-fund strategy outperforms the unconstrained two-fund one, we need to determine the EU of the two-fund strategy. Plugging $\kappa_1 = \kappa^{2f} = \theta^2/(\theta^2 + d)$ and $\kappa_2 = 0$ into Equation (2.11), we have

$$EU(\hat{\boldsymbol{w}}^{2f}) = U^\star - \frac{d}{2\gamma}\kappa^{2f}. \tag{2.131}$$

Therefore, given (2.15), the constrained three-fund strategy delivers a larger EU than that of the unconstrained two-fund strategy when $\kappa_1^c \leq \kappa^{2f}$, which is equivalent to $\gamma \leq 2\gamma_{ew}$.

Proof of Proposition 2.4

We want to find when the L_2-norm of the expected weights of unconstrained strategy is smaller than that of the expected weights of the constrained strategy, which is equivalent

to showing the same result for the squared L_2-norm because the L_2-norm is positive. We define $f(\gamma)$ the function of γ corresponding to the difference between the squared L_2-norms:

$$f(\gamma) = g(\boldsymbol{\kappa}^c) - g(\boldsymbol{\kappa}^u), \tag{2.132}$$

where

$$g(\boldsymbol{\kappa}) = ||\kappa_1 \boldsymbol{w}^\star + \kappa_2 \boldsymbol{w}_{ew}||_2^2 = \frac{\kappa_1^2}{\gamma^2}\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|_2^2 + \frac{\kappa_2^2}{N} + \frac{2\kappa_1\kappa_2}{N\gamma}\gamma_{tan}. \tag{2.133}$$

We want to find the values of γ for which $f(\gamma) \geq 0$ holds, which can be rewritten as

$$f(\gamma) = \frac{(\kappa_1^c)^2 - (\kappa_1^u)^2}{\gamma^2}\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|_2^2 + \frac{(\kappa_2^c)^2 - (\kappa_1^u)^2}{N} + 2\frac{\gamma_{tan}}{N\gamma}\left(\kappa_1^c\kappa_2^c - \kappa_1^u\kappa_2^u\right) \geq 0. \tag{2.134}$$

Proposition 2.9 shows that $\kappa_1^c \geq \kappa_1^u \geq 0$, $\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|_2^2 \geq 0$ by definition, and we assume $\gamma_{tan} \geq 0$. Therefore, a sufficient condition for (2.134) to hold is

$$(\kappa_2^c)^2 - (\kappa_2^u)^2 + \frac{2\gamma_{tan}}{\gamma}\left(\kappa_1^c\kappa_2^c - \kappa_1^u\kappa_2^u\right) \geq 0. \tag{2.135}$$

Inequality (2.135) holds for all values of γ_{tan} if the following two inequalities hold:

$$(\kappa_2^c)^2 \geq (\kappa_2^u)^2 \quad \text{and} \quad \kappa_1^c\kappa_2^c \geq \kappa_1^u\kappa_2^u. \tag{2.136}$$

Both inequalities hold whenever $0 \leq \kappa_2^u \leq \kappa_2^c$ which, from Proposition 2.9, holds when $\gamma \in \left[\gamma_{ew}, \frac{\theta^2+d}{\mu_{ew}}\right]$. Therefore, $f(\gamma) \geq 0$ in (2.132) if $\gamma \in [\gamma_{ew}, (\theta^2+d)/\mu_{ew}]$.

Proof of Proposition 2.5

Part 1. The two combination coefficients are κ_1^u in (2.14) and $\tilde{\kappa}_2^u$ in (2.24). Because $\hat{\mu}_{ew}$ is unbiased, $\tilde{\kappa}_2^u$ is unbiased as well, and the expected out-of-sample mean return of the estimated unconstrained strategy is the same as that when κ_2^u is known:

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u)'\boldsymbol{\mu}\right] = \frac{\kappa_1^u}{\gamma}\theta^2 + \kappa_2^u\mu_{ew}. \tag{2.137}$$

In comparison, the expected out-of-sample variance is larger than that when κ_2^u is known. Specifically,

$$\begin{aligned}\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u)\right] &= \mathbb{E}\left[\left(\frac{\kappa_1^u}{\gamma}\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma} + \tilde{\kappa}_2^u\boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\right)'\left(\frac{\kappa_1^u}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} + \tilde{\kappa}_2^u\boldsymbol{w}_{ew}\right)\right] \\ &= \frac{(\kappa_1^u)^2}{\gamma^2}(\theta^2+d) + \mathbb{E}\left[(\tilde{\kappa}_2^u)^2\right]\sigma_{ew}^2 + \frac{2\kappa_1^u}{\gamma}\mathbb{E}\left[\tilde{\kappa}_2^u\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\boldsymbol{w}_{ew}\right].\end{aligned} \tag{2.138}$$

From the definition of $\tilde{\kappa}_2^u$ in (2.24) and the identity $\mathbb{E}[\hat{\mu}_{ew}^2] = \mu_{ew}^2 + \sigma_{ew}^2/T$, we find that

$$\sigma_{ew}^2\mathbb{E}\left[(\tilde{\kappa}_2^u)^2\right] = (\kappa_2^u)^2\sigma_{ew}^2 + \frac{(1-\kappa_1^u)^2}{\gamma^2 T} \tag{2.139}$$

and

$$\frac{2\kappa_1^u}{\gamma}\mathbb{E}\left[\tilde{\kappa}_2^u\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\boldsymbol{w}_{ew}\right] = \frac{2\kappa_1^u\kappa_2^u}{\gamma}\mu_{ew} + \frac{2\kappa_1^u(1-\kappa_1^u)}{\gamma^2 T}. \tag{2.140}$$

Therefore, the expected out-of-sample variance is

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u,\tilde{\kappa}_2^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\kappa_1^u,\tilde{\kappa}_2^u)\right] = \mathbb{E}\left[(\hat{\boldsymbol{w}}^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^u\right] + \frac{1-(\kappa_1^u)^2}{\gamma^2 T}, \tag{2.141}$$

where $\mathbb{E}\left[(\hat{\boldsymbol{w}}^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^u\right] = \frac{(\kappa_1^u)^2}{\gamma^2}(\theta^2+d) + (\kappa_2^u)^2\sigma_{ew}^2 + \frac{2\kappa_1^u\kappa_2^u}{\gamma}\mu_{ew}$ is the expected out-of-sample variance when κ_2^u is known. Finally, using Equation (2.16), the EU loss relative to $\hat{\boldsymbol{w}}^u$ when κ_2^u is estimated by $\hat{\kappa}_2^u$ is given by (2.25).

Part 2. Using Equation (2.16), the constrained strategy delivers a larger EU than that of the estimated unconstrained strategy, $EU(\hat{\boldsymbol{w}}^c) \geq EU(\hat{\boldsymbol{w}}(\kappa_1^u,\tilde{\kappa}_2^u))$, when

$$U^\star - \frac{d}{2\gamma}\kappa_1^c \geq U^\star - \frac{d}{2\gamma}\kappa_1^u - \frac{1-(\kappa_1^u)^2}{2\gamma T}. \tag{2.142}$$

After some developments, we find that inequality (2.142) is equivalent to

$$\gamma^2[-\sigma_{ew}^2] + \gamma[2\mu_{ew}] + k \geq 0, \tag{2.143}$$

where

$$k = \frac{(\psi_{ew}^2+d)(2\psi_{ew}^2+d)}{dT(\psi_{ew}^2+d) - (2\psi_{ew}^2+d)} - \theta_{ew}^2. \tag{2.144}$$

The discriminant of the second-degree polynomial in (2.143) is $\Delta = 4\sigma_{ew}^2(k+\theta_{ew}^2)$, which is positive because we assume $N \geq 2$. Therefore, inequality (2.143) holds when the risk-aversion coefficient γ belongs to the interval in Equation (2.26). Finally, the length of the interval (2.26) decreases with d because

$$\frac{\partial}{\partial d}\left[\frac{(\psi_{ew}^2+d)(2\psi_{ew}^2+d)}{dT(\psi_{ew}^2+d) - (2\psi_{ew}^2+d)}\right] = -\frac{2\psi_{ew}^6 T + 4\psi_{ew}^4(dT+1) + 2\psi_{ew}^2 d(dT+2) + d^2}{\left(dT(\psi_{ew}^2+d) - (2\psi_{ew}^2+d)\right)^2} \leq 0. \tag{2.145}$$

Proof of Proposition 2.6

Part 1. The mean and variance of $\tilde{\kappa}_2^u(\delta)$ are

$$\mathbb{E}[\tilde{\kappa}_2^u(\delta)] = (1-\delta)\kappa_2^u + \delta(1-\kappa_1^u), \tag{2.146}$$

$$\mathbb{V}[\tilde{\kappa}_2^u(\delta)] = \frac{(1-\delta)^2(1-\kappa_1^u)^2}{\gamma^2\sigma_{ew}^2 T}. \tag{2.147}$$

Therefore, the mean squared error of $\tilde{\kappa}_2^u(\delta)$ is

$$(\mathbb{E}[\tilde{\kappa}_2^u(\delta)] - \kappa_2^u)^2 + \mathbb{V}[\tilde{\kappa}_2^u(\delta)] = \frac{(1-\kappa_1^u)^2}{\gamma^2\sigma_{ew}^2}\left(\delta^2\ell^2 + \frac{(1-\delta)^2}{T}\right). \tag{2.148}$$

Setting the derivative of (2.148) equal to zero delivers the optimal $\delta^\star$ in (2.28).

Part 2. The formula for $\tilde{\kappa}_2^u(\delta^\star)$ in (2.29) is directly obtained by plugging $\delta = \delta^\star$ into (2.27). Moreover, the two derivatives in (2.30) are

$$\frac{\partial}{\partial \hat{\mu}_{ew}} \tilde{\kappa}_2^u(\delta^\star) = \frac{1-\kappa_1^u}{\gamma \sigma_{ew}^2} \frac{T\ell^2}{1+T\ell^2}, \tag{2.149}$$

$$\frac{\partial}{\partial \hat{\mu}_{ew}} \tilde{\kappa}_2^u = \frac{1-\kappa_1^u}{\gamma \sigma_{ew}^2}, \tag{2.150}$$

and thus their ratio is equal to $1-\delta^\star \leq 1$, as desired.

Part 3. To prove the result in (2.31), we derive a formula for the EU delivered by $\tilde{\kappa}_2^u(\delta)$ for any δ, and then plug in $\delta^\star$. Using the mean and variance of $\tilde{\kappa}_2^u(\delta)$ in (2.146)–(2.147), we find that the expected out-of-sample mean return delivered by $\tilde{\kappa}_2^u(\delta)$ is

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u(\delta))' \boldsymbol{\mu}\right] = \frac{\kappa_1^u}{\gamma}\theta^2 + (1-\delta)\kappa_2^u \mu_{ew} + \delta(1-\kappa_1^u)\mu_{ew}. \tag{2.151}$$

Then, the expected out-of-sample variance is

$$\begin{aligned} &\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u(\delta))' \boldsymbol{\Sigma} \hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u(\delta))\right] \\ &= \frac{(\kappa_1^u)^2}{\gamma^2}(\theta^2 + d) + \sigma_{ew}^2 \mathbb{E}\big[\tilde{\kappa}_2^u(\delta)^2\big] + \frac{2\kappa_1^u}{\gamma} \mathbb{E}\big[\tilde{\kappa}_2^u(\delta) \hat{\boldsymbol{\mu}}' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \boldsymbol{w}_{ew}\big], \end{aligned} \tag{2.152}$$

where, using (2.139)–(2.140), the two expectations in (2.152) are

$$\mathbb{E}\big[\tilde{\kappa}_2^u(\delta)^2\big] = (1-\delta)^2(\kappa_2^u)^2 + \delta^2(1-\kappa_1^u)^2 + 2\delta(1-\delta)(1-\kappa_1^u)\kappa_2^u + \frac{(1-\delta)^2(1-\kappa_1^u)^2}{\gamma^2 \sigma_{ew}^2 T}, \tag{2.153}$$

$$\mathbb{E}\big[\tilde{\kappa}_2^u(\delta) \hat{\boldsymbol{\mu}}' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \boldsymbol{w}_{ew}\big] = (1-\delta)\kappa_2^u \mu_{ew} + \frac{(1-\delta)(1-\kappa_1^u)}{\gamma T} + \delta(1-\kappa_1^u)\mu_{ew}. \tag{2.154}$$

Next, using $\kappa_2^u = \frac{\gamma_{ew}}{\gamma}(1-\kappa_1^u)$ and the fact that the EU when κ_2^u is known, $EU(\hat{\boldsymbol{w}}^u)$ in (2.16) is obtained by plugging $(\kappa_1, \kappa_2) = (\kappa_1^u, \kappa_2^u)$ into (2.11), we obtain after some simplifications that the EU delivered by $\tilde{\kappa}_2^u(\delta)$ is

$$EU(\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u(\delta))) = EU(\hat{\boldsymbol{w}}^u) - \frac{1-\kappa_1^u}{2\gamma T}\left(\frac{\delta^2(1-\kappa_1^u)}{\delta^\star} - 2\delta + 1 + \kappa_1^u\right). \tag{2.155}$$

Now, plugging $\delta = \delta^\star$ into (2.155), we obtain

$$EU(\hat{\boldsymbol{w}}(\kappa_1^u, \tilde{\kappa}_2^u(\delta^\star))) = EU(\hat{\boldsymbol{w}}^u) - \frac{(1-\delta^\star)(1-(\kappa_1^u)^2)}{2\gamma T}, \tag{2.156}$$

which, given (2.25), is equivalent to (2.31).

Proof of Proposition 2.7

Under the elliptical model (2.38) and in the high-dimensional regime as $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, El Karoui (2010, 2013) show that under a set of technical assumptions that ensure that the distribution of the τ_t's does not put too much mass near zero (see (Assumption-BB) in El Karoui (2013) for details) we have, for any vector $\boldsymbol{v}$,

$$\boldsymbol{v}'\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} \xrightarrow{p} \eta \boldsymbol{v}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{2.157}$$

$$\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} \xrightarrow{p} \frac{\varphi \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \eta\phi}{1-\phi}, \tag{2.158}$$

provided $\boldsymbol{v}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ are bounded as $N \to \infty$, where $\eta \geq 1$ and $\varphi \geq \eta^2$ are defined in the proposition statement.[28] We exploit these results to prove the proposition.
Part 1. The out-of-sample utility of the portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ is

$$U(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = \frac{\kappa_1}{\gamma}\boldsymbol{\mu}'\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} + \kappa_2\mu_{ew} - \frac{\gamma}{2}\left(\frac{\kappa_1^2}{\gamma^2}\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\right). \tag{2.159}$$

Given (2.157)–(2.158), we directly find that the asymptotic limit of $U(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}))$ is equal to $\bar{U}(\boldsymbol{\kappa})$ in (2.44).

Part 2. We obtain the combination coefficients maximizing $\bar{U}(\boldsymbol{\kappa})$ under the convexity constraint (2.10) by setting the derivative of (2.44) with respect to κ_1 equal to zero, taking into account that $\kappa_2 = 1 - \kappa_1$. Then, to evaluate the asymptotic loss in out-of-sample utility, we use the fact that when $\kappa_2 = 1 - \kappa_1$, $\bar{U}(\boldsymbol{\kappa})$ simplifies to

$$\begin{aligned}\bar{U}(\kappa_1, 1-\kappa_1) = \kappa_1^2\left(\eta\mu_{ew} - \frac{\gamma}{2}\sigma_{ew}^2 - \frac{1}{2\gamma}\frac{\varphi\theta^2 + \eta\phi}{1-\phi}\right) + \kappa_1\left(\frac{\eta\theta^2}{\gamma} + \gamma\sigma_{ew}^2 - (1+\eta)\mu_{ew}\right) \\ + \mu_{ew} - \frac{\gamma}{2}\sigma_{ew}^2.\end{aligned} \tag{2.160}$$

Using (2.160), the asymptotic out-of-sample utility loss is

$$\begin{aligned}\bar{U}(\bar{\boldsymbol{\kappa}}_n^c) - \bar{U}(\bar{\boldsymbol{\kappa}}^c) = \left((\bar{\kappa}_{1,n}^c)^2 - (\bar{\kappa}_1^c)^2\right)\left(\eta\mu_{ew} - \frac{\gamma}{2}\sigma_{ew}^2 - \frac{1}{2\gamma}\frac{\varphi\theta^2 + \eta\phi}{1-\phi}\right) \\ + \left(\bar{\kappa}_{1,n}^c - \bar{\kappa}_1^c\right)\left(\frac{\eta\theta^2}{\gamma} + \gamma\sigma_{ew}^2 - (1+\eta)\mu_{ew}\right).\end{aligned} \tag{2.161}$$

Now, notice that

$$\begin{aligned}&\eta\mu_{ew} - \frac{\gamma}{2}\sigma_{ew}^2 - \frac{1}{2\gamma}\frac{\varphi\theta^2 + \eta\phi}{1-\phi} \\ &= -\frac{1}{2\gamma}\left(\eta\psi_{ew}^2 + \sigma_{ew}\ell(\gamma - \eta\gamma_{ew}) + (1-\eta)\gamma\mu_{ew} + \frac{(\varphi-\eta)\theta^2 + \eta\phi(1+\theta^2)}{1-\phi}\right) \\ &= -\frac{1}{2\gamma}\frac{\eta\psi_{ew}^2 + \sigma_{ew}\ell(\gamma - \eta\gamma_{ew})}{\bar{\kappa}_1^c},\end{aligned} \tag{2.162}$$

[28] Our definition of η and φ is such that $\eta = (1-\phi)s$ and $\varphi = (1-\phi)^3\xi$, where (s, ξ) are defined in Equations (3.1) and (3.2) in El Karoui (2013).

and

$$\frac{\eta\theta^2}{\gamma} + \gamma\sigma_{ew}^2 - (1+\eta)\mu_{ew} = \frac{1}{\gamma}\left(\eta\psi_{ew}^2 + \sigma_{ew}\ell(\gamma - \eta\gamma_{ew})\right). \tag{2.163}$$

Plugging (2.162)–(2.163) into (2.161), we obtain after some developments that the asymptotic utility loss simplifies to (2.46).

Part 3. The unconstrained combination coefficients maximizing $\bar{U}(\boldsymbol{\kappa})$ are found by setting the derivative of (2.44) with respect to κ_1 and κ_2 equal to zero and solving the resulting linear system of two equations with two unknowns. Then, we evaluate the asymptotic loss in out-of-sample utility. Since $\bar{\kappa}_{2,n}^u = \frac{\gamma_{ew}}{\gamma}(1 - \bar{\kappa}_{1,n}^u)$, the asymptotic out-of-sample utility delivered by the combination coefficients that are optimal under normally distributed returns is

$$\bar{U}(\bar{\boldsymbol{\kappa}}_n^u) = U^\star - \frac{1}{2\gamma}\left((\bar{\kappa}_{1,n}^u)^2\left(\frac{\varphi\theta^2 + \eta\phi}{1-\phi} - (2\eta-1)\theta_{ew}^2\right) - 2\eta\bar{\kappa}_{1,n}^u\psi_{ew}^2 + \psi_{ew}^2\right). \tag{2.164}$$

For the combination coefficients that are optimal under elliptically distributed returns, we have $\bar{\kappa}_2^u = \frac{\gamma_{ew}}{\gamma}(1 - \eta\bar{\kappa}_1^u)$ and thus the asymptotic out-of-sample utility is

$$\bar{U}(\bar{\boldsymbol{\kappa}}^u) = U^\star - \frac{1}{2\gamma}\left((\bar{\kappa}_1^u)^2\left(\frac{\varphi\theta^2 + \eta\phi}{1-\phi} - \eta^2\theta_{ew}^2\right) - 2\eta\bar{\kappa}_1^u\psi_{ew}^2 + \psi_{ew}^2\right). \tag{2.165}$$

Using (2.164)–(2.165), the asymptotic out-of-sample utility loss is

$$\begin{aligned}&\bar{U}(\bar{\kappa}_{1,n}^u, \bar{\kappa}_{2,n}^u) - \bar{U}(\bar{\kappa}_1^u, \bar{\kappa}_2^u) = \\ &-\frac{1}{2\gamma}\left((\bar{\kappa}_1^u)^2\left(\eta^2\theta_{ew}^2 - \frac{\varphi\theta^2 + \eta\phi}{1-\phi}\right) + (\bar{\kappa}_{1,n}^u)^2\left(\frac{\varphi\theta^2 + \eta\phi}{1-\phi} - (2\eta-1)\theta_{ew}^2\right) + 2\eta\psi_{ew}^2\left(\bar{\kappa}_1^u - \bar{\kappa}_{1,n}^u\right)\right).\end{aligned} \tag{2.166}$$

Now, notice that

$$\eta^2\theta_{ew}^2 - \frac{\varphi\theta^2 + \eta\phi}{1-\phi} = -\left(\eta^2\psi_{ew}^2 + \frac{(\varphi-\eta^2)\theta^2 + \eta\phi(1+\eta\theta^2)}{1-\phi}\right) = -\frac{\eta\psi_{ew}^2}{\bar{\kappa}_1^u}, \tag{2.167}$$

and

$$\begin{aligned}&\frac{\varphi\theta^2 + \eta\phi}{1-\phi} - (2\eta-1)\theta_{ew}^2 \\ &= \eta^2\psi_{ew}^2 + (\eta-1)^2\theta_{ew}^2 + \frac{(\varphi-\eta^2)\theta^2 + \eta\phi(1+\eta\theta^2)}{1-\phi} = \frac{\eta\psi_{ew}^2}{\bar{\kappa}_1^u} + (\eta-1)^2\theta_{ew}^2.\end{aligned} \tag{2.168}$$

Plugging (2.167)–(2.168) into (2.166), we obtain after some developments that the asymptotic utility loss simplifies to (2.48).

Proof of Proposition 2.8

Part 1. When $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ while $\boldsymbol{\Sigma}$ is known, we have

$$\hat{\mu}_{ew} = \boldsymbol{w}_{ew}'\hat{\boldsymbol{\mu}} = \mu_{ew} + \bar{\varepsilon}, \tag{2.169}$$

$$\hat{\gamma}_{ew} = \hat{\mu}_{ew}/\sigma_{ew}^2 = \gamma_{ew} + \bar{\varepsilon}/\sigma_{ew}^2, \tag{2.170}$$

$$\hat{\theta}^2 = \hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}} = \theta^2 + 2\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}, \tag{2.171}$$

$$\hat{\psi}_{ew}^2 = \hat{\theta}^2 - \frac{\hat{\mu}_{ew}^2}{\sigma_{ew}^2} = \psi_{ew}^2 + 2\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma_{ew} - \bar{\varepsilon}^2/\sigma_{ew}^2. \tag{2.172}$$

Using (2.169)–(2.172), the estimated combination coefficients are

$$\hat{\kappa}_1^u = \frac{\hat{\psi}_{ew}^2}{\hat{\psi}_{ew}^2 + cN/T + (c-1)\hat{\theta}^2} = \frac{\psi_{ew}^2 + 2\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma_{ew} - \bar{\varepsilon}^2/\sigma_{ew}^2}{\psi_{ew}^2 + d + 2c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma_{ew} - \bar{\varepsilon}^2/\sigma_{ew}^2}, \tag{2.173}$$

$$\hat{\kappa}_2^u = \frac{\hat{\gamma}_{ew}}{\gamma}(1 - \hat{\kappa}_1^u) = \frac{\mu_{ew} + \bar{\varepsilon}}{\gamma\sigma_{ew}^2}(1 - \hat{\kappa}_1^u), \tag{2.174}$$

$$\hat{\kappa}_1^c = \frac{\hat{\psi}_{ew}^2 + \sigma_{ew}^2(\gamma - \hat{\gamma}_{ew})^2}{\hat{\psi}_{ew}^2 + \sigma_{ew}^2(\gamma - \hat{\gamma}_{ew})^2 + cN/T + (c-1)\hat{\theta}^2} = \frac{\psi_{ew}^2 + \ell^2 + 2\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma}{\psi_{ew}^2 + \ell^2 + d + 2c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + c\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} - 2\bar{\varepsilon}\gamma}, \tag{2.175}$$

$$\hat{\kappa}_2^c = 1 - \hat{\kappa}_1^c, \tag{2.176}$$

which yields the desired estimation errors in (2.52)–(2.54). Finally, $\tilde{\kappa}_2^u = \kappa_2^u + \frac{\bar{\varepsilon}}{\gamma\sigma_{ew}^2}(1 - \kappa_1^u)$, which corresponds to (2.55).

Proof of Proposition 2.9

Part 1. In Proposition 2.2, we prove that $EU(\hat{\boldsymbol{w}}^u) - EU(\hat{\boldsymbol{w}}^c) = \frac{d}{2\gamma}(\kappa_1^c - \kappa_1^u) \geq 0$, which implies that $\kappa_1^c \geq \kappa_1^u$. Moreover, both κ_1^c and κ_1^u are between zero and one given their definition in (2.13) and (2.14). Finally, since $\pi_{tan} = \frac{\gamma_{tan}}{\gamma}\kappa_1$, we have $0 \leq \pi_{tan}^u \leq \pi_{tan}^c$ if $\gamma_{tan} \geq 0$.

Part 2. We can show that

$$\begin{aligned}
\kappa_2^u - \kappa_2^c &= \frac{\gamma_{ew}}{\gamma}\frac{d}{\psi_{ew}^2 + d} - \frac{d}{\psi_{ew}^2 + \ell^2 + d} \\
&= \frac{d}{\gamma(\psi_{ew}^2 + d)(\psi_{ew}^2 + \ell^2 + d)}\left(\gamma_{ew}(\psi_{ew}^2 + \ell^2 + d) - \gamma(\psi_{ew}^2 + d)\right) \\
&= \frac{d\mu_{ew}(\gamma - \gamma_{ew})}{\gamma(\psi_{ew}^2 + d)(\psi_{ew}^2 + \ell^2 + d)}\left(\gamma - \frac{\theta^2 + d}{\mu_{ew}}\right).
\end{aligned} \tag{2.177}$$

Therefore, the inequality $\kappa_2^u \geq \kappa_2^c$ is equivalent to

$$(\gamma - \gamma_{ew})\left(\gamma - \frac{\theta^2 + d}{\mu_{ew}}\right) \geq 0. \tag{2.178}$$

Since $(\theta^2 + d)/\mu_{ew} \geq \gamma_{ew}$ under the assumption that $\mu_{ew} \geq 0$, it directly follows that $0 \leq \kappa_2^u \leq \kappa_2^c$ if $\gamma \in \left[\gamma_{ew}, \frac{\theta^2+d}{\mu_{ew}}\right]$ and $\kappa_2^u \geq \kappa_2^c$ otherwise.

Part 3. After some developments, the inequality $\pi_{rf}^u \geq \pi_{rf}^c$ is equivalent to

$$\gamma^2 \left[\sigma_{ew}^2(\gamma_{tan} - \gamma_{ew})\right] + \gamma \left[2\mu_{ew}(\gamma_{ew} - \gamma_{tan}) + \psi_{ew}^2 + d\right] + \left[\gamma_{tan}\theta_{ew}^2 - \gamma_{ew}(\theta^2 + d)\right] \geq 0. \tag{2.179}$$

Because we assume $\gamma_{tan} > \gamma_{ew}$, this polynomial has two roots:

$$\gamma = \gamma_{ew} \quad \text{and} \quad \gamma = \frac{\gamma_{tan} - (\theta^2 + d)/\mu_{ew}}{\gamma_{tan}/\gamma_{ew} - 1}, \tag{2.180}$$

and we can then write the following about inequality (2.179):

$$\pi_{rf}^u \leq \pi_{rf}^c \quad \text{if} \quad \gamma \in \left[\frac{\gamma_{tan} - (\theta^2 + d)/\mu_{ew}}{\gamma_{tan}/\gamma_{ew} - 1}, \gamma_{ew}\right] \quad \text{and} \quad \pi_{rf}^u \geq \pi_{rf}^c \quad \text{otherwise.} \tag{2.181}$$

The lower bound of the interval is negative under the assumption $\gamma_{tan} < (\theta^2 + d)/\mu_{ew}$. Therefore, $\pi_{rf}^u \leq \pi_{rf}^c$ holds for $\gamma \leq \gamma_{ew}$ becaus any $\gamma > 0$. Finally, when $\gamma \leq \gamma_{ew}$, we have that $\pi_{rf}^c \leq 0$ because $\pi_{rf}^c = \frac{1}{\gamma}\kappa_1^c(\gamma - \gamma_{tan})$ and $\gamma_{tan} > \gamma_{ew}$ by assumption.

Proof of Proposition 2.10

The proof of parts 2 and 3 of the proposition is direct by taking the limit of $\boldsymbol{\kappa}^u$ and $\boldsymbol{\kappa}^c$ in (2.13) and (2.14) as $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$. Therefore, we turn to the proof of part 1 of the proposition. The asymptotic expectation of the out-of-sample mean return and variance of $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ can also be directly obtained by taking the limit in (2.11). We are left with the asymptotic covariance matrix of the out-of-sample mean return and variance, which are

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu} = \kappa_1 \hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\mu} + \kappa_2 \mu_{ew}, \tag{2.182}$$

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) = \kappa_1^2 \hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^\star + \kappa_2^2 \sigma_{ew}^2 + 2\kappa_1\kappa_2 \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^\star, \tag{2.183}$$

where $\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ in (2.5) is the SMV portfolio, which is unbiased and has a covariance matrix given by (2.109). Kan et al. (2024a, Theorem 3.2) implies that the asymptotic joint distribution of $(\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\mu}, \hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^\star, \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^\star)$ is normal. Therefore, using the delta method, the asymptotic joint distribution of $(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}, \hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}))$ is normal as well. Using the formulas for moments of linear and quadratic forms in normal random variables (Rencher and Schaalje, 2008), we can then find the three entries of the asymptotic covariance matrix of $(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}, \hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}))$.

First, the asymptotic variance of $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}$ as $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$ is

$$\lim_{N/T\to\phi}(T - N)\mathbb{V}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}\right] = \kappa_1^2 \lim_{N/T\to\phi}(T - N)\boldsymbol{\mu}'\mathbb{V}[\hat{\boldsymbol{w}}^\star]\boldsymbol{\mu} = \frac{\kappa_1^2}{\gamma^2}\theta^2(1 + 2\theta^2), \tag{2.184}$$

which corresponds to $\boldsymbol{V}_{11}$ in (2.60).

Second, the asymptotic covariance between $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}$ and $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ as $N, T \to \infty$ and $N/T \to \phi \in (0,1)$ is

$$\begin{aligned}&\lim_{N/T\to\phi}(T-N)\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}, \hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\right]\\ &= \lim_{N/T\to\phi}(T-N)\left(2\kappa_1^2\kappa_2\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\mu}, \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right] + \kappa_1^3\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\mu}, \hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right]\right), \end{aligned} \tag{2.185}$$

where, given $\mathbb{E}[\hat{\boldsymbol{w}}^{\star}]$ and $\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]$,

$$\lim_{N/T\to\phi}(T-N)\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\mu}, \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right] = \lim_{N/T\to\phi}(T-N)\boldsymbol{\mu}'\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]\boldsymbol{\Sigma}\boldsymbol{w}_{ew} = \frac{\mu_{ew}}{\gamma^2}(1+2\theta^2), \tag{2.186}$$

$$\lim_{N/T\to\phi}(T-N)\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\mu}, \hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right] = \lim_{N/T\to\phi}2(T-N)\boldsymbol{\mu}'\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]\boldsymbol{\Sigma}\mathbb{E}[\hat{\boldsymbol{w}}^{\star}] = \frac{2\theta^2}{\gamma^3}(1+2\theta^2). \tag{2.187}$$

Plugging (2.186)–(2.187) into (2.185) yields

$$\lim_{N/T\to\phi}(T-N)\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}, \hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\right] = \frac{2\kappa_1^2}{\gamma^2}(1+2\theta^2)\left(\frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew}\right), \tag{2.188}$$

which corresponds to $\boldsymbol{V}_{12} = \boldsymbol{V}_{21}$ in (2.61).

Third, the asymptotic variance of $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$ as $N, T \to \infty$ and $N/T \to \phi \in (0,1)$ is

$$\begin{aligned}&\lim_{N/T\to\phi}(T-N)\mathbb{V}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\right]\\ &= \lim_{N/T\to\phi}(T-N)\left(\kappa_1^4\mathbb{V}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right] + 4\kappa_1^2\kappa_2^2\boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]\boldsymbol{\Sigma}\boldsymbol{w}_{ew} + 4\kappa_1^3\kappa_2\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}, \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right]\right), \end{aligned} \tag{2.189}$$

where, given $\mathbb{E}[\hat{\boldsymbol{w}}^{\star}]$ and $\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]$,

$$\begin{aligned}&\lim_{N/T\to\phi}(T-N)\mathbb{V}[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}] = \lim_{N/T\to\phi}(T-N)\left(2\text{Tr}((\boldsymbol{\Sigma}\mathbb{V}[\hat{\boldsymbol{w}}^{\star}])^2) + 4\mathbb{E}[\hat{\boldsymbol{w}}^{\star}]'\boldsymbol{\Sigma}\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]\boldsymbol{\Sigma}\mathbb{E}[\hat{\boldsymbol{w}}^{\star}]\right)\\ &= \frac{2}{\gamma^4}(\theta^4(7+\phi) + 2\theta^2(2+\phi) + \phi), \end{aligned} \tag{2.190}$$

$$\lim_{N/T\to\phi}(T-N)\boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]\boldsymbol{\Sigma}\boldsymbol{w}_{ew} = \frac{\sigma_{ew}^2}{\gamma^2}(1+\theta_{ew}^2+\theta^2), \tag{2.191}$$

$$\lim_{N/T\to\phi}(T-N)\mathbb{C}\text{ov}\left[\hat{\boldsymbol{w}}^{\star\prime}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}, \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^{\star}\right] = \lim_{N/T\to\phi}2(T-N)\boldsymbol{w}_{ew}'\boldsymbol{\Sigma}\mathbb{V}[\hat{\boldsymbol{w}}^{\star}]\boldsymbol{\Sigma}\mathbb{E}[\hat{\boldsymbol{w}}^{\star}] = \frac{2\mu_{ew}}{\gamma^3}(1+2\theta^2). \tag{2.192}$$

Plugging (2.190)–(2.192) into (2.189) yields

$$\begin{aligned}&\lim_{N/T\to\phi}(T-N)\mathbb{V}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})\right]\\ &= \frac{2\kappa_1^2}{\gamma^2}\left(\frac{\kappa_1^2}{\gamma^2}\left(\theta^4(7+\phi) + 2\theta^2(2+\phi) + \phi\right) + \frac{4\kappa_1\kappa_2}{\gamma}\mu_{ew}(1+2\theta^2) + 2\kappa_2^2\sigma_{ew}^2(1+\theta_{ew}^2+\theta^2)\right), \end{aligned} \tag{2.193}$$

which corresponds to $\boldsymbol{V}_{22}$ in (2.62).

Proof of Proposition 2.11

Part 1. The SMV portfolio $\hat{\boldsymbol{w}}^\star$ is unbiased, and the SGMV portfolio $\hat{\boldsymbol{w}}_g$ too (Okhrin and Schmid, 2006). Therefore, the expected out-of-sample mean return of the portfolio combination (2.67) is

$$\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = \frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_g. \tag{2.194}$$

The expected out-of-sample variance decomposes as

$$\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = \frac{\kappa_1^2}{\gamma^2}(\theta^2+d) + \kappa_2^2\mathbb{E}\left[\hat{\boldsymbol{w}}_g'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] + \frac{2\kappa_1\kappa_2}{\gamma}\mathbb{E}\left[\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right]. \tag{2.195}$$

From Kan et al. (2021, Lemma 1), it holds that

$$\mathbb{E}\left[\hat{\boldsymbol{w}}_g'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] = c_1\sigma_g^2, \tag{2.196}$$

where the constant c_1 is defined in (2.70). Therefore, it remains to evaluate the expectation

$$\mathbb{E}\left[\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] = \mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}\right]. \tag{2.197}$$

In the following lemma, we prove that this expectation is equal to $c_1\mu_g$.

Lemma 2.1. $\mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}\right] = c_1\mu_g$, *where* c_1 *is defined in* (2.70).

We prove this lemma in the next section. Combining Equations (2.195)–(2.196) and Lemma 2.1, the expected out-of-sample variance is

$$\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = \frac{\kappa_1^2}{\gamma^2}(\theta^2+d) + c_1\kappa_2^2\sigma_g^2 + \frac{2c_1\kappa_1\kappa_2}{\gamma}\mu_g, \tag{2.198}$$

which results in the EU formula in (2.69).

Part 2. The unconstrained combination coefficients are found by setting the derivative of (2.69) with respect to κ_1 and κ_2 equal to zero and solving the resulting linear system of two equations with two unknowns.

Part 4. The EU of the unconstrained strategies is found by plugging the unconstrained coefficients (κ_1, κ_2) in (2.71) into the EU formula (2.69) after some developments.

Proof of Lemma 2.1

Proving the lemma amounts to show that

$$\mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{1}_N}\right] = \frac{T(T-2)\mu_g}{(T-N-1)(T-N-2)}, \tag{2.199}$$

where $\widehat{\boldsymbol{\Sigma}}_{ml} = \frac{T-N-2}{T}\widehat{\boldsymbol{\Sigma}}$ is the maximum-likelihood estimator of $\boldsymbol{\Sigma}$.

Let $\boldsymbol{P}$ be an $N \times N$ orthonormal matrix whose first two columns are

$$\boldsymbol{\nu} = \sigma_g \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathbf{1}_N, \tag{2.200}$$

$$\boldsymbol{\eta} = \frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{\mu} - \mu_g \mathbf{1}_N)}{\psi_g}, \tag{2.201}$$

where ψ_g is defined in (2.68). Let

$$\boldsymbol{z} = \sqrt{T}\boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\hat{\boldsymbol{\mu}} \sim \mathcal{N}(\boldsymbol{\mu}_z, \boldsymbol{I}_N), \tag{2.202}$$

$$\boldsymbol{W} = T\boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_{ml}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{P} \sim \mathcal{W}_N(T-1, \boldsymbol{I}_N), \tag{2.203}$$

and they are independent of each other. We can show that

$$\boldsymbol{\mu}_z = \sqrt{T}\boldsymbol{P}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} = \sqrt{T}\theta_g \boldsymbol{e}_1 + \sqrt{T}\psi_{ew}\boldsymbol{e}_2, \tag{2.204}$$

where $\boldsymbol{e}_i$ is the ith column of the identity matrix $\boldsymbol{I}_N$. Using $\boldsymbol{W}$ and $\boldsymbol{z}$, we can write

$$\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\hat{\boldsymbol{\mu}} = \frac{T^{\frac{3}{2}}}{\sigma_g}\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{z}, \tag{2.205}$$

$$\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{1}_N = \frac{T\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1}{\sigma_g^2}. \tag{2.206}$$

It follows that

$$\frac{\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}_{ml}^{-1}\mathbf{1}_N} = \frac{\sqrt{T}\sigma_g\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{z}}{\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1} =: q. \tag{2.207}$$

We are interested in obtaining the exact mean of q, which is equal to

$$\mathbb{E}[q] = T\mu_g\mathbb{E}\left[\frac{\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{e}_1}{\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1}\right] + T\sigma_g\psi_g\mathbb{E}\left[\frac{\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{e}_2}{\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1}\right]. \tag{2.208}$$

Partition $\boldsymbol{W}$ into four blocks such that $\boldsymbol{W}_{11}$ is its (1,1)th element. Using Muirhead (1982, Theorem 3.2.10), we can show that

$$x := (\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1)^{-1} = \boldsymbol{W}_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{21} \sim \chi^2_{T-N}, \tag{2.209}$$

$$\boldsymbol{y} := -\boldsymbol{W}_{22}^{-\frac{1}{2}}\boldsymbol{W}_{21} \sim \mathcal{N}(\mathbf{0}_{N-1}, \boldsymbol{I}_{N-1}), \tag{2.210}$$

$$\boldsymbol{W}_{22} \sim \mathcal{W}_{N-1}(T-1, \boldsymbol{I}_{N-1}), \tag{2.211}$$

and they are independent of each other. Let $\boldsymbol{Q} = [\boldsymbol{e}_2, \ldots, \boldsymbol{e}_N]$. From the partition matrix inverse formula, we can verify that

$$\boldsymbol{Q}'\boldsymbol{W}^{-1}\boldsymbol{e}_1 = \frac{\boldsymbol{W}_{22}^{-\frac{1}{2}}\boldsymbol{y}}{x}, \tag{2.212}$$

$$\boldsymbol{Q}'\boldsymbol{W}^{-1}\boldsymbol{Q} = \boldsymbol{W}_{22}^{-1} + \frac{\boldsymbol{W}_{22}^{-\frac{1}{2}}\boldsymbol{y}\boldsymbol{y}'\boldsymbol{W}_{22}^{-\frac{1}{2}}}{x}, \tag{2.213}$$

so we have

$$\boldsymbol{e}_2'\boldsymbol{W}^{-1}\boldsymbol{e}_1 = \frac{[1,\ \mathbf{0}_{N-2}']\boldsymbol{W}_{22}^{-\frac{1}{2}}\boldsymbol{y}}{x}. \tag{2.214}$$

It follows that

$$\begin{aligned}
\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{e}_1 &= \boldsymbol{e}_1'\boldsymbol{W}^{-1}[\boldsymbol{e}_1,\ \boldsymbol{Q}][\boldsymbol{e}_1,\ \boldsymbol{Q}]'\boldsymbol{W}^{-1}\boldsymbol{e}_1 \\
&= (\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1)^2 + \boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{W}^{-1}\boldsymbol{e}_1 \\
&= \frac{1+\boldsymbol{y}'\boldsymbol{W}_{22}^{-1}\boldsymbol{y}}{x^2}, \qquad (2.215) \\
\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{e}_2 &= \boldsymbol{e}_1'\boldsymbol{W}^{-1}[\boldsymbol{e}_1,\ \boldsymbol{Q}][\boldsymbol{e}_1,\ \boldsymbol{Q}]'\boldsymbol{W}^{-1}\boldsymbol{e}_2 \\
&= (\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1)(\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_2) + \boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{W}^{-1}\boldsymbol{e}_2 \\
&= \frac{[1,\ \boldsymbol{0}_{N-2}']\boldsymbol{W}_{22}^{-\frac{1}{2}}\boldsymbol{y}}{x^2} + \frac{\boldsymbol{y}'\boldsymbol{W}_{22}^{-\frac{1}{2}}}{x}\left(\boldsymbol{W}_{22}^{-1} + \frac{\boldsymbol{W}_{22}^{-\frac{1}{2}}\boldsymbol{y}\boldsymbol{y}'\boldsymbol{W}_{22}^{-\frac{1}{2}}}{x}\right)\begin{bmatrix}1\\ \boldsymbol{0}_{N-2}\end{bmatrix}. \qquad (2.216)
\end{aligned}$$

Using the above expressions, we obtain

$$\begin{aligned}
\mathbb{E}\left[\frac{\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{e}_1}{\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1}\right] &= \mathbb{E}\left[\frac{1+\boldsymbol{y}'\boldsymbol{W}_{22}^{-1}\boldsymbol{y}}{x}\right] \\
&= \mathbb{E}[1+\boldsymbol{y}'\boldsymbol{W}_{22}^{-1}\boldsymbol{y}]\mathbb{E}[x^{-1}] \\
&= \frac{1+\frac{N-1}{T-N-1}}{T-N-2} \\
&= \frac{T-2}{(T-N-1)(T-N-2)}, \qquad (2.217) \\
E\left[\frac{\boldsymbol{e}_1'\boldsymbol{W}^{-2}\boldsymbol{e}_2}{\boldsymbol{e}_1'\boldsymbol{W}^{-1}\boldsymbol{e}_1}\right] &= 0, \qquad (2.218)
\end{aligned}$$

and therefore we have from (2.208) that, as desired,

$$\mathbb{E}[q] = \frac{T(T-2)\mu_g}{(T-N-1)(T-N-2)}. \qquad (2.219)$$

Proof of Proposition 2.12

The two combination coefficients are κ_1^u in (2.71) and $\hat{\kappa}_2^u$ in (2.73). Because $\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$ is unbiased, $\hat{\kappa}_2^u$ is unbiased as well, and the expected out-of-sample mean return of the estimated unconstrained strategy is the same as that when κ_2^u is known:

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u, \hat{\kappa}_2^u)'\boldsymbol{\mu}\right] = \frac{\kappa_1^u}{\gamma}\theta^2 + \kappa_2^u\mu_g. \qquad (2.220)$$

In comparison, the expected out-of-sample variance is larger than that when κ_2^u is known. Specifically,

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u, \hat{\kappa}_2^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\kappa_1^u, \hat{\kappa}_2^u)\right] = \frac{(\kappa_1^u)^2}{\gamma^2}(\theta^2+d) + \mathbb{E}\left[(\hat{\kappa}_2^u)^2\hat{\boldsymbol{w}}_g'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] + \frac{2\kappa_1^u}{\gamma}\mathbb{E}\left[\hat{\kappa}_2^u\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right]. \qquad (2.221)$$

Applying (2.107), $\mathbb{E}\left[\hat{\boldsymbol{w}}_g'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] = c_1\sigma_g^2$ in (2.196), and Lemma 2.1, we have that

$$\begin{aligned}\mathbb{E}\left[(\hat{\kappa}_2^u)^2\hat{\boldsymbol{w}}_g'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] &= \frac{(1/c_1-\kappa_1^u)^2}{\gamma^2}\mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{(\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N)^2}\right] \\ &= c_1\sigma_g^2(\kappa_2^u)^2 + \frac{c_1(1/c_1-\kappa_1^u)^2}{\gamma^2 T} \end{aligned} \tag{2.222}$$

and

$$\begin{aligned}\frac{2\kappa_1^u}{\gamma}\mathbb{E}\left[\hat{\kappa}_2^u\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_g\right] &= \frac{2\kappa_1^u(1/c_1-\kappa_1^u)}{\gamma^2}\mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}\right] \\ &= \frac{2c_1\kappa_1^u\kappa_2^u}{\gamma}\mu_g + \frac{2c_1\kappa_1^u(1/c_1-\kappa_1^u)}{T\gamma^2}. \end{aligned} \tag{2.223}$$

Therefore, the expected out-of-sample variance is

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^u,\hat{\kappa}_2^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\kappa_1^u,\hat{\kappa}_2^u)\right] = \mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)\right] + \frac{c_1(1/c_1^2-(\kappa_1^u)^2)}{\gamma^2 T}, \tag{2.224}$$

where $\mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)\right] = \frac{(\kappa_1^u)^2}{\gamma^2}(\theta^2+d) + c_1\sigma_g^2(\kappa_2^u)^2 + \frac{2c_1\kappa_1^u\kappa_2^u}{\gamma}\mu_g$ is the expected out-of-sample variance when κ_2^u is known. Finally, the EU loss relative to $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)$ when κ_2^u is estimated by $\hat{\kappa}_2^u$ is given by (2.74).

Proof of Proposition 2.13

The mean and variance of $\tilde{\kappa}_2^u(\delta)$ are

$$\mathbb{E}[\tilde{\kappa}_2^u(\delta)] = (1-\delta)\kappa_2^u + \delta(1/c_1-\kappa_1^u), \tag{2.225}$$

$$\mathbb{V}[\tilde{\kappa}_2^u(\delta)] = \frac{(1-\delta)^2(1/c_1-\kappa_1^u)^2}{\gamma^2\sigma_g^2 T}. \tag{2.226}$$

Therefore, the mean squared error of $\tilde{\kappa}_2^u(\delta)$ is

$$(\mathbb{E}[\tilde{\kappa}_2^u(\delta)]-\kappa_2^u)^2 + \mathbb{V}[\tilde{\kappa}_2^u(\delta)] = \frac{(1/c_1-\kappa_1^u)^2}{\gamma^2}\left(\delta^2(\gamma-\gamma_{tan})^2 + \frac{(1-\delta)^2}{\sigma_g^2 T}\right). \tag{2.227}$$

Setting the derivative of (2.227) equal to zero delivers the optimal $\delta^\star$ in (2.13).

Proof of Proposition 2.14

The EU of the combination of the SMV and SGMV portfolios,

$$\hat{\boldsymbol{w}}(\boldsymbol{\kappa}) = \frac{\kappa_1}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} + \frac{\kappa_2}{\gamma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N, \tag{2.228}$$

is obtained by plugging $\kappa_2 = 0$ into (2.90), which gives

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = \frac{\kappa_1}{\gamma}\theta^2 + \frac{\kappa_2}{\gamma}\gamma_{tan} - \frac{c\gamma}{2}\left(\frac{\kappa_1^2}{\gamma^2}\left(\theta^2+\frac{N}{T}\right) + \frac{\kappa_2^2}{\gamma^2}\frac{1}{\sigma_g^2} + \frac{2\kappa_1\kappa_2}{\gamma^2}\gamma_{tan}\right). \tag{2.229}$$

The EU of the portfolio combination of Kan and Zhou (2007) in (2.83) is then obtained by plugging κ_1^{kz} and κ_2^{kz} in (2.82) into the EU formula (2.229) after some developments.

Concerning the fully invested combination of the SMV and SGMV portfolios, $\hat{\boldsymbol{w}}(\kappa)$ in (2.79), we have from Kan et al. (2021) that the EU depends on κ as

$$EU(\hat{\boldsymbol{w}}(\kappa)) = \mu_g - \frac{\gamma}{2}c_1\sigma_g^2 + \frac{1}{\gamma}\frac{T-N-2}{T-N-1}\left(\kappa\psi_g^2 - \frac{\kappa^2}{2}\left(\psi_g^2 + \frac{N-1}{T}\right)\frac{(T-2)(T-N-2)}{(T-N)(T-N-3)}\right). \tag{2.230}$$

The EU of the optimal portfolio combination is then obtained by plugging κ^{kwz} in (2.80) into the EU formula (2.230).

Proof of Proposition 2.15

Part 1. The two combination coefficients are κ_1^{kz} in (2.82) and $\hat{\kappa}_2^{kz}$ in (2.85). Because $\hat{\mu}_g$ is unbiased, $\hat{\kappa}_2^{kz}$ is unbiased as well, and the expected out-of-sample mean return of the estimated unconstrained strategy is the same as that when κ_2^{kz} is known in (2.229):

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^{kz}, \hat{\kappa}_2^{kz})'\boldsymbol{\mu}\right] = \frac{\kappa_1^{kz}}{\gamma}\theta^2 + \kappa_2^{kz}\gamma_{tan}. \tag{2.231}$$

In comparison, the expected out-of-sample variance is larger than that when κ_2^{kz} is known. Specifically,

$$\begin{aligned}\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^{kz}, \hat{\kappa}_2^{kz})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\kappa_1^{kz}, \hat{\kappa}_2^{kz})\right] &= \frac{(\kappa_1^{kz})^2}{\gamma^2}(\theta^2 + d) + \frac{1}{\gamma^2}\mathbb{E}\left[(\hat{\kappa}_2^{kz})^2\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\right] \\ &\quad + \frac{2\kappa_1^{kz}}{\gamma^2}\mathbb{E}\left[\hat{\kappa}_2^{kz}\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\right].\end{aligned} \tag{2.232}$$

Applying (2.107) and $\mathbb{E}\left[\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\right] = c\boldsymbol{\Sigma}^{-1}$, we have that

$$\begin{aligned}\frac{1}{\gamma^2}\mathbb{E}\left[(\hat{\kappa}_2^{kz})^2\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\right] &= \frac{(1/c - \kappa_1^{kz})^2}{\gamma^2}\mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{(\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N)^2}\right] \\ &= \frac{c(\kappa_2^{kz})^2}{\gamma^2\sigma_g^2} + \frac{c(1/c - \kappa_1^{kz})^2}{\gamma^2 T}\end{aligned} \tag{2.233}$$

and

$$\begin{aligned}\frac{2\kappa_1^{kz}}{\gamma^2}\mathbb{E}\left[\hat{\kappa}_2^{kz}\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}\right] &= \frac{2\kappa_1^{kz}(1/c - \kappa_1^{kz})}{\gamma^2}\mathbb{E}\left[\frac{\hat{\boldsymbol{\mu}}'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_N'\boldsymbol{\Sigma}^{-1}\mathbf{1}_N}\right] \\ &= \frac{2c\kappa_1^{kz}\kappa_2^{kz}}{\gamma^2}\gamma_{tan} + \frac{2c\kappa_1^{kz}(1/c - \kappa_1^{kz})}{\gamma^2 T}.\end{aligned} \tag{2.234}$$

Therefore, the expected out-of-sample variance is

$$\mathbb{E}\left[\hat{\boldsymbol{w}}(\kappa_1^{kz}, \hat{\kappa}_2^{kz})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\kappa_1^{kz}, \hat{\kappa}_2^{kz})\right] = \mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})\right] + \frac{c(1/c^2 - (\kappa_1^{kz})^2)}{\gamma^2 T}, \tag{2.235}$$

where $\mathbb{E}\left[\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})\right] = c\left(\frac{(\kappa_1^{kz})^2}{\gamma^2}(\theta^2+\frac{N}{T})+\frac{(\kappa_2^{kz})^2}{\gamma^2}\frac{1}{\sigma_g^2}+\frac{2\kappa_1^{kz}\kappa_2^{kz}}{\gamma^2}\gamma_{tan}\right)$ is the expected out-of-sample variance when κ_2^{kz} is known. Finally, the loss in EU relative to $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^{kz})$ when κ_2^{kz} is estimated by $\hat{\kappa}_2^{kz}$ is given by (2.86).

Part 2. Using Equations (2.84) and (2.86), the fully invested strategy of Kan et al. (2021) delivers a larger EU than the estimated three-fund strategy of Kan and Zhou (2007), $EU(\hat{\boldsymbol{w}}(\kappa^{kwz})) \geq EU(\hat{\boldsymbol{w}}(\kappa_1^{kz},\hat{\kappa}_2^{kz}))$, when

$$\mu_g - \frac{\gamma}{2}c_1\sigma_g^2 + \frac{T-N-2}{T-N-1}\frac{\psi_g^2}{2\gamma}\kappa^{kwz} \geq U^\star - \frac{d}{2\gamma}\kappa_1^{kz} - \frac{c-1}{2\gamma}\gamma_{tan}\kappa_2^{kz} - \frac{c(1/c^2-(\kappa_1^{kz})^2)}{2\gamma T}. \tag{2.236}$$

After some developments, we find that inequality (2.236) is equivalent to

$$\gamma^2[-c_1\sigma_g^2] + \gamma[2\mu_g] + \frac{b-\theta_g^2}{c_1} \geq 0, \tag{2.237}$$

where b is defined in (2.88). The discriminant of this second-degree polynomial is

$$\Delta = 4\sigma_g^2 b. \tag{2.238}$$

If $b<0$, the polynomial (2.237) has no real roots and the Kan-Zhou three-fund strategy outperforms for any γ. Otherwise, the two real roots of the polynomial are those in (2.87).

Proof of Proposition 2.16

Part 1. Because the four-fund portfolio combination in (2.89) is unbiased, the out-of-sample mean return is unbiased as well,

$$\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\mu}] = \frac{\kappa_1}{\gamma}\theta^2 + \kappa_2\mu_{ew} + \frac{\kappa_3}{\gamma}\gamma_{tan}. \tag{2.239}$$

To find the expected out-of-sample variance in (2.107), we need the covariance matrix of $\hat{\boldsymbol{w}}(\boldsymbol{\kappa})$. Using the result in Javed et al. (2021, Corollary 2.2), it is

$$\begin{aligned}\mathbb{V}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = {} & c\frac{\kappa_1^2}{\gamma^2}\left(\frac{\theta^2}{T-2}+\frac{1}{T}\right)\boldsymbol{\Sigma}^{-1} \\ & + c_2\left(\frac{\kappa_1^2}{\gamma^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1} + \frac{\kappa_3^2}{\gamma^2}\boldsymbol{\Sigma}^{-1}\mathbf{1}_N\mathbf{1}_N'\boldsymbol{\Sigma}^{-1} + \frac{\kappa_1\kappa_3}{\gamma^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}\mathbf{1}'+\mathbf{1}\boldsymbol{\mu}')\boldsymbol{\Sigma}^{-1}\right),\end{aligned} \tag{2.240}$$

where

$$c_2 = \frac{c(T-N)}{(T-2)(T-N-2)} = \frac{T-N}{(T-N-1)(T-N-4)}. \tag{2.241}$$

The expected out-of-sample variance is then given by (2.107), where

$$\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})]'\boldsymbol{\Sigma}\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = \frac{\kappa_1^2}{\gamma^2}\theta^2 + \kappa_2^2\sigma_{ew}^2 + \frac{\kappa_3^2}{\gamma^2}\frac{1}{\sigma_g^2} + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew} + \frac{2\kappa_1\kappa_3}{\gamma^2}\gamma_{tan} + \frac{2\kappa_2\kappa_3}{\gamma}, \tag{2.242}$$

$$\text{Trace}(\boldsymbol{\Sigma}\mathbb{V}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})]) = c\frac{\kappa_1^2}{\gamma^2}\frac{N}{T} + \left(c_2+\frac{cN}{T-2}\right)\left(\frac{\kappa_1^2}{\gamma^2}\theta^2 + \frac{\kappa_3^2}{\gamma^2}\frac{1}{\sigma_g^2} + \frac{2\kappa_1\kappa_3}{\gamma^2}\gamma_{tan}\right). \tag{2.243}$$

Noticing that $1 + c_2 + \frac{cN}{T-2} = c$, the expected out-of-sample variance is

$$\mathbb{E}[\hat{\boldsymbol{w}}(\boldsymbol{\kappa})'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}(\boldsymbol{\kappa})] = c\left(\frac{\kappa_1^2}{\gamma^2}\left(\theta^2 + \frac{N}{T}\right) + \frac{\kappa_3^2}{\gamma^2}\frac{1}{\sigma_g^2} + \frac{2\kappa_1\kappa_3}{\gamma^2}\gamma_{tan}\right) + \kappa_2^2\sigma_{ew}^2 + \frac{2\kappa_1\kappa_2}{\gamma}\mu_{ew} + \frac{2\kappa_2\kappa_3}{\gamma}, \tag{2.244}$$

and thus the EU is given by (2.90).

Part 2. The EU of the four-fund portfolio in (2.90) can be rewritten in matrix format as

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = \boldsymbol{\kappa}'\boldsymbol{\eta} - \frac{\gamma}{2}\boldsymbol{\kappa}'\boldsymbol{S}\boldsymbol{\kappa}, \tag{2.245}$$

where

$$\boldsymbol{\eta} = \begin{pmatrix} \theta^2/\gamma \\ \mu_{ew} \\ \gamma_{tan}/\gamma \end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix} \frac{c}{\gamma^2}\left(\theta^2 + \frac{N}{T}\right) & \frac{\mu_{ew}}{\gamma} & \frac{c\gamma_{tan}}{\gamma^2} \\ \frac{\mu_{ew}}{\gamma} & \sigma_{ew}^2 & \frac{1}{\gamma} \\ \frac{c\gamma_{tan}}{\gamma^2} & \frac{1}{\gamma} & \frac{c}{\gamma^2\sigma_g^2} \end{pmatrix}. \tag{2.246}$$

The matrix $\boldsymbol{S}$ is positive definite and thus invertible. This is because the covariance matrix $\boldsymbol{\Sigma}$ is assumed positive definite, and thus any expected out-of-sample variance is strictly positive, $\boldsymbol{\kappa}'\boldsymbol{S}\boldsymbol{\kappa} > 0$ for any $\boldsymbol{\kappa} \neq \mathbf{0}_2$. As a result, the combination coefficients $\boldsymbol{\kappa}$ maximizing the EU in (2.90) are

$$\boldsymbol{\kappa}^u = \arg\max_{\boldsymbol{\kappa}} \ \boldsymbol{\kappa}'\boldsymbol{\eta} - \frac{\gamma}{2}\boldsymbol{\kappa}'\boldsymbol{S}\boldsymbol{\kappa} = \frac{1}{\gamma}\boldsymbol{S}^{-1}\boldsymbol{\eta}, \tag{2.247}$$

which from (2.246) simplify to (2.91)–(2.93) after some developments.

Part 3. Just like the utility of the optimal mean-variance portfolio $\boldsymbol{w}^\star$ is $U^\star = \theta^2/(2\gamma) = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/(2\gamma)$, the EU achieved by the optimal four-fund portfolio combination $\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)$ obtained from the combination coefficients (2.247) is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)) = \frac{\boldsymbol{\eta}'\boldsymbol{S}^{-1}\boldsymbol{\eta}}{2\gamma}, \tag{2.248}$$

where $\boldsymbol{\eta}$ and $\boldsymbol{S}$ are defined in (2.246). After some developments, we can show that (2.248) corresponds to (2.95).

Proof of Proposition 2.17

We begin with the correlation between the out-of-sample return of the SMV and EW portfolios, which is

$$\mathbb{C}\text{orr}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}] = \frac{\mathbb{C}\text{ov}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}]}{\sqrt{\mathbb{V}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}]\mathbb{V}[\boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}]}}. \tag{2.249}$$

We must evaluate the three terms appearing in (2.249). It is clear that $\mathbb{V}[\boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}] = \sigma_{ew}^2$. Moreover, applying (2.107),

$$\mathbb{C}\text{ov}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}] = \frac{1}{\gamma}\mathbb{E}[\boldsymbol{r}_{T+1}'\boldsymbol{w}_{ew}\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{r}_{T+1}] - \mathbb{E}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}]\mathbb{E}[\boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}] = \mu_{ew}/\gamma. \tag{2.250}$$

Finally, from the law of total variance,

$$\mathbb{V}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}] = \mathbb{E}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^\star] + \boldsymbol{\mu}'\mathbb{V}[\hat{\boldsymbol{w}}^\star]\boldsymbol{\mu}, \tag{2.251}$$

where $\mathbb{E}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}^\star] = \frac{\theta^2+d}{\gamma^2}$ from the proof of Proposition 2.1 and $\mathbb{V}[\hat{\boldsymbol{w}}^\star]$ is given by (2.109). This gives

$$\mathbb{V}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}] = \frac{1}{\gamma^2}\left(\frac{c}{T}(\theta^2(T+1)+N) + \frac{2\theta^4}{T-N-4}\right). \tag{2.252}$$

Plugging $\mathbb{C}\text{ov}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \boldsymbol{w}_{ew}'\boldsymbol{r}_{T+1}]$, $\mathbb{V}[\boldsymbol{w}_{ew}'R_{T+1}]$, and $\mathbb{V}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}]$ into (2.249) gives (2.96). This correlation tends to θ_{ew}/θ as $T \to \infty$.

Let us now consider the correlation between the out-of-sample return of the SMV and SGMV portfolios, which is

$$\mathbb{C}\text{orr}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}] = \frac{\mathbb{C}\text{ov}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}]}{\sqrt{\mathbb{V}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}]\mathbb{V}[\hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}]}}. \tag{2.253}$$

Applying (2.107) and $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}\widehat{\boldsymbol{\Sigma}}^{-1}] = c\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}$ for any constant matrix $\boldsymbol{A}$, we have that

$$\begin{aligned}\mathbb{C}\text{ov}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}] &= \frac{1}{\gamma}\mathbb{E}[\boldsymbol{r}_{T+1}'\hat{\boldsymbol{w}}_g\hat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{r}_{T+1}] - \mathbb{E}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}]\mathbb{E}[\hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}] \\ &= \frac{\gamma_{tan}}{\gamma^2}(c + (c-1)\theta^2). \end{aligned} \tag{2.254}$$

Similarly, we can show that

$$\mathbb{V}[\hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}] = \frac{1}{\gamma^2}\left(\mathbb{E}[\boldsymbol{r}_{T+1}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N\mathbf{1}_N'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{r}_{T+1}] - \gamma_{tan}^2\right) = \frac{1}{\gamma^2}\left(c/\sigma_g^2 + (c-1)\gamma_{tan}^2\right). \tag{2.255}$$

Finally, plugging $\mathbb{C}\text{ov}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}, \hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}]$, $\mathbb{V}[\hat{\boldsymbol{w}}_g'\boldsymbol{r}_{T+1}]$, and $\mathbb{V}[(\hat{\boldsymbol{w}}^\star)'\boldsymbol{r}_{T+1}]$ into (2.253) gives (2.97). This correlation tends to θ_g/θ as $T \to \infty$.

Proof of Proposition 2.18

Part 1. Given Equation (2.11), the ESR of the three-fund portfolio in (2.7) can be rewritten in matrix format as

$$ESR(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = \frac{\boldsymbol{\kappa}'\boldsymbol{\eta}}{\sqrt{\boldsymbol{\kappa}'\boldsymbol{S}\boldsymbol{\kappa}}}, \tag{2.256}$$

where

$$\boldsymbol{\eta} = \begin{pmatrix}\theta^2/\gamma \\ \mu_{ew}\end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix}\frac{c}{\gamma^2}\left(\theta^2 + \frac{N}{T}\right) & \frac{\mu_{ew}}{\gamma} \\ \frac{\mu_{ew}}{\gamma} & \sigma_{ew}^2\end{pmatrix}. \tag{2.257}$$

The matrix $\boldsymbol{S}$ is positive definite and thus invertible. This is because the covariance matrix $\boldsymbol{\Sigma}$ is assumed positive definite, and thus any expected out-of-sample variance is strictly

positive, $\boldsymbol{\kappa}'\boldsymbol{S}\boldsymbol{\kappa} > 0$ for any $\boldsymbol{\kappa} \neq \mathbf{0}_2$. As a result, just like any maximum-utility portfolio $\frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ provides the maximum Sharpe ratio $\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$, any vector of unconstrained combination coefficients $\boldsymbol{\kappa}^u = \frac{1}{\gamma}\boldsymbol{S}^{-1}\boldsymbol{\eta}$ provides the maximum ESR:

$$\max_{\boldsymbol{\kappa}} ESR(\hat{\boldsymbol{w}}(\boldsymbol{\kappa})) = ESR(\hat{\boldsymbol{w}}(\boldsymbol{\kappa}^u)) = \sqrt{\boldsymbol{\eta}'\boldsymbol{S}^{-1}\boldsymbol{\eta}}. \tag{2.258}$$

After some developments, we find that

$$\sqrt{\boldsymbol{\eta}'\boldsymbol{S}^{-1}\boldsymbol{\eta}} = \sqrt{\theta_{ew}^2 + (\theta^2 - \theta_{ew}^2)\kappa_1^u}, \tag{2.259}$$

which is in between θ_{ew} and θ because $\kappa_1^u \in [0, 1]$.

Part 2. The constrained combination coefficients maximizing the ESR in (2.100) are

$$\boldsymbol{\kappa} = \arg\max_{\boldsymbol{\kappa}'\mathbf{1}_N=1} \frac{\boldsymbol{\kappa}'\boldsymbol{\eta}}{\sqrt{\boldsymbol{\kappa}'\boldsymbol{S}\boldsymbol{\kappa}}} = \frac{\boldsymbol{S}^{-1}\boldsymbol{\eta}}{\mathbf{1}_N'\boldsymbol{S}^{-1}\boldsymbol{\eta}} = \left(\frac{\psi_{ew}^2}{\psi_{ew}^2 + \frac{\gamma_{ew}}{\gamma}d}, \frac{\frac{\gamma_{ew}}{\gamma}d}{\psi_{ew}^2 + \frac{\gamma_{ew}}{\gamma}d}\right), \tag{2.260}$$

and they correspond to the unconstrained combination coefficients in (2.13) for $\gamma = \gamma_{ew}$ and $\gamma = \theta^2/\mu_{ew}$, in which case the constrained strategy delivers the maximum ESR.

To find when the ESR of the constrained strategy $\hat{\boldsymbol{w}}^c$ is minimized, we first simplify its ESR by plugging (2.13) into (2.100), which gives

$$ESR(\hat{\boldsymbol{w}}^c) = \frac{d\mu_{ew}\gamma + \theta^2(\psi_{ew}^2 + \ell^2)}{\sqrt{(\psi_{ew}^2 + \ell^2 + d)(d\sigma_{ew}^2\gamma^2 + \theta^2(\psi_{ew}^2 + \ell^2))}}. \tag{2.261}$$

From (2.261), it is easy to check that

$$\lim_{\gamma\to 0} ESR(\hat{\boldsymbol{w}}^c) = \lim_{\gamma\to\infty} ESR(\hat{\boldsymbol{w}}^c) = \frac{\theta^2}{\sqrt{\theta^2 + d}}. \tag{2.262}$$

To see under which condition no value of γ between 0 and ∞ can give a smaller ESR than (2.262), we differentiate $ESR(\hat{\boldsymbol{w}}^c)$ with respect to γ:

$$\frac{\partial}{\partial\gamma} ESR(\hat{\boldsymbol{w}}^c) = \frac{d^2(\theta^2 - \mu_{ew}\gamma)(\mu_{ew} - \sigma_{ew}^2\gamma)(\theta^2 - \sigma_{ew}^2\gamma^2)}{\big((\psi_{ew}^2 + \ell^2 + d)(d\sigma_{ew}^2\gamma^2 + \theta^2(\psi_{ew}^2 + \ell^2))\big)^{3/2}}. \tag{2.263}$$

This derivative is equal to zero when $\gamma = \gamma_{ew}$ and $\gamma = \theta^2/\mu_{ew}$, which correspond to the two maxima identified above, and when $\gamma = \theta/\sigma_{ew}$, which corresponds to a local minimum. To conclude the proof, we must thus find when the value of $ESR(\hat{\boldsymbol{w}}^c)$ with $\gamma = \theta/\sigma_{ew}$ is larger than $\theta^2/\sqrt{\theta^2 + d}$ in (2.262). By plugging $\gamma = \theta/\sigma_{ew}$ into (2.261), we find that

$$ESR(\hat{\boldsymbol{w}}^c) = \theta\frac{\theta_{ew}d + 2\theta^2(\theta - \theta_{ew})}{\theta d + 2\theta^2(\theta - \theta_{ew})} \quad \text{when} \quad \gamma = \theta/\sigma_{ew}, \tag{2.264}$$

and after some developments this value is larger than $\theta^2/\sqrt{\theta^2 + d}$ when

$$d > \frac{\theta^2}{\theta_{ew}^2}(\theta - 3\theta_{ew})(\theta - \theta_{ew}). \tag{2.265}$$

Chapter 3

Optimal Portfolio Size under Parameter Uncertainty

Abstract

We introduce a method to determine the investor's optimal portfolio size that maximizes the expected out-of-sample utility. This portfolio size trades off between accessing investment opportunities and limiting the number of parameters to be estimated. Unlike sparse methods such as lasso that exclude assets during the optimization step, our approach fixes the optimal number of assets before computing the portfolio weights, which improves robustness and provides greater flexibility in practical implementations. Empirically, our restricted portfolios outperform their counterparts applied to all available assets. Our methodology renders portfolio theory valuable even when the dataset dimension and sample size are comparable.

Reference

This chapter is based on:

Lassance, N., Vanderveken, R. and Vrins, F. (2025a). Optimal portfolio size under parameter uncertainty. *SSRN working paper*.

3.1 Introduction

Given a chosen universe of M assets, conventional wisdom argues that an unconstrained rational investor should consider investing in all M assets, so as to diversify idiosyncratic risk and improve the efficient frontier. However, this is at odds with the concentrated portfolios typically observed in practice among individual investors (Campbell, 2006; Calvet et al., 2007; Goetzmann and Kumar, 2008) and institutions (Koijen and Yogo, 2019).[1] As reviewed by, e.g., Boyle et al. (2012), various papers aim to rationalize the concentrated portfolios of investors or to explain them with behavioral biases. Existing explanations consider, e.g., transaction costs, short-sale constraints, prospect theory, overconfidence, cost of information, geographical closeness and familiarity, private incentives, or preferences for higher moments.

Besides behavioral and financial arguments, another motivation for reducing portfolio size is *parameter uncertainty*, i.e., the parameters of the asset return distribution are unknown and must be estimated. In theory, the optimal portfolio can only benefit from increased investment opportunities associated to a higher portfolio size. In practice, however, more parameters and portfolio weights must be estimated, which increases estimation risk and hurts out-of-sample performance (DeMiguel et al., 2009b; Barroso and Saxena, 2021).

Sparse portfolio selection aims at alleviating parameter uncertainty precisely by reducing the portfolio size. This is typically achieved by imposing portfolio norm penalties, constraints based on lasso (DeMiguel et al., 2009a; Fan et al., 2012; Yen, 2016; Ao et al., 2019) or cardinality constraints (Gao and Li, 2013; Du et al., 2023). However, such methods suffer from five main drawbacks. First, the optimization programs must be solved in dimension M, hence, face large estimation risk, even if the output portfolio is ultimately of lower dimension. Second, they can be computationally intensive to implement especially in high dimension. For instance, cardinality constraints render the portfolio problem NP-hard. Third, they entangle the selection of the assets with the computation of portfolio weights in one single step. Therefore, sparse methods do not allow for different rebalancing frequencies for portfolio weights and the asset selection, or for flexibility in the asset selection. Fourth, they typically feature extra parameters such as a penalty strength often estimated by cross-validation, which is time consuming and adds an extra layer of estimation risk. Lastly, the resulting portfolio size N is implicit from the portfolio weights and does not have a clear meaning.

We propose a sequential *three-stage process* that addresses the main drawbacks of sparse portfolio selection listed above. First, select a portfolio strategy. In this chapter, we consider different combinations of mean-variance (MV), global-minimum-variance (GMV), and equally weighted (EW) portfolios. Second, for the chosen portfolio strategy, compute

[1] Institutional investors are typically more diversified than individual investors, but still, Koijen and Yogo (2019) find in their dataset of US stocks that the median institution held 67 stocks in the period 2015–2017.

an *optimal* portfolio size $N \leq M$. Third, select which N assets to invest in among the M available ones and compute the weights on the N assets according to the strategy in step one. Interestingly, Ao et al. (2019) propose a similar sequential process for reasons of computational efficiency, in which they select a chosen number of N assets to maximize the Sharpe ratio and then implement their MAXSER strategy on this subset of assets. Specifically, among the S&P500 constituents, they *arbitrarily* select $N = 50$ of them. In contrast, we show how to select an *optimal* N for several portfolio rules based on estimation risk considerations.

Two key questions remain in our three-stage process: How do we find the optimal portfolio size N? How do we then select the N assets? We focus on the first question. Specifically, we introduce a methodology for finding the optimal N and leave flexibility to the investor as to which N assets to select. In our empirical analysis, we evaluate the performance of our optimal N on 10 simple and sensible asset selection rules as an illustration.

To find the optimal N, we consider a classical setting in which investors are expected utility maximizers with MV preferences, there are no investment constraints, and returns are multivariate elliptically distributed. In this setting, parameter uncertainty stems from the unknown mean, covariance matrix, and fat tails of returns. We consider several MV portfolio rules and find, for each of them, the optimal portfolio size N that strikes a tradeoff between two opposite goals: increasing investment opportunities and reducing estimation risk. This N is *optimal* in the sense that it maximizes the *expected out-of-sample utility* (EU), the standard portfolio performance measure under parameter uncertainty (Kan and Zhou, 2007; Tu and Zhou, 2011; Kan et al., 2021), and it varies across portfolio rules depending on their estimation risk. We observe empirically that determining the optimal N using our theory and then selecting the N assets using different selection rules significantly outperforms investing in all M assets in about 90% of considered configurations.

Our approach has five main benefits compared to the aforementioned sparse methods. First, because we restrict the portfolio size *before* computing the portfolio weights, these weights depend on a limited number of parameters, and thus, face less estimation risk. Second, our approach is less computationally expensive because our optimal N can be found very efficiently and the portfolio weights are then computed on a universe of smaller dimension. Third, our optimal N depends on the data, the portfolio strategy, and the objective function, but is agnostic with respect to the selection rule determining which N assets to invest in, i.e., the investor has flexibility regarding asset selection. This allows different rebalancing frequencies for portfolio weights and asset selection, which is valuable

in practice.[2] Fourth, our approach does not require the calibration of any extra parameters such as a penalty strength. Lastly, the optimal N has a clear meaning from trading off between diversification benefits and estimation risk when maximizing expected utility.

Our theoretical analysis starts with the sample mean-variance (SMV) portfolio, which is optimal *in sample* but not out of sample. Assuming that asset returns are equicorrelated as advocated by Engle and Kelly (2012) and Clements et al. (2015),[3] we express the EU of this portfolio as a function of the portfolio size, the sample size, the correlation, the assets' Sharpe ratios, and three parameters that measure the impact of fat tails.[4] We show that because the SMV portfolio is highly sensitive to estimation errors, its optimal portfolio size N is typically very small.

As a remedy for the large estimation risk faced by the SMV portfolio, Kan and Zhou (2007) introduce a *two-fund rule* that scales down the SMV portfolio by combining it with the risk-free asset so as to maximize the EU; a portfolio rule that Kan and Lassance (2025) extend to the case with elliptical returns. Building on this research, we also derive the optimal N for the two-fund rule. Remarkably, for typical levels of rather high correlations in equity data, we can show that this optimal N is slightly below *half the sample size*. Given commonly used sample sizes (e.g., 120 months), this result means that it is optimal to hold a limited number of assets in the portfolio to optimize out-of-sample performance.

In addition to the two-fund rule, we derive the optimal N for *three-fund rules*, i.e., the combination of the SMV portfolio, the sample global minimum-variance (SGMV) portfolio, and the risk-free asset in Kan and Zhou (2007), DeMiguel et al. (2015), and Yuan and Zhou (2024), as well as the combination of the SMV portfolio, the equally weighted (EW) portfolio, and the risk-free asset in Tu and Zhou (2011), Kan and Wang (2023), and Lassance et al. (2024b). Our approach is flexible and can accommodate other portfolio rules whose EU has an analytical expression. Similarly to the two-fund rule, we find that the optimal N for these rules is slightly below half the sample size for typical levels of equity return correlations. These results extend the literature on portfolio choice with estimation risk by showing that it remains beneficial to substantially reduce the portfolio size even after optimally combining the SMV portfolio with robust strategies to reduce

[2]We rebalance the weights every month but the selected assets every year, which reduces asset turnover and outperforms selecting assets every month. Nonetheless, we show in Section 3.F.7 of the supplementary material that our optimal N outperforms investing in all M assets also when the selected assets change every month. One cannot achieve this with norm or cardinality constraints because the selection of assets and their weights is simultaneous.

[3]Equicorrelation covariance matrices achieve a satisfying tradeoff between specification and estimation error (Engle and Kelly, 2012) and allow us to express the EU of our different MV portfolio rules as an explicit function of N. We employ this assumption only for finding the optimal N, not the portfolio weights. We show that the optimal N found under the equicorrelation assumption is similar to that under a single-factor model assumption. Moreover, we find that our method performs well also when returns are not equicorrelated.

[4]Most of the literature on portfolio choice with estimation risk assumes multivariate normally distributed returns. Our incorporation of elliptical fat tails builds on the recent work by Kan and Lassance (2025).

estimation risk.

We test our theory first via a simulation experiment in which we evaluate the performance of our optimal N when (i) the optimal N is subject to estimation errors (bona fide estimator), (ii) asset returns are not equicorrelated, and (iii) the decision of which N assets to select out of the M available ones is done randomly. The main conclusion is that the two- and three-fund rules implemented with our optimal portfolio size N deliver an EU close to the maximum and substantially larger than that obtained when investing either in all M assets or in too few assets. This shows that our method delivers satisfying performance even when relaxing the assumptions underlying our theory, highlighting its practical relevance.

Finally, we turn to empirical data. We consider six datasets of characteristic and industry-sorted test assets, with M around 100. Using a rolling window of 120 months, we compare the out-of-sample utility net of costs of different strategies, including our optimally restricted two-fund and three-fund rules. For the latter, we propose 10 simple and sensible *selection rules* to decide which N assets to keep in the portfolio. We find large benefits from restricting the portfolio size N with our theory, as the two-fund and three-fund rules implemented with the 10 asset selection rules nearly consistently outperform the same portfolio rules applied to all M assets. The improvement is particularly large and statistically significant when we construct the portfolios with a shrinkage covariance matrix. Moreover, for some of the asset selection rules, the optimal two-fund and three-fund rules manage to consistently outperform EW and SGMV portfolios (with or without a risk-free asset), which is a notoriously difficult task when the dataset dimension M is comparable to the sample size, as well as benchmark MV portfolios that reduce the number of assets using weight thresholds or norm constraints.

The chapter is structured as follows. In Section 3.2, we study the optimal portfolio size for the MV portfolio with no parameter uncertainty. In Section 3.3, we show how to derive an optimal N for the SMV portfolio and the two-fund rule under parameter uncertainty. Sections 3.4 and 3.5 contain our simulation and empirical analysis, respectively. Section 3.6 concludes.

3.2 Optimal portfolio size without parameter uncertainty

In this section, we study the optimal portfolio size for an unconstrained MV investor who knows the parameters of asset returns without uncertainty. We suppose the investor starts from a universe of M assets and wishes to find an optimal number $N \leq M$ of assets.

Given a fixed number N of assets, let $\boldsymbol{r}$ be the $N \times 1$ vector of asset excess returns, which has a mean $\boldsymbol{\mu}_N$ and a positive-definite covariance matrix $\boldsymbol{\Sigma}_N$. We denote by μ_i, σ_i^2, and $s_i = \mu_i/\sigma_i$ the mean excess return, variance, and Sharpe ratio of asset i. An MV investor with risk-aversion coefficient $\gamma > 0$ selects the portfolio weights on the risky assets

as

$$\boldsymbol{w}^\star = \arg\max_{\boldsymbol{w}\in\mathbb{R}^N} U(\boldsymbol{w}) = \arg\max_{\boldsymbol{w}\in\mathbb{R}^N} \boldsymbol{w}'\boldsymbol{\mu}_N - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}_N\boldsymbol{w}, \tag{3.1}$$

where $U(\boldsymbol{w})$ is the MV utility of portfolio $\boldsymbol{w}$. Solving (3.1) yields the MV portfolio and the maximum achievable utility,

$$\boldsymbol{w}^\star = \frac{1}{\gamma}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N \quad \text{and} \quad U(\boldsymbol{w}^\star) = \frac{\theta_N^2}{2\gamma} \quad \text{with} \quad \theta_N^2 = \boldsymbol{\mu}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N, \tag{3.2}$$

where θ_N is the maximum Sharpe ratio. Clearly, $U(\boldsymbol{w}^\star)$ is non-decreasing with N because, at worst, one can set the weight of an additional asset to zero and keep the same utility. Thus, without parameter uncertainty, it is optimal to consider the whole investment universe and set $N = M$. To relate $U(\boldsymbol{w}^\star)$ and N more explicitly, we assume the following about $\boldsymbol{\Sigma}_M$. We denote $\boldsymbol{I}_M$ the identity matrix of dimension M, and $\mathbf{1}_M$ the $M \times 1$ vector of ones.

Assumption 1. *The covariance matrix $\boldsymbol{\Sigma}_M$ takes the form $\boldsymbol{\Sigma}_M(\rho) = \boldsymbol{D}_M\boldsymbol{P}_M(\rho)\boldsymbol{D}_M$ with $\boldsymbol{D}_M$ the diagonal matrix of standard deviations and $\boldsymbol{P}_M(\rho)$ the equicorrelation matrix,*

$$\boldsymbol{P}_M(\rho) = (1-\rho)\boldsymbol{I}_M + \rho\mathbf{1}_M\mathbf{1}_M', \tag{3.3}$$

where $-\frac{1}{M-1} < \rho < 1$ so that $\boldsymbol{\Sigma}_M$ is positive definite.

Under Assumption 1, the dependence structure is controlled by a single parameter, ρ. It follows that for any $N \leq M$, $\boldsymbol{\Sigma}_N$ has the same form as in Assumption 1. This assumption is commonly used in portfolio selection to trade off between specification and estimation error. Elton and Gruber (1973) find that assuming equicorrelation improves portfolio performance relative to a wide range of alternatives. Ledoit and Wolf (2003a) shrink the sample covariance matrix toward an equicorrelation model. Boyle et al. (2012) study an ambiguity-averse portfolio choice model and assume the parameters, including correlations, are homogeneous. Engle and Kelly (2012) propose a dynamic equicorrelation covariance matrix and find that it "[...] improves portfolio selection compared to an unrestricted dynamic correlation structure." Finally, Clements et al. (2015) "[...] find evidence in favour of assuming equicorrelation across various portfolio sizes, particularly during times of crisis." Chung et al. (2022) find that errors in correlations dominate those in means and variances for MV optimization.

Note that we use Assumption 1 *only* as a way to express the EU of our different portfolio rules as an explicit function of N, which, as we detail in Section 3.3, then allows us to find and estimate the optimal N. We do *not* need this assumption for the portfolio weights. In Section 3.A of the supplementary material, we find the optimal N under another commonly used covariance matrix approximation, the *single-factor model*, which we show is essentially equivalent to the optimal N under equicorrelation, highlighting the robustness of our method.

The next proposition provides the analytical expression of $U(\boldsymbol{w}^\star)$ under Assumption 1.[5]

Proposition 3.1. *Under Assumption 1, the utility of the MV portfolio is* $U(\boldsymbol{w}^\star) = \frac{\theta_N^2}{2\gamma}$ *with*

$$\theta_N^2 = \frac{N}{1-\rho}\delta_N(\bar{\boldsymbol{\theta}}_N), \quad \delta_N(\bar{\boldsymbol{\theta}}_N) = \bar{\theta}_{N,2} - \frac{N\rho}{1-\rho+N\rho}\bar{\theta}_{N,1}^2 \geq 0, \tag{3.4}$$

where $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ *is defined as*

$$\bar{\theta}_{N,k} = \frac{1}{N}\sum_{i=1}^{N} s_i^k. \tag{3.5}$$

Moreover, $U(\boldsymbol{w}^\star)$ *is non-decreasing in* N *and is strictly increasing from* N *to* $N+1$ *whenever* $s_{N+1} \neq \rho N\bar{\theta}_{N,1}/(1-\rho+N\rho)$.

Proposition 3.1 shows that the maximum utility depends on the assets' correlation structure via the correlation parameter ρ and on the assets' Sharpe ratios via $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$. Moreover, it is non-decreasing in N because including an additional asset in the portfolio cannot decrease the utility. In Figure 3.1, we depict $U(\boldsymbol{w}^\star)$ as a function of $N \in \{1, \ldots, M\}$ for $\rho \in \{0.2, 0.5, 0.8\}$ and assets' monthly Sharpe ratios θ_i calibrated to a dataset of $M = 96$ portfolios[6] sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.[7]

3.3 Optimal portfolio size under parameter uncertainty

We show in Section 3.2 that the utility of the MV portfolio without parameter uncertainty is non-decreasing with the portfolio size N. Intuitively, this is because excluding *a priori* some assets from the investment strategy amounts to adding a constraint to the optimization problem, which can only deteriorate the solution. In this section, we show that this no longer holds when the assets' parameters are unknown and the investor relies on sample estimates.

3.3.1 Sample estimates and distributional assumption

Given a sample of i.i.d. asset excess returns of size T, $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$, let

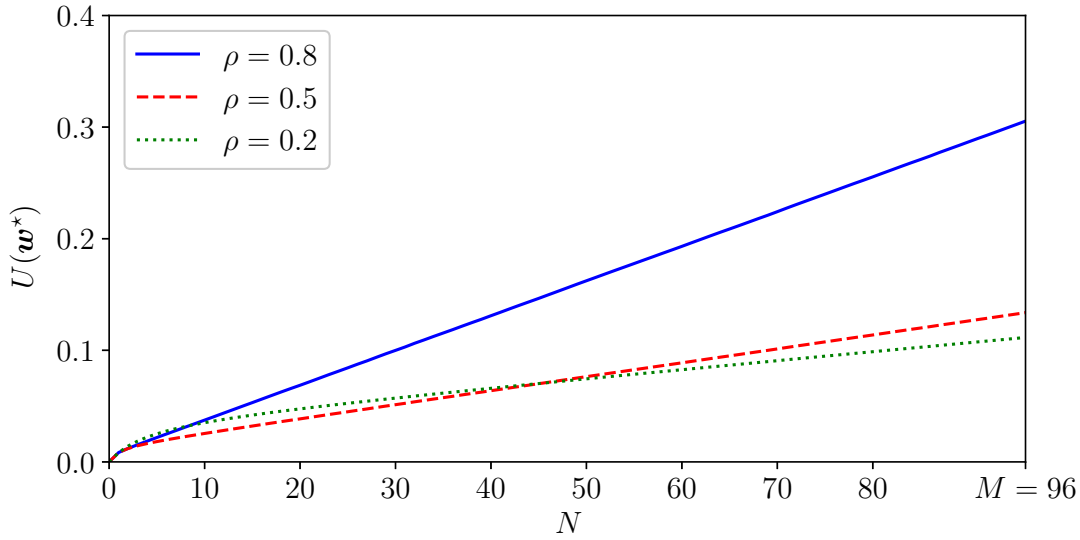
$$\hat{\boldsymbol{\mu}}_N = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_N = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)' \tag{3.6}$$

[5]The proofs of all results are available in Section 3.G of the supplementary material.

[6]We remove four portfolios sorted on size and book-to-market for which there is missing data.

[7]Figure 3.1 also shows that $U(\boldsymbol{w}^\star)$ is not a monotonic function of the correlation ρ. In Section 3.B of the supplementary material, we investigate the relationship between $U(\boldsymbol{w}^\star)$ and ρ and demonstrate that $U(\boldsymbol{w}^\star)$ is a convex function of ρ, extending the results of Gandy and Veraart (2013).

Figure 3.1: Utility of the mean-variance portfolio as a function of N and ρ



Notes. This figure depicts $U(\boldsymbol{w}^\star)$ in Equation (3.4) as a function of the portfolio size N under the assumption that asset returns are equicorrelated with a correlation equal to $\rho \in \{0.2, 0.5, 0.8\}$. We calibrate the assets' monthly Sharpe ratios to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023. Starting with $N = 1$ asset chosen randomly, we compute $U(\boldsymbol{w}^\star)$. Then, we add a randomly selected asset not previously selected, and compute $U(\boldsymbol{w}^\star)$ again. We continue this procedure until $N = M$. We repeat this procedure $10,000$ times and depict the average $U(\boldsymbol{w}^\star)$ over all draws. We consider a risk-aversion coefficient $\gamma = 1$.

be sample estimates of $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$, and we require $T > N$ so that $\hat{\boldsymbol{\Sigma}}_N$ is almost surely invertible. We study the optimal portfolio size N for *sample portfolios* implemented by relying on $\hat{\boldsymbol{\mu}}_N$ and $\hat{\boldsymbol{\Sigma}}_N$. Given a sample portfolio $\hat{\boldsymbol{w}}$, we define its optimal size $N^\star$ as the N maximizing the *expected out-of-sample utility (EU)* performance measure pioneered by Kan and Zhou (2007),

$$EU(\hat{\boldsymbol{w}}) = \mathbb{E}[U(\hat{\boldsymbol{w}})] = \mathbb{E}\left[\hat{\boldsymbol{w}}'\boldsymbol{\mu}_N - \frac{\gamma}{2}\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}_N\hat{\boldsymbol{w}}\right]. \tag{3.7}$$

Kan and Zhou (2007) and follow-up papers like Tu and Zhou (2011), DeMiguel et al. (2015), Kan et al. (2021), Lassance et al. (2024a,b), and Yuan and Zhou (2024) all take the expectation assuming that asset returns are normally distributed, which is analytically useful but unrealistic in practice. Therefore, we follow Kan and Lassance (2025) who study the EU of various sample portfolio rules when asset returns are *elliptically* distributed. Like them, we use the representation of the elliptical distribution in El Karoui (2010, 2013).

Assumption 2. *Asset returns are elliptically distributed, namely*

$$\boldsymbol{r} \overset{d}{=} \boldsymbol{\mu}_M + (\tau\boldsymbol{\Sigma}_M)^{1/2}\boldsymbol{z}_M, \tag{3.8}$$

where $\boldsymbol{z}_M \sim \mathcal{N}(\boldsymbol{0}_M, \boldsymbol{I}_M)$, τ *is a positive random variable satisfying* $\mathbb{E}[\tau] = 1$, *and* $\boldsymbol{z}_M \perp \tau$.[8]

[8]Because the multivariate elliptical distribution is closed under marginalization, the joint distribution of any subset of N assets belongs to the same family. The corresponding mean and covariance parameters $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ are obtained by selecting the appropriate entries of $\boldsymbol{\mu}_M$ and $\boldsymbol{\Sigma}_M$, respectively, while the distribution of τ remains unchanged.

The distribution of τ in Assumption 2 captures the effect of fat tails. For example, we recover the normal distribution when $\tau = 1$ collapses to a constant, and the t-distribution when $\tau \sim (\nu-2)/\chi^2_\nu$, where χ^2_ν denotes a chi-square distribution with $\nu > 2$ degrees of freedom. We focus on the elliptical distribution because it is consistent with MV portfolios (Chamberlain, 1983; Schuhmacher et al., 2021) and because Kan and Lassance (2025) show that it is crucial to account for fat tails in determining optimal portfolio combination rules.

3.3.2 Optimal portfolio size for sample portfolios

The first strategy for which we study the optimal portfolio size is the estimated counterpart of the MV portfolio in (3.2), i.e., the *sample mean-variance (SMV)* portfolio,

$$\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N. \tag{3.9}$$

However, the SMV portfolio is not robust to parameter uncertainty because it is only optimal *in sample*. Therefore, we also consider the *two-fund rule* of Kan and Zhou (2007) that scales down the SMV portfolio to maximize its EU. This two-fund rule is of the form

$$\hat{\boldsymbol{w}}(\alpha) = \alpha\hat{\boldsymbol{w}}^\star = \frac{\alpha}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N, \tag{3.10}$$

where $\alpha \in \mathbb{R}$ is the *combination coefficient*. Note that the SMV portfolio in (3.9) corresponds to $\alpha = 1$, i.e., $\hat{\boldsymbol{w}}(1) = \hat{\boldsymbol{w}}^\star$.

In the next proposition, we derive the EU of the SMV portfolio and the two-fund rule when asset returns are elliptically distributed, the optimal combination coefficient $\alpha^\star$, and which EU it delivers. These results then allow us to determine the optimal portfolio size N. In Section 3.C of the supplementary material, we follow the same approach for three-fund rules. Specifically, a three-fund rule based on the GMV portfolio as in, e.g., Kan and Zhou (2007), DeMiguel et al. (2015), and Yuan and Zhou (2024), which we call the *GMV-three-fund rule*, and a three-fund rule based on the EW portfolio as in, e.g., Tu and Zhou (2011), Kan and Wang (2023), and Lassance et al. (2024b), which we call the *EW-three-fund rule*.

Proposition 3.2. *Let $T > N+4$, $\boldsymbol{M} = \boldsymbol{I}_T - \mathbf{1}_T\mathbf{1}_T'/T$, $\boldsymbol{Z}_N$ be a $T \times N$ matrix of independent standard normal variables, $\boldsymbol{T}$ be a diagonal matrix of T independent copies of $\tau^{1/2}$, and $\boldsymbol{T} \perp \boldsymbol{Z}_N$. Then, under Assumption 2, the EU of the two-fund rule $\hat{\boldsymbol{w}}(\alpha)$ in* (3.10) *is*

$$EU(\hat{\boldsymbol{w}}(\alpha)) = \frac{1}{2\gamma}\frac{T}{T-N-2}\left[\left(2\alpha k_{N,1} - \frac{c_N\alpha^2 k_{N,2}T}{T-N-2}\right)\theta_N^2 - \frac{c_N\alpha^2 k_{N,3}N}{T-N-2}\right], \tag{3.11}$$

where θ_N^2 is the maximum squared Sharpe ratio given in (3.4)*, c_N is given by*

$$c_N = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}, \tag{3.12}$$

and $k_{N,1}$, $k_{N,2}$, and $k_{N,3}$ are functions of only N, T, and the distribution of τ:

$$k_{N,1} = \frac{T-N-2}{N}\mathbb{E}\left[\mathrm{tr}\big((\boldsymbol{Z}_N'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z}_N)^{-1}\big)\right], \tag{3.13}$$

$$k_{N,2} = \frac{(T-N-2)^2}{c_N N}\mathbb{E}\left[\mathrm{tr}\big((\boldsymbol{Z}_N'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z}_N)^{-2}\big)\right], \tag{3.14}$$

$$k_{N,3} = \frac{(T-N-2)^2}{c_N NT}\mathbb{E}\left[\mathbf{1}_T'\boldsymbol{T}\boldsymbol{Z}_N(\boldsymbol{Z}_N'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z}_N)^{-2}\boldsymbol{Z}_N'\boldsymbol{T}\mathbf{1}_T\right]. \tag{3.15}$$

The EU of the SMV portfolio is obtained as $EU(\hat{\boldsymbol{w}}^\star) = EU(\hat{\boldsymbol{w}}(1))$. Moreover, the optimal combination coefficient $\alpha^\star$ maximizing (3.11) *is*

$$\alpha^\star = \frac{T-N-2}{c_N T}\frac{k_{N,1}\theta_N^2}{k_{N,2}\theta_N^2 + k_{N,3}\frac{N}{T}}, \tag{3.16}$$

and the resulting EU of the optimal two-fund rule is

$$EU(\hat{\boldsymbol{w}}(\alpha^\star)) = \frac{1}{2\gamma c_N}\frac{k_{N,1}^2\theta_N^4}{k_{N,2}\theta_N^2 + k_{N,3}\frac{N}{T}}. \tag{3.17}$$

Three comments are in order. First, Proposition 3.2 is valid for any covariance matrix $\boldsymbol{\Sigma}_N$, not only those of the form $\boldsymbol{\Sigma}_N(\rho)$. Second, $k_{N,1}$, $k_{N,2}$, and $k_{N,3}$ do not have a closed-form expression, but can be evaluated via Monte Carlo simulations given the distribution of τ. We explain how we estimate them in Section 3.E of the supplementary material. Kan and Lassance (2025) show that they increase with estimation risk N/T and the tail heaviness. Third, because uncertainty in the parameters $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ impacts the EU, we no longer expect a larger N to always be favorable as it was the case in Section 3.2. Instead, when selecting the value of N maximizing (3.11), there is a tradeoff between improving the population performance (i.e., increasing θ_N^2) and limiting estimation risk (i.e., decreasing N/T).

Next, we proceed as in Section 3.2 and relate the EU with N more explicitly by assuming that the covariance matrix $\boldsymbol{\Sigma}_N$ complies with Assumption 1, i.e., $\boldsymbol{\Sigma}_N = \boldsymbol{\Sigma}_N(\rho)$.

Corollary 3.1. *Under Assumption 1, the EU of the SMV portfolio $\hat{\boldsymbol{w}}^\star$ and the optimal two-fund rule $\hat{\boldsymbol{w}}(\alpha^\star)$ are given by*

$$EU(\hat{\boldsymbol{w}}^\star) = \frac{1}{2\gamma}\frac{NT}{T-N-2}\left[\frac{\delta_N(\bar{\boldsymbol{\theta}}_N)}{1-\rho}\left(2k_{N,1} - \frac{c_N k_{N,2} T}{T-N-2}\right) - \frac{c_N k_{N,3}}{T-N-2}\right] \tag{3.18}$$

$$=: EU_{smv}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N), \text{ and} \tag{3.19}$$

$$EU(\hat{\boldsymbol{w}}(\alpha^\star)) = \frac{N}{2\gamma c_N(1-\rho)} \times \frac{k_{N,1}^2\delta_N(\bar{\boldsymbol{\theta}}_N)^2}{k_{N,2}\delta_N(\bar{\boldsymbol{\theta}}_N) + k_{N,3}\frac{1-\rho}{T}} =: EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N). \tag{3.20}$$

Corollary 3.1 shows that both EU_{smv} and EU_{2f} functions depend on several parameters.[9] First, the sample size T, directly but also via c_N and $(k_{N,1}, k_{N,2}, k_{N,3})$. Second, the

[9]The EU of the SMV portfolio and the two-fund rule is proportional to $1/\gamma$, and thus, the N optimizing the EU does not depend on γ. Therefore, we do not highlight the dependence on γ in the functions EU_{smv} and EU_{2f}. This is also true for the three-fund rules we consider in Section 3.C of the supplementary material.

portfolio size N, directly but also via c_N, the function δ_N, and $(k_{N,1}, k_{N,2}, k_{N,3})$. Third, *which* N assets are involved in the portfolio via $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$.

Now, to maximize $EU_{smv}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ or $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ with respect to N, one would also need to optimize the selection of assets at the same time because $\bar{\boldsymbol{\theta}}_N$ depends on N via the chosen subset of assets. This is a difficult optimization problem because the number of possible subsets of N assets with N going from one to M grows very quickly with M.[10] Moreover, this would introduce potentially severe estimation risk and time instability in the selection of assets, which is not desirable in practice. This is why, in order to disentangle the determination of the optimal N from the asset selection, we replace $\bar{\boldsymbol{\theta}}_N$ by its counterpart for the whole investment universe, i.e., $\bar{\boldsymbol{\theta}}_M$. This approximation may not be accurate when N is small relative to M.[11] In this case, however, the estimated $\bar{\boldsymbol{\theta}}_N$ will be particularly noisy, and thus estimating it from $\bar{\boldsymbol{\theta}}_M$ can help too. All in all, the way we determine the optimal portfolio size N for the SMV portfolio and the optimal two-fund rule becomes

$$N^\star_{smv} = \underset{N\in\{1,\ldots,\min(M,T-5)\}}{\operatorname{argmax}} EU_{smv}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M), \text{ and} \tag{3.21}$$

$$N^\star_{2f} = \underset{N\in\{1,\ldots,\min(M,T-5)\}}{\operatorname{argmax}} EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M), \tag{3.22}$$

where we search N up to $\min(M, T-5)$ because we assume $T > N+4$ in Proposition 3.2.[12]

We now illustrate the optimal portfolio sizes $N^\star_{smv}$ and $N^\star_{2f}$. Using the full sample of the 96S-BM dataset considered previously, we compute the sample estimators $\hat{\boldsymbol{\mu}}_{96}$ and $\hat{\boldsymbol{\Sigma}}_{96}$ using (3.6). Then, we set the population mean $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and the population covariance matrix $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\rho) = \hat{\boldsymbol{D}}_{96}\boldsymbol{P}_{96}(\rho)\hat{\boldsymbol{D}}'_{96}$ with $\hat{\boldsymbol{D}}_{96} = \operatorname{diag}(\hat{\boldsymbol{\Sigma}}_{96})^{1/2}$, which yields $\bar{\boldsymbol{\theta}}_{96} = (0.125, 0.0169)$, and we take $\rho \in \{0.2, 0.5, 0.8\}$. We consider two elliptical distributions. First, the normal distribution, i.e., $\tau = 1$ and $k_{N,1} = k_{N,2} = k_{N,3} = 1$. Second, the t-distribution, i.e., $\tau^2 \sim (\nu-2)/\chi^2_\nu$, with $\nu = 6$. Figure 3.2 depicts how $N^\star_{smv}$ and $N^\star_{2f}$ depend on the sample size T. We expect both to increase with T because, as T increases, the estimated portfolios become better estimates of the true optimal MV portfolio for which the optimal N is unbounded.

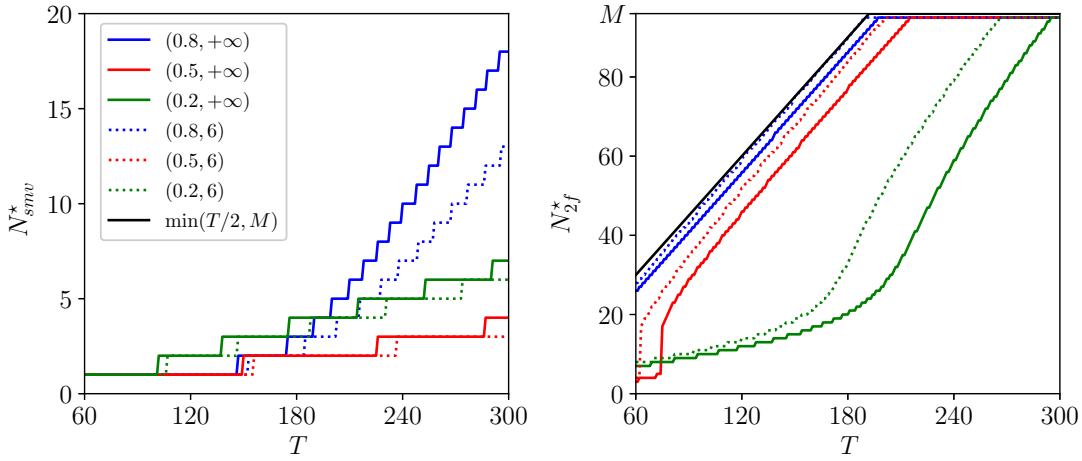
We make four observations from Figure 3.2. First, we find that $N^\star_{smv}$ is very small compared to T and M. For instance, with $T = 240$ and $\rho = 0.5$, we have $N^\star_{smv} = 3$ both for the normal and t-distributions. For a larger $\rho = 0.8$, we find a larger $N^\star_{smv}$ but it is still

[10] The number of possible subsets of N assets out of $M = 100$ ones is equal to $\sum_{N=1}^{M}\binom{M}{N} \approx 1.26 \times 10^{30}$.

[11] In Section 3.D of the supplementary material, we show that $\bar{\boldsymbol{\theta}}_N$ quickly converges to $\bar{\boldsymbol{\theta}}_M$ in practice.

[12] In Section 3.D of the supplementary material, we compare for the optimal two-fund rule the dependence on N of the EU obtained with $\bar{\boldsymbol{\theta}}_M$ and $\bar{\boldsymbol{\theta}}_N$, i.e., $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ and $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ in (3.20). The function $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ depends on N via the sequence of assets considered, and thus, we generate a large number of random asset sequences, evaluate $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ for each sequence, and compare it to $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$. We find that $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ provides a good approximation and that the dispersion of $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ across the asset sequences is reasonable.

Figure 3.2: Optimal portfolio size for the SMV portfolio and the optimal two-fund rule



Notes. This figure depicts the optimal portfolio size for the SMV portfolio, $N^\star_{smv}$ in (3.21) (left panel), and for the optimal two-fund rule, $N^\star_{2f}$ in (3.22) (right panel), as a function of the sample size T. Each line represents a different choice of the correlation, $\rho = \{0.2, 0.5, 0.8\}$, and the degrees of freedom of the t-distribution, $\nu = \{6, \infty\}$. We calibrate $\bar{\boldsymbol{\theta}}_M = (0.125, 0.0169)$ to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.

small. For example, when $\rho = 0.8$ and $T = 120$, 180 and 240, we have $N^\star_{smv} = 1$, 3, and 10 for the normal distribution, and $N^\star_{smv} = 1$, 2, and 7 for the t-distribution, respectively.

Second, both $N^\star_{smv}$ and $N^\star_{2f}$ tend to increase with ρ. This can be explained because, as we show in Section 3.B of the supplementary material, the maximum achievable utility without parameter uncertainty, $U(\boldsymbol{w}^\star)$ in Proposition 3.1, is a convex function of ρ, and thus, is increasing in ρ for ρ being large enough. Therefore, given that $U(\boldsymbol{w}^\star)$ is an increasing function of N, a larger ρ tends to increase the optimal N, everything else held equal.

Third, $N^\star_{2f}$ is higher than $N^\star_{smv}$, and is below $T/2$, or equal to M if $T/2$ is enough above M. Also, $N^\star_{2f}$ gets closer to $T/2$ as ρ increases. These results can be explained based on Equations (3.20) and (3.22). Specifically, we can show that in the normal case, i.e., $k_{N,1} = k_{N,2} = k_{N,3} = 1$, $N^\star_{2f}$ is below half of the sample size $T/2$ (or equal to M if M is sufficiently below $T/2$), and close to $T/2$ as ρ is either close to zero or one. To see this, consider the N that maximizes the first ratio of EU_{2f}, i.e., N/c_N. We have

$$\frac{N}{c_N} = \frac{N(T-N-1)(T-N-4)}{(T-2)(T-N-2)} \approx \frac{N(T-N-4)}{(T-2)}, \tag{3.23}$$

and the latter is maximized by

$$\underset{N\in\{1,\ldots,\min(M,T-5)\}}{\operatorname{argmax}} \frac{N(T-N-4)}{(T-2)} = \min([T/2-2], M). \tag{3.24}$$

Turning to the second ratio of EU_{2f} in (3.20), it is easy to show that if $k_{N,1} = k_{N,2} = k_{N,3} = 1$, it is a decreasing function of N and thus is maximized by $N = 1$, which moves $N^\star_{2f}$ below (3.24). However, the sensitivity of this second term to N depends on how sensitive $N\rho/(1-\rho+N\rho)$ is to N, and it is less so as ρ approaches zero or one. In equity

data where correlations are typically large, we thus expect $N^\star_{2f}$ to stay close to (3.24). Finally, in the elliptical case, $k_{N,1}$, $k_{N,2}$, and $k_{N,3}$ are different from one and depend on N, but Figure 3.2 and later simulations suggest that $N^\star_{2f}$ still remains close to that under normality.

Fourth, fat tails positively impact $N^\star_{2f}$, while we observe the opposite for $N^\star_{smv}$. This can be explained because the two-fund rule optimally scales down the SMV portfolio according to tail heaviness via $k_{N,1}$, $k_{N,2}$, and $k_{N,3}$. Therefore, as shown by Kan and Lassance (2025, Proposition 5), the two-fund rule often performs better when asset returns are elliptically instead of normally distributed, and thus we find a larger $N^\star_{2f}$ under elliptical returns.

3.4 Simulation analysis

We now run simulations to analyze two questions. First, in Section 3.4.1, what is the magnitude and variability of estimated optimal values of N for the SMV portfolio and the two-fund and three-fund rules. Second, in Section 3.4.2, whether we obtain a portfolio performance close to optimal even when we use *estimated* optimal values of N and that Assumption 1, i.e., asset returns are equicorrelated, is not fulfilled. To answer both questions, we draw monthly excess returns from a t-distribution, i.e., $\tau^2 \sim (\nu-2)/\chi^2_\nu$, with degrees of freedom $\nu \in \{6, \infty\}$. We calibrate the mean and covariance matrix of asset returns from the 96S-BM dataset as previously.[13] Specifically, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and consider $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, where $\bar{\rho} = 0.74$ is the average of the correlations in $\hat{\boldsymbol{\Sigma}}_{96}$.[14] The second choice of $\boldsymbol{\Sigma}_M$ allows us to compare the estimated optimal values of N with the true optimal one that, for each portfolio strategy, is known theoretically only under Assumption 1, i.e., when $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$.

3.4.1 Estimated optimal portfolio size

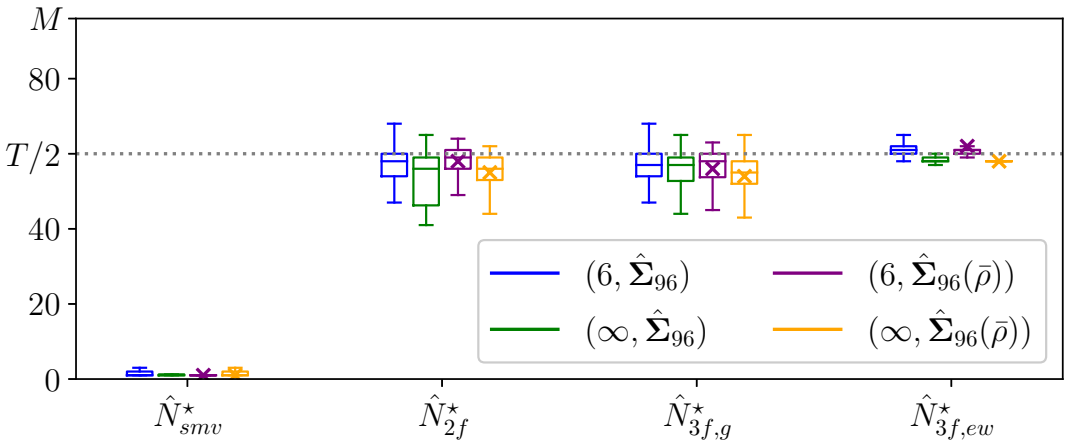
We now analyze the estimated optimal portfolio size for the SMV portfolio ($\hat{N}^\star_{smv}$), the two-fund rule ($\hat{N}^\star_{2f}$), the GMV-three-fund rule ($\hat{N}^\star_{3f,g}$), and the EW-three-fund rule ($\hat{N}^\star_{3f,ew}$). These are the estimated counterparts of $N^\star_{smv}$ in (3.21), $N^\star_{2f}$ in (3.22), $N^\star_{3f,g}$ in (3.51), and $N^\star_{3f,ew}$ in (3.62) following the methodology detailed in Section 3.E of the supplementary material.

We simulate $K = 10,000$ times $T = 120$ returns and estimate, for each strategy, the optimal portfolio sizes in each of the K simulations. Figure 3.3 depicts boxplots of the estimated optimal N for the different choices of $(\nu, \boldsymbol{\Sigma}_M)$ across all simulations. We make a number of observations. First, in the case $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, our $\hat{N}^\star$ are close to the true $N^\star$ on average for each strategy, highlighting the quality of the estimation procedure

[13] Running our simulation analysis on the five additional datasets we consider in the empirical analysis of Section 3.5 yields similar insights, and thus, we focus on the 96S-BM dataset for conciseness.

[14] In Section 3.F.1 of the supplementary material, instead of fixing $\rho = \bar{\rho}$, we study how the estimated optimal portfolio size varies with ρ.

Figure 3.3: Estimated optimal portfolio size in simulated data



Notes. This figure depicts boxplots of the estimated optimal portfolio size N for the SMV portfolio ($\hat{N}^{\star}_{smv}$), two-fund rule ($\hat{N}^{\star}_{2f}$), GMV-three-fund rule ($\hat{N}^{\star}_{3f,g}$), and EW-three-fund rule ($\hat{N}^{\star}_{3f,ew}$). These are the estimated counterparts of $N^{\star}_{smv}$ in (3.21), $N^{\star}_{2f}$ in (3.22), $N^{\star}_{3f,g}$ in (3.51), and $N^{\star}_{3f,ew}$ in (3.62) following the estimation methodology in Section 3.E of the supplementary material. The boxplots are obtained by simulating 10,000 times $T = 120$ t-distributed returns. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\boldsymbol{\Sigma}}_{96}$, and consider $\nu \in \{6, \infty\}$ degrees of freedom. Each boxplot corresponds to a different choice of $(\nu, \boldsymbol{\Sigma}_M)$. We depict with a '$\times$' marker the oracle value of the optimal N known under Assumption 1, i.e., when $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$.

proposed in Section 3.E of the supplementary material. Second, the $\hat{N}^{\star}$ are similar under $\hat{\boldsymbol{\Sigma}}_{96}$ and $\hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, meaning that drawing returns from a distribution that does not fulfill Assumption 1 does not significantly affect the estimation of N. Third, the $\hat{N}^{\star}$ for the SMV portfolio is very small, in line with Figure 3.2. Fourth, the $\hat{N}^{\star}$ are similar for the two-fund and three-fund rules, meaning that they are robust to considering different portfolio combination rules, and are close to but typically below $T/2$, which is because $\bar{\rho}$ is rather large as discussed in Section 3.3.2. Finally, the lower the ν, and thus the fatter the tails, the higher the optimal N in the two-fund and three-fund rules, which is consistent with Figure 3.2.

3.4.2 Optimality of restricted portfolios

Section 3.4.1 shows the estimated optimal portfolio size $\hat{N}^{\star}$ for various MV combination rules. We now analyze the performance of the corresponding restricted portfolios using $N = \hat{N}^{\star}$. We observe that even in the unfavorable setting where (i) Assumption 1 is violated, (ii) the optimal N is estimated from the data and (iii) the N assets selected in the restricted portfolios are chosen randomly,[15] the restricted portfolio delivers an EU close to the maximum achievable one, and outperforms on average the naive equally-weighted portfolio which is not subject to estimation risk. For conciseness, we only consider the

[15]In the empirical analysis of Section 3.5, we use different selection rules to improve upon a random selection. In the current simulation exercise we are interested in general conclusions, independent of one specific asset selection rule, averaged over a large number of random selections.

two-fund rule in this section; the results are similar for the three-fund rules in Section 3.C of the supplementary material.

We conduct this analysis as follows. For each choice of $(\nu, \boldsymbol{\Sigma}_M)$, where $\nu \in \{6, \infty\}$ and $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, we draw $K = 10{,}000$ times $L = 300$ t-distributed returns. We then use the L returns in a rolling window exercise. Specifically, for each simulation k, given the sample size $T = 120$ and a portfolio size N, we randomly select at each rolling window a subset of N assets, compute the estimated two-fund rule $\hat{\boldsymbol{w}}(\hat{\alpha}^\star)$[16] over these N assets, and evaluate the out-of-sample portfolio return $r_{t,k}$ over the next month. We also implement the naive EW portfolio $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$ on the same N assets for comparison purposes. We then roll the window by one month and proceed similarly until we have used the whole sample of L returns. This procedure gives us, for any N, a time series of out-of-sample portfolio returns $r_{t,k}$, where $t = 1, \ldots, L-T$ and $k = 1, \ldots, K$, from which we compute the EU,

$$EU = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2\right), \quad \text{where} \tag{3.25}$$

$$\hat{\mu}_k = \frac{1}{L-T}\sum_{t=1}^{L-T} r_{t,k}, \quad \hat{\sigma}_k^2 = \frac{1}{L-T}\sum_{t=1}^{L-T}(r_{t,k} - \hat{\mu}_k)^2. \tag{3.26}$$

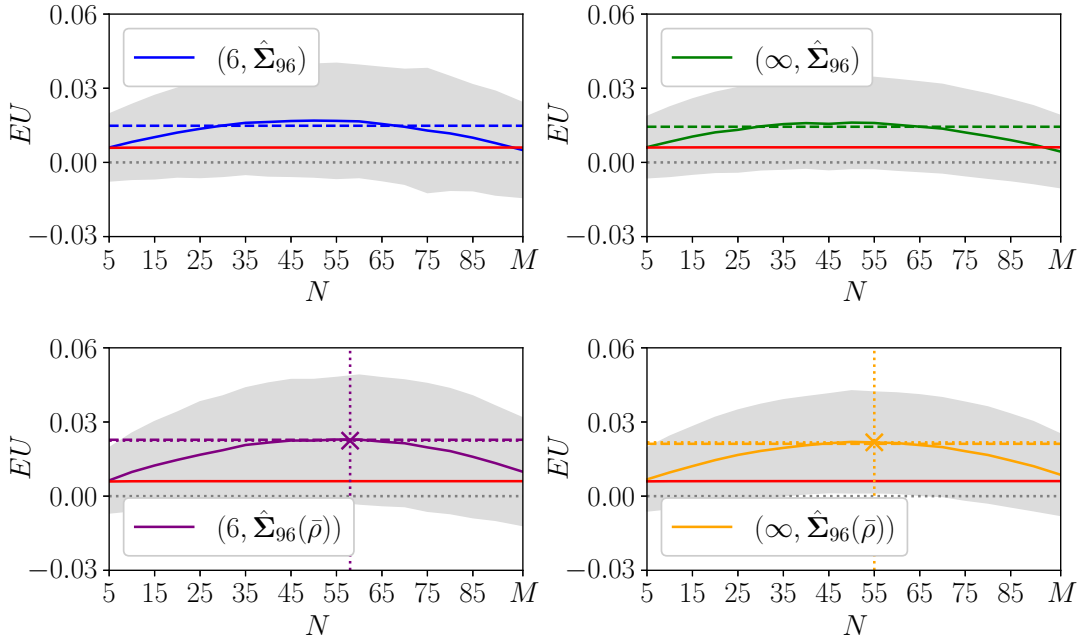
We implement this exercise with various portfolio sizes and depict the results in Figure 3.4. First, for both $\hat{\boldsymbol{w}}(\hat{\alpha}^\star)$ and $\boldsymbol{w}_{ew}$, we consider fixed values of N ranging from $N = 5$ to $N = M = 96$ with a step size of five (six for the last step), and depict the EU as a function of N with a solid line for $\hat{\boldsymbol{w}}(\hat{\alpha}^\star)$ and with a dash-dotted red line for $\boldsymbol{w}_{ew}$. Second, for $\hat{\boldsymbol{w}}(\hat{\alpha}^\star)$ only, we consider the estimated optimal portfolio size $N = \hat{N}^\star_{2f,t,k}$ which is not fixed but instead varies depending on the rolling window and the simulation, and depict its EU with a horizontal dashed line. Third, for $\hat{\boldsymbol{w}}(\hat{\alpha}^\star)$ only and for the two bottom panels, i.e., when $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, we also consider the oracle $N = N^\star_{2f}$ obtained under the true parameters $(\hat{\boldsymbol{\mu}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho}), \nu)$. In this case, we depict the value of $N^\star_{2f}$ with a vertical dotted line and its EU with a horizontal dotted line, and mark the intersection with a '$\times$' marker.[17]

Several observations can be made from Figure 3.4. First, our estimated optimal portfolio size for the two-fund rule, $\hat{N}^\star_{2f}$, delivers an EU that is close to the maximum, *both* for the case in which returns are equicorrelated and in which they are not. Second, we can see that for both $\nu = 6$ and $\nu = \infty$, the EU is maximized when N is slightly below $T/2$, in line with the theoretical results derived in Section 3.3.2 and the fact that the average correlation is large for the 96S-BM dataset ($\bar{\rho} = 0.74$). Third, $\hat{N}^\star_{2f}$ delivers an EU that is substantially larger than that when N is either too small or too close to M which highlights the value

[16] $\hat{\alpha}^\star$ is the estimate of the optimal combination coefficient in the two-fund rule, $\alpha^\star$ in (3.16). We detail our estimation methodology in Section 3.E of the supplementary material.

[17] In Section 3.F.2 of the supplementary material, we implement a similar exercise in which the correlation matrix has a block structure. Even in this setup that strongly violates Assumption 1, there are significant gains from reducing the portfolio size from M to $\hat{N}^\star_{2f}$.

Figure 3.4: Expected out-of-sample utility of the two-fund rule in simulated data



Notes. This figure depicts the expected out-of-sample utility (EU) of the two-fund rule as a function of the portfolio size N. We simulate 10,000 samples of t-distributed returns with $\nu \in \{6, \infty\}$ degrees of freedom. For each N, we conduct a rolling window exercise as described in Section 3.4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\boldsymbol{\Sigma}}_{96}$. Each panel corresponds to a different choice of $(\nu, \boldsymbol{\Sigma}_M)$. In each panel, the solid line depicts the EU in (3.25), the shaded gray area depicts the one standard deviation interval around the EU across all simulations, and the dashed horizontal line depicts the EU obtained when using the estimated optimal N for the two-fund rule, i.e., the estimate of $N^{\star}_{2f}$ in (3.22) according to the estimation method in Section 3.E of the supplementary material. The dash-dotted red line depicts the EU delivered by the equally weighted portfolio. In the bottom panels, the vertical and horizontal dotted lines depict the oracle N and its EU, respectively, with the '×' marker showing the intersection. The dotted horizontal gray line depicts the zero EU level. We set a risk-aversion coefficient of $\gamma = 1$.

of choosing the right portfolio size and the performance risk of blindly investing in all available assets. Lastly, $\hat{N}^{\star}_{2f}$ substantially outperforms the naive EW portfolio on average, which is a challenging benchmark strategy when M is comparable to T.

3.5 Empirical analysis

We now test the performance of our optimally restricted MV portfolios in empirical data. In Section 3.5.1 and 3.5.2, we present the datasets and portfolio strategies, respectively. In Section 3.5.3, we propose different rules to select in which assets to invest after obtaining the optimal portfolio size. In Section 3.5.4, we detail the out-of-sample methodology and performance measures. Finally, we report and discuss the results in Section 3.5.5.

3.5.1 Datasets

The cross-section of equities is composed of several thousands of stocks, which is a challenging high-dimensional environment. To explain the cross-section more parsimoniously, the

standard practice in the asset pricing literature is to value-weight stocks into characteristic portfolios according to firm characteristics that have been shown to explain average returns. Another standard practice is to group stocks into a smaller number of industry portfolios.

Therefore, our empirical analysis builds on six datasets of monthly excess returns of characteristic and industry portfolios, which we list in Table 3.1. Given that our objective is to optimally reduce the portfolio size, the full investment universe must not be too small in the first place, and thus we consider datasets with M being around 100 assets.

The first dataset already considered previously, 96S-BM, is composed of decile portfolios sorted on size and book-to-market. The second dataset, 108CHA, is composed of the long and short legs of the 54 characteristics considered in Lassance and Martín-Utrera (2024), collected from the authors. The third dataset, 100S-OP, is composed of decile portfolios sorted on size and operating profitability. The fourth dataset, 94IN-NV, is composed of 48 industry portfolios and of the long and short legs obtained from the 23 characteristics in Novy-Marx and Velikov (2015), which are available on Robert Novy-Marx's website. The fifth dataset, 107IN-CHA, is composed of 47 industry portfolios,[18] of quintile portfolios sorted on size and book-to-market and on operating profitability and investment, and of decile portfolios sorted on momentum. The sixth dataset, 98IN-CHA-NV, is composed of 47 industry portfolios, of quintile portfolios sorted on size and book-to-market ratio, of decile portfolios sorted on momentum, and of the long and short legs obtained from the eight low-turnover characteristics in Novy-Marx and Velikov (2015). The time period for each dataset is reported in Table 3.1.

In Section 3.F.3 of the supplementary material, we look at how Assumption 1 (equicorrelated returns) and Assumption 2 (elliptical returns) transpire in the six datasets. First, we find that the equicorrelation assumption is reasonable for the 96S-BM, 100S-OP, and 108CHA datasets, with a root mean squared error (RMSE) between the observed and equicorrelation matrices of around 5-10%, but less so for the three other datasets, with a RMSE around 10-20%. Thus, we can assess the performance of our optimal portfolio size across different correlation structures. Second, we find that the parameters $(k_{M,1}, k_{M,2}, k_{M,3})$ defined in (3.13)–(3.15), which control the impact of elliptical fat tails of asset returns on the EU, substantially depart from one, i.e., from normality. Thus, it is crucial to account for fat tails when implementing combination rules given their impact on optimal combination coefficients.

3.5.2 Portfolio strategies

We evaluate the out-of-sample performance of the 10 portfolio strategies in Table 3.2. The first three are the optimal combination rules in Section 3.3.2 and Section 3.C of the supplementary material: the two-fund rule (2F), the GMV-three-fund rule (3FGMV),

[18]We consider 47 industry portfolios instead of 48 because of missing data in the healthcare portfolio.

Table 3.1: List of datasets considered in the empirical analysis

#	Name	Description	M	Time period
1	96S-BM	96 portfolios sorted on size and book-to-market	96	07/1963–08/2023
2	108CHA	108 characteristic portfolios built on long and short legs of 54 characteristics in Lassance and Martín-Utrera (2024)	108	09/1966–12/2020
3	100S-OP	100 portfolios sorted on size and operating profitability	100	07/1963–08/2023
4	94IN-NV	48 industry portfolios and 46 characteristic portfolios built on long and short legs of 23 characteristics in Novy-Marx and Velikov (2015)	94	07/1973–12/2013
5	107IN-CHA	47 industry portfolios, 25 portfolios sorted on size and book-to-market, 25 portfolios sorted on operating profitability and investment, 10 portfolios sorted on momentum	107	07/1963–12/2022
6	98IN-CHA-NV	47 industry portfolios, 25 portfolios sorted on operating profitability and investment, 10 portfolios sorted on momentum, 16 characteristic portfolios built on long and short legs of eight low-turnover characteristics in Novy-Marx and Velikov (2015)	98	07/1963–12/2013

Notes. This table lists the datasets of monthly excess returns considered in the empirical analysis of Section 3.5, which we detail in Section 3.5.1. The columns report, in order, the dataset number, the dataset name, the dataset description, the size of the dataset M, and the time period. The industry portfolios and the portfolios sorted on size, book-to-market, operating profitability, investment, and momentum are downloaded from Kenneth French's website. The characteristic portfolios in Novy-Marx and Velikov (2015) are downloaded from Robert Novy-Marx's website. In the construction of these datasets, we have removed assets with missing data over the time period considered. The 108CHA dataset is the same as that used by Lassance and Martín-Utrera (2024). In the construction of the 108CHA dataset, Lassance and Martín-Utrera (2024) drop characteristics with more than five percent of missing observations for more than five percent of firms with CRSP returns available for the entire sample.

and the EW-three-fund-rule (3FEW). We apply these strategies to all M assets or to a subset of N assets, with N chosen optimally as $\hat{N}^{\star}_{2f}$ in (3.22) for 2F, $\hat{N}^{\star}_{3f,g}$ in (3.51) for 3FGMV, and $\hat{N}^{\star}_{3f,ew}$ in (3.62) for 3FEW, which we estimate following Section 3.E of the supplementary material.

The next two portfolio strategies are individual portfolios: the EW portfolio in (3.52) and the SGMV portfolio in (3.41). We then consider two portfolio strategies that combine EW and SGMV with the risk-free asset to maximize the EU, EWRF and GMVRF, which might be preferred to the 2F, 3FGMV, and 3FEW combination rules because they disregard the SMV portfolio that faces high estimation risk. We implement those as

$$\hat{w}_{ewrf} = \frac{\hat{\lambda}_{ew,N}}{\gamma}\frac{\mathbf{1}_N}{N} \qquad (3.27) \qquad \text{and} \qquad \hat{w}_{gmvrf} = \frac{\hat{\mu}_{g,N}}{c_N\gamma}\hat{\mathbf{\Sigma}}^{-1}\mathbf{1}, \qquad (3.28)$$

where $\hat{\lambda}_{ew,N}$ and $\hat{\mu}_{g,N}$ are defined in (3.70)–(3.71).

The last portfolio strategy we consider is the SMV portfolio, which we implement in three different ways. In the first case, we compute the SMV portfolio on all M assets using Equation (3.9). In the second and third case, we reduce the portfolio size using two methods in which either the estimation of weights and the choice of N are achieved in

Table 3.2: List of portfolio strategies considered in the empirical analysis

Name	Description
2F	Two-fund rule that combines the sample mean-variance portfolio and the risk-free asset. See Equation (3.10) and Proposition 3.2.
3FGMV	Three-fund rule that combines the sample mean-variance portfolio, the sample global minimum-variance portfolio, and the risk-free asset. See Equation (3.43) and Proposition 3.5.
3FEW	Three-fund rule that combines the sample mean-variance portfolio, the equally weighted portfolio, and the risk-free asset. See Equation (3.54) and Proposition 3.7.
EW	Equally weighted portfolio. See Equation (3.52).
EWRF	Combination of EW and risk-free asset. See Equation (3.27).
SGMV	Sample global minimum-variance portfolio. See Equation (3.41).
GMVRF	Combination of SGMV and the risk-free asset. See Equation (3.28).
SMV	Sample mean-variance portfolio. See Equation (3.9).
SMV-L_1	SMV portfolio subject to a L_1-norm constraint with the threshold selected via cross-validation. See Equation (3.29).
SMV-TW	SMV portfolio with a threshold on weights selected via cross-validation.

Notes. This table lists the portfolio strategies we consider in the empirical analysis, which we detail in Section 3.5.2. We apply the 2F, 3FGMV, 3FEW, and EW portfolio strategies on a subset of all M assets, and the asset selection rules we implement for that purpose are described in Section 3.5.3.

one step, or N is chosen after having first computed the SMV portfolio weights on all M assets.[19]

The first SMV portfolio with a reduced size is SMV-L_1, which solves the mean-variance portfolio problem subject to an L_1-norm constraint,

$$\hat{\boldsymbol{w}}_{L_1} = \underset{\boldsymbol{w}\in\mathbb{R}^M}{\operatorname{argmax}} \ \boldsymbol{w}'\hat{\boldsymbol{\mu}}_M - \frac{\gamma}{2}\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}_M\boldsymbol{w} \quad \text{subject to} \quad ||\boldsymbol{w}||_1 = \sum_{i=1}^{M}|w_i| \leq \delta. \tag{3.29}$$

The L_1-norm performs asset selection and limits the amount of short-selling, which helps improve out-of-sample performance (DeMiguel et al., 2009a). We calibrate the threshold δ using cross-validation with out-of-sample utility as a decision criterion. Specifically, we search the optimal δ in an interval of 1,000 equally spaced values ranging from zero to $\delta_{\max}$, defined as the norm of the unconstrained solution to (3.29), i.e., $\delta_{\max} = ||\frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_M^{-1}\hat{\boldsymbol{\mu}}_M||_1$.

The second SMV portfolio with a reduced size is SMV-TW, which operates a trimming on the weights. The idea is that after having computed the SMV portfolio on all M assets, investors may want to select assets whose absolute portfolio weights are above a given

[19] In Section 3.F.5 of the supplementary material, we also report the performance of the restricted SMV portfolio whose portfolio size is determined using our theory in Section 3.F.5, i.e., using $\hat{N}^{\star}_{smv}$, and the asset selection rules in Section 3.5.3. We find that for all selection rules, the restricted SMV portfolio outperforms the unrestricted SMV on all M assets. However, we don't report these results in the main text because, even with a reduced size, the SMV portfolio largely underperforms the 2F, 3FGMV, and 3FEW portfolios.

threshold $\bar{w}$, in which case the value of N is determined implicitly by $\bar{w}$ as

$$N_{\bar{w}} = \sum_{i=1}^{M} \mathbb{1}_{\{|\hat{w}_i^\star| \geq \bar{w}\}} \quad \text{where} \quad \hat{\boldsymbol{w}}^\star = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\boldsymbol{\mu}}_M. \tag{3.30}$$

Once the $N_{\bar{w}}$ assets are selected, we recompute the SMV portfolio on these $N_{\bar{w}}$ assets, which delivers better performance than keeping the original weights that are estimated under a high level of estimation risk. We select the threshold $\bar{w}$ via cross-validation with out-of-sample utility as a decision criterion. Specifically, we search the optimal $\bar{w}$ in an interval of 10 equally spaced values ranging from zero (all assets selected) to $\bar{w}_{\max}$ (no asset selected), where $\bar{w}_{\max} = \max|\frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\boldsymbol{\mu}}_M|$ is the largest absolute weight of the unrestricted SMV portfolio.[20]

Finally, we consider three different estimates of the covariance matrix. First, the sample estimator in (3.6) for consistency with our theory. Second, the linear shrinkage estimator of Ledoit and Wolf (2004). Third, the nonlinear shrinkage estimator of Ledoit and Wolf (2020), for which we report the results in Section 3.F.4 of the supplementary material for conciseness.[21]

3.5.3 Asset selection rules

Given a portfolio size N for the 2F, 3FGMV, and 3FEW portfolios, we must determine a *selection rule* to decide which N assets to select among all M assets. Our purpose is not to find which selection rule is most effective. Instead, our main objective is to illustrate the added value of optimally reducing the portfolio size, and that even simple but sensible selection rules can deliver substantial gains over choosing all M assets.

For comparison purposes, we will also apply some of the proposed selection rules to the EW portfolio to evaluate their intrinsic value without them being affected by estimation errors in optimized portfolio weights as it is the case for the 2F, 3FGMV, and 3FEW portfolios. In that case, we will consider a subset of assets of size $N = T/2$, where we use a sample size $T = 120$ months as explained in Section 3.5.4, because the optimal N is generally close to $T/2$ for typical levels of equity correlations as explained in our theory and our simulations.

The asset selection rules we consider are listed in Table 3.3. We propose 11 rules that we select according to two criteria. First, they are *simple* to understand and implement. Second, they are *sensible*, i.e., they all have an economic intuition. The first selection rule consists in choosing all assets and setting $N = M$. The 10 remaining selection rules

[20] Given the exposure of the SMV to estimation risk, using more than 10 values yields similar values since the cross-validation typically yields $\bar{w}$ close to $\bar{w}_{\max}$ such that we keep a very small number of assets.

[21] Analytical expressions for the EU of portfolio strategies implemented with shrinkage estimators of $\boldsymbol{\Sigma}_N$ are not available. Therefore, we cannot obtain the optimal combination coefficients and optimal portfolio size N in those cases. As an alternative, we follow Kan et al. (2021) and implement the portfolios by relying on the obtainable combination coefficients and optimal N computed under the sample estimator of the covariance matrix $\boldsymbol{\Sigma}_N$, but use shrinkage estimators of $\boldsymbol{\Sigma}_N$ to compute the portfolio weights.

Table 3.3: List of asset selection rules considered in the empirical analysis

Name	Description
All	All assets are included in the portfolio: $N = M$.
MaxSR	Select N assets with the maximum Sharpe ratios.
MinSR	Select N assets with the minimum Sharpe ratios.
BWSR	Select $\lceil N/2 \rceil$ assets with the best Sharpe ratios and $\lfloor N/2 \rfloor$ assets with the worst Sharpe ratios.
MaxVar	Select N assets with the maximum variances.
MinVar	Select N assets with the minimum variances.
BWVar	Select $\lceil N/2 \rceil$ assets with the best variances and $\lfloor N/2 \rfloor$ assets with the worst variances.
MaxPC	Select the N assets with the maximum correlations with the first principal component.
MinPC	Select the N assets with the minimum correlations with the first principal component.
Best θ_N^2	Select the N assets which deliver the best portfolio Sharpe ratio as in Ao et al. (2019).
MaxW	Select the N assets with the maximum absolute weights.

Notes. This table lists the asset selection rules we use in the empirical analysis of Section 3.5, which we detail in Section 3.5.3. That is, in an investment universe of M assets, which N assets we select given an optimal $N \leq M$. The notation $\lceil N/2 \rceil$ ($\lfloor N/2 \rfloor$) means $N/2$ rounded above (below) to the nearest integer.

select a subset of $N \leq M$ assets. Among those, the first six are based solely on marginal information about the assets, whereas the last four rules also consider information about their dependence. Moreover, the first nine rules select the assets before computing the portfolio weights (ex-ante approach), whereas the last rule selects the subset of assets after computing the weights on all assets (ex-post approach). We now describe these 10 asset selection rules.[22]

Maximum Sharpe ratio (MaxSR). We select the N assets with the maximum in-sample Sharpe ratios over the estimation window, because the EU of the SMV portfolio in (3.18) and of the two-fund rule in (3.16) both depend on the assets' Sharpe ratios. This selection rule is expected to work particularly well when there is momentum in the assets.[23] However, as a downside, it ranks the assets according to average returns that are notoriously noisy (Merton, 1980). We can anticipate two other issues with this rule. First, the selected assets have the best Sharpe ratios, but estimation risk may lead us to assign a wrong sign to portfolio weights and mistakenly short some good assets. Second, we do not include the assets with the worst Sharpe ratios although shorting them might deliver performance gains.

Minimum Sharpe ratio (MinSR). We select the N assets with the minimum in-sample Sharpe ratios. This rule will work well under mean-reversion in the assets, or from using short positions. We expect MinSR to face the same issues as MaxSR, i.e., not having

[22] Our three-stage approach, as described in Section 3.1, is flexible in that it allows any type of selection rule, and we propose a specific set as an illustration. One may also consider more sophisticated rules based on asset clustering, time-series models, or machine learning.

[23] We use datasets of characteristic and industry portfolios, and portfolios exhibit more momentum (i.e., positive return autocorrelation) than individual stocks (Campbell et al., 1997, p.74).

access to the assets with the best Sharpe ratios to form long positions, and ending up with mistaken long positions on assets that have bad Sharpe ratios as a result of estimation noise.

Best-worst Sharpe ratio (BWSR). We blend the MaxSR and MinSR rules by selecting the $\lceil N/2 \rceil$ assets with the best in-sample Sharpe ratios and the $\lfloor N/2 \rfloor$ assets with the worst in-sample Sharpe ratios, which can be combined with long and short positions. This is in line with the standard way of constructing individual characteristics as long-short portfolios of stocks in the top and bottom quantiles of a given firm characteristic. However, BWSR may lead to more extreme portfolio weights that do not generalize well out of sample.

The MaxSR, MinSR, and BWSR selection rules rank assets based on their in-sample Sharpe ratios.[24] Given that average returns are notoriously noisy, these rankings are bound to be unstable.[25] As a result, the set of selected assets is likely to change substantially from one period to another, increasing portfolio turnover and transaction costs. To address this point, the next three asset selection rules in Table 3.3 are the following.

Maximum, minimum, and best-worst variance (MaxVar, MinVar, BWVar). These selection rules rank assets based on the in-sample variance instead of the Sharpe ratio. The resulting rankings of assets are expected to be more stable over time than those under the Sharpe ratio because the variance is quite persistent over time. The MaxVar rule selects assets with the maximum variances, which are bound to also have larger expected returns if a positive risk-return tradeoff stands in the data. However, the low-risk anomaly implies that assets and portfolios with lower variances often have larger expected returns (Ang et al., 2006; Moreira and Muir, 2017), an anomaly that the MinVar rule can exploit.

The selection rules considered up to now exploit marginal information about the assets. We now propose four other selection rules that account for the dependence between assets.

Maximum and minimum correlation with the first principal component (MaxPC, MinPC). We select those assets whose returns have the maximum or the minimum correlations with the first principal component (PC). On the one hand, those assets with minimum correlations with the first PC are those least explained by the remaining assets, and thus, that offer higher potential for diversification gains. On the other hand, those assets with maximum correlations may lead to less extreme weights than those optimized on more dissimilar assets.

Best portfolio Sharpe ratio (Best θ_N^2). This selection rule follows Ao et al. (2019) and consists in choosing the assets so that they deliver a high maximum squared Sharpe ratio, θ_N^2 in (3.2). We implement this rule in three steps. First, we form 1,000 random selections

[24] In unreported results, we implement selection rules that rank assets according to their in-sample utilities. We find that these rules deliver an out-of-sample performance close to the Sharpe ratio based rules.

[25] In unreported results, we implement selection rules that rank assets according to average returns. We find that they typically underperform selection rules based on the Sharpe ratio and variance.

of N assets, where N is our optimal portfolio size. Second, for each selection, we estimate θ_N^2 as detailed in Section 3.E of the supplementary material. Third, we pick the selection which corresponds to the 95% quantile of all estimated values of θ_N^2, to avoid outliers.

The selection rules above are *ex-ante* rules because the selection of the N assets is done prior to and independently of the portfolio strategy. However, the in-sample portfolio weights on all M assets can be a valuable information for deciding which N assets to select. Therefore, the last selection rule we consider in Table 3.3 is an *ex-post* selection rule in which we select the assets based on the portfolio strategy that will then be applied to the selected assets.

Maximum absolute weights (MaxW). This selection rule works in three steps. First, compute $\hat{\boldsymbol{w}} \in \mathbb{R}^M$, the portfolio weights on all M assets using a given strategy (2F, 3FGMV, or 3FEW). Second, select the N assets with the maximum absolute weights $|\hat{w}_i|$. Third, recompute the portfolio on the N selected assets using the same strategy as before to obtain $\hat{\boldsymbol{w}} \in \mathbb{R}^N$.[26] MaxW selects the assets whose weights are large, and thus, are identified as relevant by the portfolio. We can anticipate the issue for MaxW that the selection of assets inherits large estimation errors because it is based on portfolio weights optimized under a high level of estimation risk, i.e., a high value of M/T.

In Section 3.F.8, we study how different is the selection of assets with these different rules. Specifically, we introduce a measure that determines how much more or less overlap there is between two asset selection rules relative to a pair of independent random selection rules. We find that for essentially all pairs of selection rules we consider, there is either around the same or less overlap than the random selection case. That is, our selection rules do select dissimilar assets, which allows us to test our portfolio strategies on different investment universes.

3.5.4 Out-of-sample methodology

We evaluate the out-of-sample portfolio performance with a standard monthly rebalancing. Specifically, at the end of month t, we estimate portfolio k over the T previous months, and we compute its out-of-sample return in month $t+1$. We consider a fixed sample size of $T = 120$ months.[27] We repeat this process iteratively, resulting in $T_{tot} - T$ out-of-sample gross returns $r_{gross,k,t}$, where T_{tot} is the total number of months in the dataset. We then compute the net out-of-sample returns, $r_{net,k,t} = r_{gross,k,t}$ if $t = T+1$ and

$$r_{net,k,t} = (1 + r_{gross,k,t})\,(1 - p \times \text{turnover}_{k,t-1}) - 1 \quad \text{if} \quad t = T+2, \ldots, T_{tot}, \tag{3.31}$$

[26] We find that dropping this third step and keeping the initial portfolio weights computed on the M assets, with or without rescaling, delivers a poor performance because these weights are computed under a high level of estimation risk, i.e., a high value of M/T.

[27] Given that our theoretical results require $T > M + 4$, and that the largest $M = 108$, we need $T > 112$. If we choose a sample size that is too large, such as $T = 240$, then the optimal N for the 2F, 3FGMV, and 3FEW portfolios, which is close to $T/2$ as shown in Section 3.4.1, will often be larger than the total size of the investment universe, M. In that case, we cannot assess the effect of reducing the portfolio size.

where p is the proportional transaction cost parameter and

$$\text{turnover}_{k,t} = \sum_{i=1}^{N} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, T_{tot}, \tag{3.32}$$

with $w_{i,k,t}$ the weight of asset i in month t and $w_{i,k,(t-1)^+}$ the prior-month weight before rebalancing in month t. We set $p = 10$ basis points in line with average bid-ask spreads reported by Engle et al. (2012) and Frazzini et al. (2018). Finally, we compare the portfolio strategies in terms of *annualized out-of-sample utility net of transaction costs*,

$$U_k = 12 \times \left(\hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2\right), \tag{3.33}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the sample mean and variance of $r_{net,k,t}$. We set a risk-aversion coefficient of $\gamma = 1$; the intuition behind our results remains very similar for other values of γ.[28] In Section 3.F.6 of the supplementary material, we also consider the out-of-sample Sharpe ratio.

For the 2F, 3FGMV, and 3FEW portfolios, the selected assets change as we re-estimate the optimal N and apply an asset selection rule. Because our methodology disentangles the determination of the optimal N from the choice of the N assets, we can also disentangle the rebalancing of portfolio weights from changes in the N assets. Therefore, to make our method less costly and more practically implementable, we rebalance the portfolio weights every month but change the optimal N and select assets once a year. In Section 3.F.7 of the supplementary material, we also consider changing the selected assets every one, six and 24 months, and analyze the impact on performance. We find that setting a lower frequency than monthly substantially reduces turnover and improves the net-of-cost performance. Moreover, frequencies of 6, 12, or 24 months yield similar results overall.

Finally, we compute two-sided p-values for the statistical test of the difference between the net utility of the 2F, 3FGMV, and 3FEW portfolios computed on all assets ('All' selection rule) versus under each of the 10 other selection rules in Table 3.3. The p-values for SMV-TW and SMV-L_1 are computed by taking SMV as a benchmark. To compute the p-values, we generate 10,000 bootstrap samples using the stationary block bootstrap approach of Politis and Romano (1994) with an average block size of five, and then use the methodology of Ledoit and Wolf (2008, Remark 3.2) to produce the resulting p-values. We use the symbols ○, ◑, and ● to indicate that the p-value is less than 10%, 5%, and 1%, respectively.

3.5.5 Discussion of results

In Table 3.4, we report the annualized net out-of-sample utility, in percentage points, of the 10 portfolio strategies listed in Table 3.2. For the 2F, 3FGMV, 3FEW, and EW portfolio

[28]The out-of-sample utility net of costs of the 2F, 3FGMV, 3FEW, GMVRF, EWRF, and SMV portfolios is proportional to $1/\gamma$, and thus, their ranking is not affected by the chosen γ. However, the ranking of the SGMV and EW portfolios, which are fully invested in risky assets, is affected by the chosen γ.

strategies, we report the performance across the 11 asset selection rules in Table 3.3.[29]

The results in Table 3.4 show that portfolio strategies estimated with the shrinkage covariance matrix generally outperform those estimated with the sample one. Therefore, albeit the theory for finding optimal portfolio strategies considers the sample covariance matrix for tractability, it is beneficial for investors to use the shrinkage covariance matrix. Overall, the ranking of portfolio strategies and selection rules is similar for both covariance matrix estimators, hence the discussion of results below essentially applies to either case.

First, we address our main objective which is to assess the value of optimally reducing the portfolio size. For all three portfolio strategies, i.e., 2F, 3FGMV, and 3FEW, we find that optimally reducing the portfolio size delivers substantial performance gains. Specifically, the asset selection rule 'All', which sets $N = M$, is nearly systematically outperformed by the remaining selection rules that select a subset of assets whose size is determined using our theory.[30] The differences are often statistically significant, particularly when using a shrinkage covariance matrix. Moreover, the differences are systematically positive for the MinSR, BWSR, MaxVar, BWVar, MinPC, and MaxPC asset selection rules, and are often substantial. For instance, under the shrinkage covariance matrix, and on average across datasets, the 2F strategy applied to all M assets delivers an annualized net utility of 11.30 percentage points, whereas it delivers 58.17 percentage points with the BWSR selection rule. We depict this main finding in Figure 3.5 which shows, across the six datasets and for the shrinkage covariance matrix, the difference between the annualized net out-of-sample utility of 2F, 3FGMV, and 3FEW implemented on a subset of $\hat{N}^\star$ assets versus all M assets. Having addressed our main objective, we now discuss other findings from Table 3.4.

Second, we find that applying the MinSR, MaxVar, and MaxPC asset selection rules to the 2F, 3FGMV, and 3FEW strategies delivers consistently positive net out-of-sample utilities that are almost systematically larger than those of the seven other rules. Focusing on the shrinkage covariance matrix, the BWSR, BWVar, MinPC, and Best-θ_N^2 selection rules also consistently deliver positive utilities that generally outperform the benchmarks. These results are quite remarkable because the EW portfolio, in particular, is difficult to outperform using mean-variance portfolios when estimation risk, i.e., M/T, is large, which is the case in all our datasets where M/T ranges from 0.78 to 0.9. These results are the consequence of our optimal portfolio size because the 2F, 3FGMV and 3FEW portfolios with the 'All' selection rule ($N = M$) underperform the EW portfolio for four out of six datasets.

Third, we turn to the asset selection rules based on the Sharpe ratio, i.e., MaxSR, MinSR,

[29]The MaxW asset selection rule is not applicable to the EW portfolio since all weights are equal.

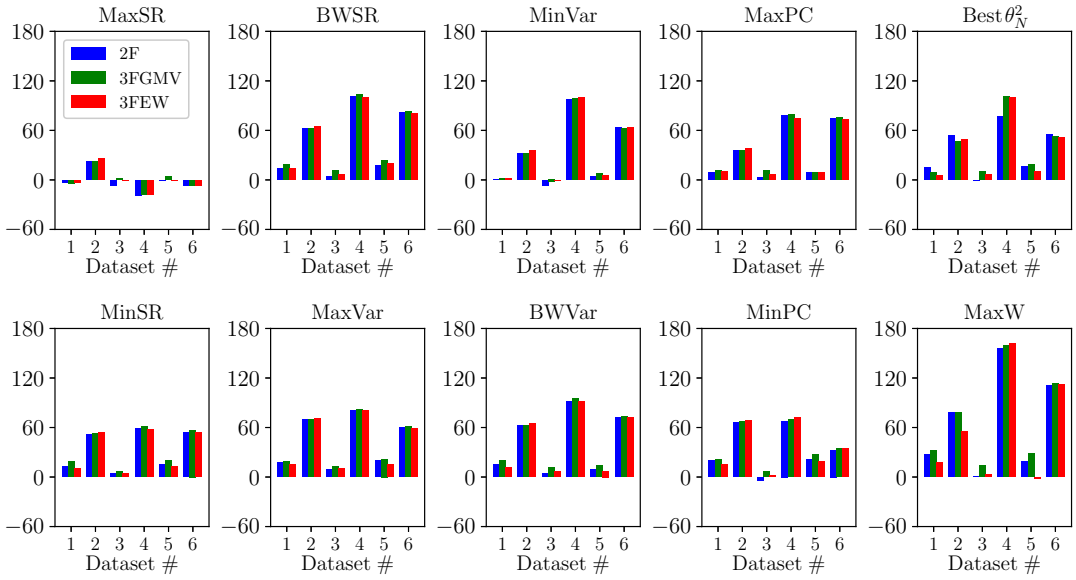
[30]Out of the 180 configurations (10 selection rules × 6 datasets × 3 portfolio strategies), the restricted portfolios outperform the 'All' selection rule in 153, 161, and 165 cases under the sample, linear shrinkage, and nonlinear shrinkage covariance matrix, respectively. If we restrict to cases in which the p-value for the performance gain is below 10%, these numbers reduce to 93, 147, and 152, respectively.

Table 3.4: Annualized net out-of-sample utility in empirical data (in percentage points)

		Sample covariance matrix						Linear shrinkage covariance matrix					
Port-	Asset	Dataset						Dataset					
folio	select.	96	108	100	94	107	98IN-	96	108	100	94	107	98IN-
strat.	rule	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV
2F	All	−0.31	−59.6	−15.3	99.4	−18.8	56.8	5.11	1.20	1.09	41.6	1.09	17.7
2F	MaxSR	1.38	11.6●	−7.48	23.6○	−10.5	−27.1◑	1.93○	23.1●	−6.10	23.1●	0.69	10.8○
2F	MinSR	25.4●	73.2●	7.10●	146	24.9●	139◑	18.0●	53.2●	5.21○	101●	16.0●	72.1●
2F	BWSR	7.39	84.4●	−4.42	133	11.7○	242●	19.2◑	63.8●	4.74	143●	18.8◑	99.5●
2F	MaxVar	25.0◑	124●	17.9●	114	21.7●	83.8	23.2●	70.5●	9.99◑	122●	21.1●	77.7●
2F	MinVar	2.47	12.5●	−13.1	219◑	−1.88	225●	6.08	33.1●	−5.25○	140●	5.40○	81.4●
2F	BWVar	13.0	89.3●	−6.76	182○	0.14	180●	19.8●	63.4●	5.39	133●	10.6◑	89.9●
2F	MaxPC	8.80	18.3●	3.67◑	169○	9.62●	237●	14.7●	36.9●	4.67	119●	9.81●	91.7●
2F	MinPC	27.4●	112●	−6.23	115	17.9◑	94.1	24.9●	67.7●	−2.55	109●	22.8●	50.5●
2F	Best θ_N^2	−1.22	14.7●	−8.77	182○	4.73	137◑	24.4●	59.2●	−4.03	149●	23.0●	95.7●
2F	MaxW	−127●	−305●	−179●	−581●	−129●	−366●	32.3●	79.0●	2.19	197●	19.8○	129●
3FGMV	All	0.47	−60.5	−12.5	99.1	−19.7	55.4	5.77	1.21	1.79	41.9	1.31	17.9
3FGMV	MaxSR	−4.06	10.8●	−1.93○	15.6◑	−6.02	−27.4◑	1.63	23.3●	3.04	24.6●	5.32	10.8○
3FGMV	MinSR	36.0●	71.6●	12.9●	150	31.5●	141◑	24.3●	53.4●	8.94◑	103●	21.5●	74.8●
3FGMV	BWSR	10.9	82.4●	1.02	143	19.9◑	253●	24.7●	64.0●	13.0◑	146●	25.2●	101●
3FGMV	MaxVar	25.1◑	126●	24.9●	117	21.7●	85.0	24.6●	70.9●	14.3●	123●	23.1●	79.6●
3FGMV	MinVar	6.13	13.3●	−14.9	222◑	0.35○	219●	7.94	33.2●	−0.72	141●	8.69◑	80.7●
3FGMV	BWVar	27.4◑	89.0●	2.74	183○	11.2◑	190●	25.8●	63.5●	12.9◑	137●	15.4●	90.8●
3FGMV	MaxPC	12.2	16.7●	10.8●	176○	7.50◑	225●	17.7●	36.9●	13.1●	122●	11.0●	93.1●
3FGMV	MinPC	33.3●	113●	6.47◑	115	25.9●	91.3	27.0●	68.2●	8.74○	112●	28.7●	52.0●
3FGMV	Best θ_N^2	12.3	7.39●	−0.08	122	1.33○	165●	15.1○	48.2●	11.9●	144●	19.6●	71.0●
3FGMV	MaxW	−136●	−300●	−147●	−650●	−110●	−325●	38.2●	79.4●	15.7	201●	30.4●	131●
3FEW	All	1.83	−62.8	−10.8	102	−16.8	60.8	7.21	−2.56	3.01	40.3	2.38	18.5
3FEW	MaxSR	3.68	12.6●	−0.61	23.6◑	−12.1	−20.4◑	4.05	23.4●	2.23	22.4◑	1.79	12.0
3FEW	MinSR	24.8●	70.8●	11.3●	145	24.8●	146◑	17.6◑	51.0●	7.15	98.2●	15.3◑	72.0●
3FEW	BWSR	8.43	84.4●	3.41	135	17.9◑	237●	20.6●	62.6●	9.48○	140●	22.0●	98.8●
3FEW	MaxVar	26.0◑	122●	25.3●	124	16.0●	98.9	22.2●	68.7●	12.8●	121●	18.0●	77.0●
3FEW	MinVar	4.93	12.3●	−5.87	220◑	0.40	230●	8.94	33.5●	2.79	140●	8.08○	82.2●
3FEW	BWVar	14.2	88.6●	−3.30	195○	−0.15	187●	19.0●	62.4●	9.50○	132●	9.65○	90.4●
3FEW	MaxPC	9.42	18.1●	11.9●	162	5.87◑	247●	17.0●	35.7●	9.98◑	115●	11.5●	92.0●
3FEW	MinPC	26.0◑	114●	2.26	120	15.9◑	103	22.1●	66.3●	5.27	113●	21.4●	52.8●
3FEW	Best θ_N^2	12.5	4.58●	−4.45	136	−8.83	166●	12.5	46.2●	10.1◑	140●	13.2●	69.5●
3FEW	MaxW	−124●	−305●	−157●	−650●	−139●	−361●	24.8●	52.9●	5.80	202●	0.52	131●
EW	All	8.11	2.80	8.00	3.98	7.13	5.61	8.11	2.80	8.00	3.98	7.13	5.61
EW	MaxSR	8.77○	4.72●	8.49○	5.97●	7.40	6.25◑	8.77○	4.72●	8.49○	5.97●	7.40	6.25◑
EW	MinSR	7.39◑	0.83●	7.55○	2.75●	6.77	5.18	7.39◑	0.83●	7.55○	2.75●	6.77	5.18
EW	BWSR	7.94	2.42●	7.72	2.86●	7.35	5.28	7.94	2.42●	7.72	2.86●	7.35	5.28
EW	MaxVar	7.95	1.38●	7.97	3.52	6.68	5.31	7.95	1.38●	7.97	3.52	6.68	5.31
EW	MinVar	8.60	4.23●	8.18	4.36	7.44	5.86	8.60	4.23●	8.18	4.36	7.44	5.86
EW	BWVar	7.60●	2.34●	7.52◑	3.99	6.88	5.58	7.60●	2.34●	7.52◑	3.99	6.88	5.58
EW	MaxPC	8.32	2.97	8.83◑	2.19●	7.37	5.08◑	8.32	2.97	8.83◑	2.19●	7.37	5.08◑
EW	MinPC	7.87	2.48●	7.20●	5.00◑	6.97	6.22◑	7.87	2.48●	7.20●	5.00◑	6.97	6.22◑
EW	Best θ_N^2	7.88○	2.72	7.92	3.82	7.02	6.03◑	8.27○	2.75	7.82	3.95	6.90	5.92◑
EWRF		1.42	−4.35	1.26	−3.06	0.80	−0.29	1.42	−4.35	1.26	−3.06	0.80	−0.29
SGMV		1.76	−49.7	4.25	−0.34	−2.28	−4.67	10.3	2.33	7.75	4.02	8.48	6.65
GMVRF		1.81	−26.3	−0.98	−11.4	−12.3	−5.34	4.01	−0.00	1.47	0.14	0.66	0.49
SMV		−4E+05	−3E+11	−1E+06	−6E+07	−8E+08	−8E+07	−1443	−107	−1399	−2360	−1457	−1143
SMV-L_1		15.8●	−108●	11.7●	−579●	−2.91●	0.88●	23.9●	67.8	5.13●	−195●	10.4●	16.4●
SMV-TW		−35.3●	−302●	−16.8●	−105●	−25.9●	−143●	−29.9●	−52.1	−17.8●	−376●	−33.6●	−349●

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 10 portfolio strategies in Table 3.2, the 11 asset selection rules in Table 3.3, across the six datasets in Table 3.1. The table is constructed following the out-of-sample methodology described in Section 3.5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under each selection rule with the 'All' selection rule as benchmark, using the block bootstrap methodology described in Section 3.5.4. We also report p-values for SMV-TW and SMV-L_1, using the SMV as benchmark. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

Figure 3.5: Difference in net out-of-sample utility relative to investing in all assets



Notes. This figure depicts, for three different portfolio strategies, the difference between the annualized net out-of-sample utility in percentage points obtained when implementing the portfolios on a subset of N assets, where the optimal N is estimated using our theory, using 10 asset selection rules relative to the case in which the portfolios are implemented on all M assets. The three portfolio strategies are the two-fund rule (2F, blue), the GMV-three-fund rule (3FGMV, green), and the EW-three-fund rule (3FEW, red). Each panel considers one of the 10 asset selection rules described in Table 3.3. The figure is constructed following the methodology described in Section 3.5.4. We consider six datasets described in Table 3.1. We construct the portfolios with the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

and BWSR. When these are applied to the EW portfolio, MaxSR consistently outperforms and delivers better performance than the EW portfolio of all M assets. In contrast, MinSR and BWSR underperform the EW portfolio of all assets. Specifically, on average across datasets, the EW portfolio of all assets delivers an annualized net out-of-sample utility of 5.94 percentage points, whereas the EW portfolio with the MaxSR, MinSR, and BWSR selection rules delivers average utilities of 6.93, 5.08, and 5.60, respectively. These results indicate return momentum as forming an EW portfolio of assets with maximum in-sample Sharpe ratios helps improve out-of-sample performance relative to selecting all assets.

However, when we turn to MV optimized portfolios, namely, 2F, 3FGMV and 3FEW, we reach an opposite conclusion. Specifically, MinSR consistently outperforms MaxSR as well as the portfolio implemented on all M assets, and MaxSR is one the worst-performing selection rules. These results mean that when we use optimized portfolio weights, the performance gains from asset selection rules do not originate from selecting assets with the best marginal performance, but from selecting assets that can be best exploited by the portfolio, e.g., with short positions for the MinSR rule. Finally, the BWSR selection rule, which capitalizes on long and short positions by selecting both good and bad assets, achieves its objective as it consistently outperforms MaxSR and MinSR when using a

shrinkage covariance matrix.[31]

Fourth, we turn to the asset selection rules based on the variance, i.e., MaxVar, MinVar, and BWVar. When these are applied to the EW portfolio, MinVar consistently outperforms and also delivers better performance than the EW portfolio of all M assets. Specifically, on average across datasets, the EW portfolio of all assets delivers an annualized net out-of-sample utility of 5.94 percentage points, whereas the EW portfolio with the MaxVar, MinVar, and BWVar selection rules delivers average utilities of 5.47, 6.44, and 5.65, respectively. This finding indicates a low-risk anomaly in our datasets. However, as noted above, this does not necessarily mean that MinVar is also the best selection rule for the 2F, 3FGMV and 3FEW strategies. And indeed, Figure 3.5 shows that for these portfolios, MinVar delivers a worse performance overall than MaxVar and BWVar. This can be explained because the assets with the maximum variances, which are selected by MaxVar and BWVar, often have low or negative average returns as the low-risk anomaly predicts, and the 2F, 3FGMV and 3FEW portfolios are able to profit from this using small or short positions.[32]

Fifth, the MaxPC and MinPC asset selection rules, which account for the dependence between the assets by selecting those that are most or least correlated with the first PC, offer a consistently good performance. In particular, the resulting 2F, 3FGMV, and 3FEW portfolios have positive net out-of-sample utilities in all cases. However, there is no clear trend regarding which of the two selection rules outperforms the other.

Sixth, the Best θ_N^2 selection rule, which attempts to maximize the portfolio Sharpe ratio, outperforms investing in all assets for the 2F, 3FGMV and 3FEW portfolios in all cases but one. Interestingly, although they are quite similar in spirit, the Best θ_N^2 selection rule systematically and significantly outperforms the MaxSR rule under the shrinkage covariance matrix, indicating that considering information about the dependence of assets is crucial in selecting assets that deliver a large portfolio Sharpe ratio.

Seventh, the MaxW selection rule outperforms on average all other rules under the shrinkage covariance matrix. For instance, considering the 3FGMV portfolio, MaxW delivers an annualized net utility of 82.69 percentage points on average across datasets, whereas the second-best selection rule, BWSR, delivers an average utility of 62.32 percentage points. However, MaxW underperforms all other selection rules under the sample covariance matrix. This can be explained because under the sample estimator, the 2F, 3FGMV and 3FEW portfolios are more subject to error maximization (Michaud, 1989) than under

[31] However, BWSR is often outperformed by MinSR under the sample covariance matrix because, in that case, the long-short positions are too extreme, and thus, do not work well out of sample and increase turnover.

[32] It seems that the low-risk anomaly is strong enough in the 94IN-NV and 98IN-CHA-NV datasets for the MinVar selection rule to outperform MaxVar and BWVar when applied to the 2F, 3FGMV and 3FEW portfolios. This can be explained because Moreira and Muir (2017) show that the low-risk anomaly does not apply to the size factor, and this factor is used to construct all datasets except 94IN-NV and 98IN-CHA-NV.

the shrinkage estimator. Therefore, selecting the assets with large absolute weights, i.e., identified as most relevant, is more desirable under the shrinkage estimator than under the sample estimator.[33]

Finally, we investigate our three ways of implementing the SMV portfolio: SMV, SMV-L_1, and SMV-TW. First, the plain SMV delivers by far the worst performance because we consider settings with high M/T. Second, SMV-L_1, which selects assets by maximizing the MV utility subject to an L_1-norm constraint, greatly improves out-of-sample performance relative to SMV and SMV-TW. However, in most cases, SMV-L_1 underperforms the optimally restricted 2F, 3FGMV, and 3FEW portfolios. Third, SMV-TW, which keeps only the assets whose SMV portfolio weights are above a threshold selected via cross-validation, outperforms the plain SMV but still delivers negative net utilities. In Section 3.F.5 of the supplementary material, we compare the portfolio size obtained with SMV-L_1 and SMV-TW to that obtained with our theory, $\hat{N}^{\star}_{smv}$. We find that the three approaches set a small N on average, and that our approach displays much less variability across windows.

3.6 Conclusion

We offer a novel perspective on the optimal *portfolio size* and challenge the conventional wisdom that investors should invest in as many assets as possible to diversify away idiosyncratic risk. Specifically, because of *parameter uncertainty*, we show that it is optimal to invest in a limited number N of assets. For several mean-variance portfolio combination rules, we derive the optimal value $N^{\star}$ of N maximizing the expected *out-of-sample* performance. This optimal N strikes a tradeoff between accessing additional investment opportunities and limiting the number of parameters and optimal weights to estimate. For typical levels of equity return correlations, we find that the optimal N is slightly below *half the sample size*, which can be quite small. Our approach is flexible in the sense that it disentangles the computation of the optimal N from the choice of which N assets to include in the portfolio.

We empirically test our theory in various environments and analyze the results. To this end, we propose an estimator $\hat{N}^{\star}$ of $N^{\star}$ and suggest a set of simple, sensible and intrinsically different selection rules to determine which $\hat{N}^{\star}$ assets to select from the initial investment universe. We show that our optimally restricted portfolios outperform their counterparts that invest in all available assets. This result prevails in nearly all cases, across different portfolio strategies, datasets, and asset selection rules, and holds with

[33]To confirm this, we implement in unreported results an opposite selection rule denoted MinW which selects the assets with the lowest absolute weights instead of the highest absolute weights. The idea behind MinW is to exclude assets with extreme portfolio weights because these are most subject to error maximization. As expected, we find that MinW systematically outperforms MaxW by a substantial margin when using the sample covariance matrix, but the opposite happens under the shrinkage covariance matrix.

transaction costs. For some asset selection rules, our optimally restricted portfolios also systematically outperform the robust EW and minimum-variance portfolios, which are hard to beat in high dimension.

Overall, our methodology renders portfolio theory valuable in the challenging but practically relevant case in which the sample size is close to the size of the investment universe.

Supplementary Material

This supplementary material contains seven sections. In Section 3.A, we study the optimal portfolio size under a single-factor model assumption for the covariance matrix. In Section 3.B, we analyze the relation between the correlation of assets and the maximum utility of the MV portfolio. In Section 3.C, we present the theory we use to determine the optimal portfolio size for the three-fund rules. In Section 3.D, we study the impact of the sequence of assets on the EU and the optimal portfolio size. In Section 3.E, we explain how we estimate the different parameters on which the combination coefficients and the optimal values of N depend. In Section 3.F, we report and discuss additional empirical results. In Section 3.G, we provide the proofs of all theoretical results.

3.A Single-factor model assumption

In the main body of the chapter, we find the optimal portfolio size by assuming that asset returns are equicorrelated, i.e., Assumption 1. In this section, we show that we obtain a very similar portfolio size under another assumption for the covariance matrix, which is that asset returns follow a single-factor model. This is a commonly used assumption in asset pricing and portfolio selection to obtain parsimonious representations of expected returns and covariance matrices; see, e.g., MacKinlay and Pástor (2000), Tu and Zhou (2011), and Kan et al. (2021).

Assumption 3. *The asset returns follow a single-factor model,*

$$\boldsymbol{r} = \boldsymbol{\alpha}_N + \boldsymbol{\beta}_N f + \boldsymbol{\varepsilon}, \tag{3.34}$$

where $f \in \mathbb{R}$ is the factor with zero mean and variance σ_f^2, $\boldsymbol{\beta}_N \in \mathbb{R}^N$ is the vector of factor loadings, $\mathbb{E}[\boldsymbol{\varepsilon}] = \boldsymbol{0}_N$, $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma_\varepsilon^2 \boldsymbol{I}_N$, and f is uncorrelated with $\boldsymbol{\varepsilon}$.

Under Assumption 3, the covariance matrix of the asset returns is of the form

$$\boldsymbol{\Sigma}_N = \sigma_f^2 \boldsymbol{\beta}_N \boldsymbol{\beta}_N' + \sigma_\varepsilon^2 \boldsymbol{I}_N. \tag{3.35}$$

This parametrization of the covariance matrix entails $N + 2$ parameters: the N factor sensitivities $\boldsymbol{\beta}_N$, σ_f^2, and σ_ε^2. This is similar to the parametrization of the covariance matrix under Assumption 1, which entails $N + 1$ parameters: N asset variances and one correlation coefficient. However, the equicorrelation and single-factor assumptions restrict the covariance matrix in different ways. Assumption 1 fixes the correlation coefficients to a single ρ, while allowing individual variances to vary freely. In contrast, Assumption 3 entangles variances and correlations, setting the variance of asset i to $\sigma_f^2 \beta_i^2 + \sigma_\varepsilon^2$, and the correlation between assets i and j to $\beta_i \beta_j \sigma_f^2$.

Given these differences, if both assumptions result in similar optimal portfolio sizes, then this would support the robustness of our approach. Therefore, we proceed by comparing the optimal portfolio sizes under the two assumptions, both the oracle and estimated ones. We focus on the optimal portfolio size for the two-fund rule for conciseness; similar results can be obtained for the SMV portfolio and the three-fund rules considered in Section 3.C

First, in order to find the optimal N for the two-fund rule, in the next proposition we derive the maximum squared Sharpe ratio, $\theta_N^2 = \boldsymbol{\mu}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N$, and the EU of the optimal two-fund rule, $EU(\hat{\boldsymbol{w}}(\alpha^\star))$, under the single-factor covariance matrix in (3.35)

Proposition 3.3. *Under Assumption 3, the maximum squared Sharpe ratio is*

$$\theta_N^2 = \frac{N}{\sigma_\varepsilon^2} \Delta_N(\bar{\boldsymbol{\ell}}_N) \quad \textit{with} \quad \Delta_N(\bar{\boldsymbol{\ell}}_N) = \bar{\ell}_N^{\mu} - \frac{N\sigma_f^2}{\sigma_\varepsilon^2 + N\sigma_f^2 \bar{\ell}_N^{\beta}} (\bar{\ell}_N^{\mu,\beta})^2, \tag{3.36}$$

where $\bar{\boldsymbol{\ell}}_N = (\bar{\ell}_N^{\mu}, \bar{\ell}_N^{\beta}, \bar{\ell}_N^{\mu,\beta})$, $\bar{\ell}_N^{\mu} = \frac{1}{N}\sum_{i=1}^N \mu_i^2$, $\bar{\ell}_N^{\beta} = \frac{1}{N}\sum_{i=1}^N \beta_i^2$, *and* $\bar{\ell}_N^{\mu,\beta} = \frac{1}{N}\sum_{i=1}^N \mu_i \beta_i$. *Moreover, the EU of the optimal two-fund rule* $\hat{\boldsymbol{w}}(\alpha^\star)$ *is given by*

$$EU(\hat{\boldsymbol{w}}(\alpha^\star)) = \frac{N}{2\gamma c_N \sigma_\varepsilon^2} \times \frac{k_{N,1}^2 \Delta_N(\bar{\boldsymbol{\ell}}_N)^2}{k_{N,2}\Delta_N(\bar{\boldsymbol{\ell}}_N) + k_{N,3}\sigma_\varepsilon^2/T} =: EU_{2f}^{sf}(N, T, \sigma_f^2, \sigma_\varepsilon^2, \tau, \bar{\boldsymbol{\ell}}_N). \tag{3.37}$$

As in the main body of the chapter, we use the result in Proposition 3.3 to find the optimal N for the two-fund rule as

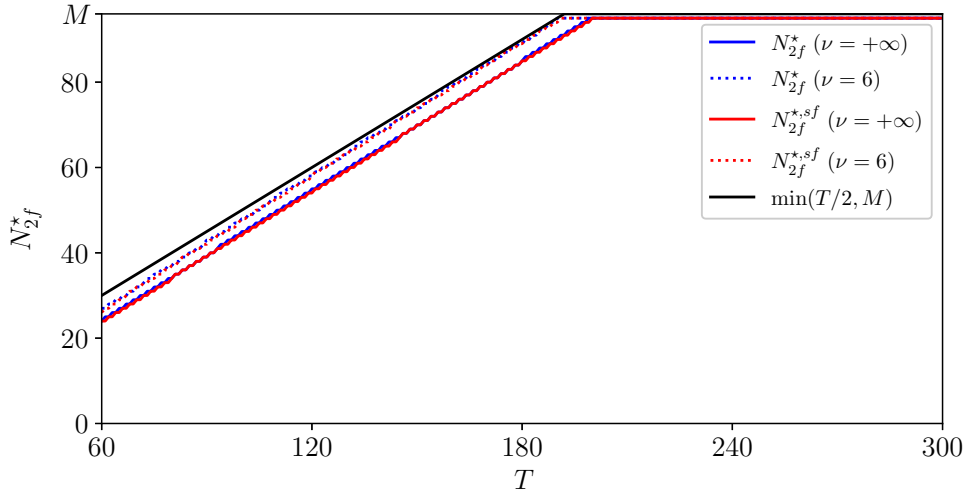
$$N_{2f}^{\star,sf} = \underset{N \in \{1,\ldots,\min(M,T-5)\}}{\operatorname{argmax}} EU_{2f}^{sf}(N, T, \sigma_f^2, \sigma_\varepsilon^2, \tau, \bar{\boldsymbol{\ell}}_M). \tag{3.38}$$

To determine $N_{2f}^{\star,sf}$, we must compute $\bar{\boldsymbol{\ell}}_M$, σ_f^2, and σ_ε^2. To do so, we use the Fama-French market excess return as the factor f and we regress the M asset returns on f via OLS to obtain the β_i's. For each asset, we obtain a residual variance $\sigma_{\varepsilon_i}^2$, and we set $\sigma_\varepsilon^2 = \frac{1}{M}\sum_{i=1}^M \sigma_{\varepsilon_i}^2$.

Having derived the oracle optimal portfolio size for the optimal two-fund rule obtained under the single factor model assumption, $N_{2f}^{\star,sf}$ in (3.38), we now compare it to the oracle optimal portfolio size under the equicorrelation assumption, $N_{2f}^{\star}$ in (3.22). We proceed similarly to Figure 3.2 in the main body of the chapter and calibrate the parameters to the full sample of the 96S-BM dataset, which yields $\bar{\boldsymbol{\ell}}_M = (6.07 \times 10^{-5}, 1.229, 0.0082)$, $\sigma_\varepsilon^2 = 0.0014$, $\bar{\boldsymbol{\theta}}_M = (0.125, 0.0169)$, $\rho = 0.74$. Moreover, the variance of the market factor over the period July 1963 to August 2023 is $\sigma_f^2 = 0.0020$. As in Figure 3.2, we depict the optimal N as a function of the sample size T between 60 and 300 when the returns are multivariate t-distributed with $\nu = \infty$ (i.e., normality) or $\nu = 6$ (i.e., an excess kurtosis of three).

The results are depicted in Figure 3.6 and show that, remarkably, the equicorrelation and single-factor model assumptions yield essentially equivalent optimal portfolio sizes for all sample sizes T. This can be explained because, under both assumptions, the maximum

Figure 3.6: Optimal portfolio size for the two-fund rule under the single-factor model and equicorrelation assumptions



Notes. This figure depicts the optimal portfolio size for the optimal two-fund rule as a function of the sample size T under two different assumptions. First, in blue, the equicorrelation assumption, i.e., Assumption 1, which yields $N^\star_{2f}$ in (3.22). Second, in red, the single-factor model assumption, i.e., Assumption 3, which yields $N^{\star,sf}_{2f}$ in (3.38). We assume the asset are multivariate t-distributed with $\nu = \infty$ (solid lines) and $\nu = 6$ (dashed lines) degrees of freedom. We calibrate the parameters on which $N^\star_{2f}$ and $N^{\star,sf}_{2f}$ depend to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023, which yields $\bar{\boldsymbol{\ell}}_M = (6.07 \times 10^{-5}, 1.229, 0.0082)$, $\sigma^2_\varepsilon = 0.0014$, $\bar{\boldsymbol{\theta}}_M = (0.125, 0.0169)$, $\rho = 0.74$. The single factor is the Fama-French market excess return, which has a variance $\sigma^2_f = 0.0020$. We set a risk-aversion coefficient $\gamma = 1$.

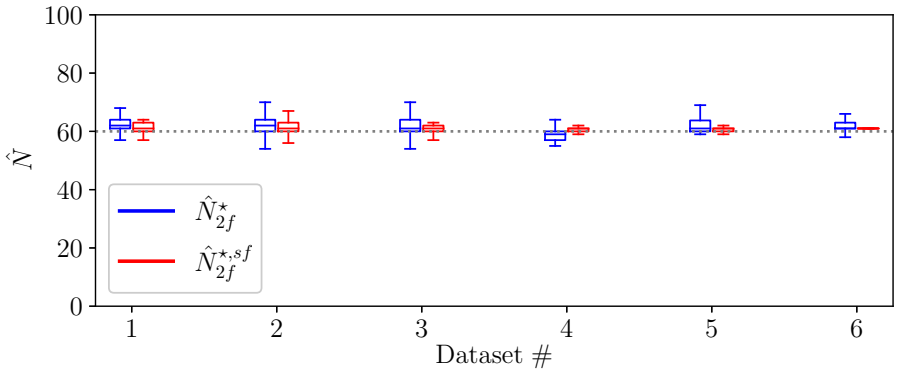
squared Sharpe ratio θ^2_N is equal to N multiplied by a factor that has little sensitivity to N. Engle and Kelly (2012, p. 213) also note that equicorrelation and single-factor covariance matrices are consistent: "In a one-factor world, [...] if the cross-sectional dispersion of β_j is small and idiosyncrasies have similar variance over each period, then the system is well described by Dynamic Equicorrelation."

Finally, we test whether the estimated optimal portfolio sizes under the two assumptions are similar in empirical data too. That is, in each rolling window of size $T = 120$ months and for all datasets listed in Table 3.1, we compute the estimated counterpart of $N^{\star,sf}_{2f}$ in (3.38), $\hat{N}^{\star,sf}_{2f}$. We obtain it by estimating $(\sigma^2_f, \sigma^2_\varepsilon, \bar{\boldsymbol{\ell}}_M)$ as explained above and the parameters $(k_{N,1}, k_{N,2}, k_{N,3})$ as explained in Section 3.E, and compare it to $\hat{N}^\star_{2f}$ in the main body of the chapter.

We depict boxplots of the estimated optimal portfolio sizes under the single-factor model and equicorrelation assumptions in Figure 3.7. As in Figure 3.6, we find that the two assumptions deliver similar results, which confirms the robustness of the optimal N found under equicorrelation. Also, we observe that the single-factor estimator, $\hat{N}^{\star,sf}_{2f}$, has a lower variance than $\hat{N}^\star_{2f}$, because the parameters $(\sigma^2_f, \sigma^2_\varepsilon, \bar{\boldsymbol{\ell}}_M)$ have less variability than $(\rho, \bar{\boldsymbol{\theta}}_M)$.[34]

[34]In unreported results, we find that the net out-of-sample utility delivered by the restricted two-fund rule is similar using either $\hat{N}^{\star,sf}_{2f}$ or $\hat{N}^\star_{2f}$ for all selection rules in Table 3.3.

Figure 3.7: Estimated optimal portfolio size in empirical data



Notes. This figure depicts boxplots of the estimated optimal portfolio size for the two-fund rule under two different assumptions: the equicorrelation assumption ($\hat{N}^{\star}_{2f}$) and the single factor assumption ($\hat{N}^{\star,sf}_{2f}$). These are the estimated counterparts of $N^{\star}_{2f}$ in (3.22) and $N^{\star,sf}_{2f}$ in (3.38), respectively. The boxplots are obtained by implementing the estimation methodology in Section 3.E of the supplementary material for each dataset listed in Table 3.1. We use a sample size $T = 120$ and set $\gamma = 1$.

3.B Equicorrelation and the maximum utility

Proposition 3.1 shows that under Assumption 1, i.e., all correlations among assets are equal to ρ, the maximum utility delivered by the MV portfolio is

$$U(\boldsymbol{w}^{\star}) = \frac{N}{2\gamma(1-\rho)}\left(\bar{\theta}_{N,2} - \frac{N\rho}{1-\rho+N\rho}\bar{\theta}^2_{N,1}\right), \tag{3.39}$$

where $-\frac{1}{N-1} < \rho < 1$. As discussed in Section 3.2, $U(\boldsymbol{w}^{\star})$ is an increasing function of N. However, it is not a monotonic function of the correlation ρ. In the next proposition, we demonstrate that $U(\boldsymbol{w}^{\star})$ is a convex function of ρ and we derive the minimizer.

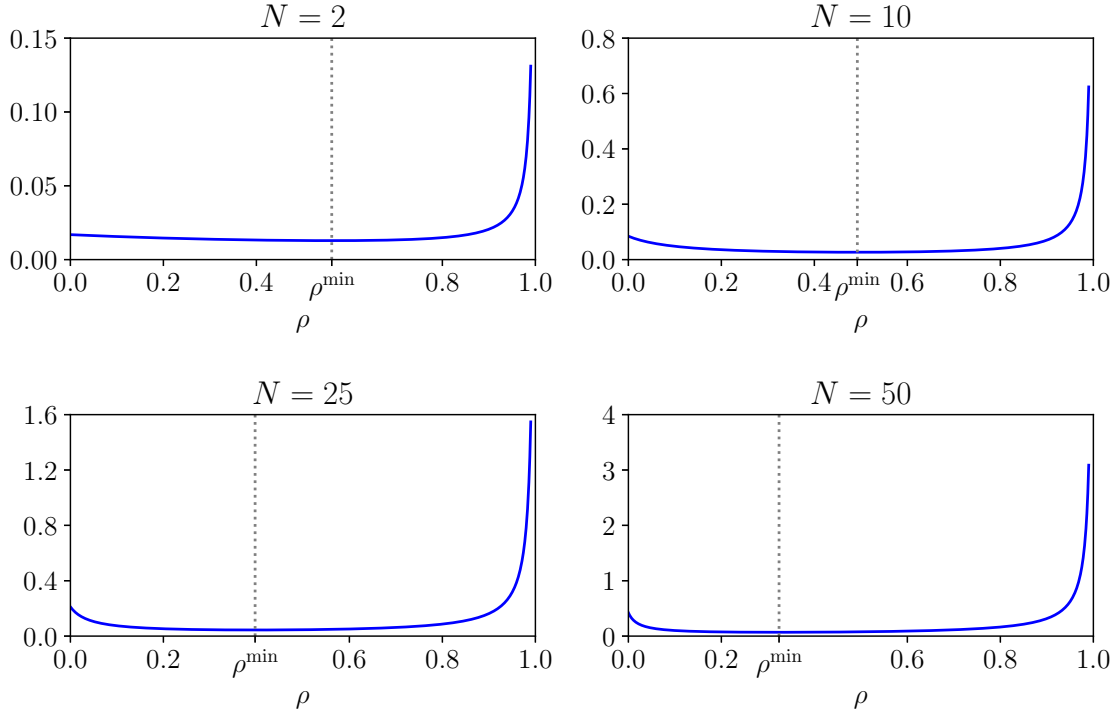
Proposition 3.4. *The maximum utility $U(\boldsymbol{w}^{\star})$ in (3.39) is a convex function of $\rho \in (-\frac{1}{N-1}, 1)$ and is minimized by*[35]

$$\rho_{\min} = \frac{\frac{N}{\sqrt{N-1}}\sqrt{\bar{\theta}^2_{N,1}(\bar{\theta}_{N,2} - \bar{\theta}^2_{N,1})} - \bar{\theta}_{N,2}}{N(\bar{\theta}_{N,2} - \bar{\theta}^2_{N,1}) - \bar{\theta}_{N,2}}. \tag{3.40}$$

Figure 3.8 illustrates Proposition 3.4. We calibrate $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ to the 96S-BM dataset, which yields $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2}) = (0.125, 0.0169)$, and we depict $U(\boldsymbol{w}^{\star})$ in (3.39) as a function of ρ for $N \in \{2, 10, 25, 50\}$ and $\gamma = 1$. The figure shows that $U(\boldsymbol{w}^{\star})$ is a convex function of ρ, decreases with ρ for $\rho < \rho_{\min}$, and increases with ρ for $\rho > \rho_{\min}$. This result extends a similar finding in Gandy and Veraart (2013, p.536), who also find that the maximum utility "[...] is initially decreasing in ρ and increasing afterwards.", and which we extend to the case where asset returns have different variances.

[35] It holds that $\lim_{N \to \bar{\theta}_{N,2}/(\bar{\theta}_{N,2} - \bar{\theta}^2_{N,1})} \rho_{\min} = 1 - \bar{\theta}_{N,2}/(2\bar{\theta}^2_{N,1})$, see the proof of Proposition 3.4.

Figure 3.8: Impact of correlation on maximum utility



Notes. This figure depicts the maximum utility $U(\boldsymbol{w}^\star)$ in (3.39) under the assumption that all assets have equal correlations ρ. We calibrate $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ to a dataset of 96 portfolios sorted on size and book-to-market spanning July 1963 to August 2023 to the 96S-BM dataset, which yields $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2}) = (0.125, 0.0169)$. We depict $U(\boldsymbol{w}^\star)$ as a function of ρ for $N \in \{2, 10, 25, 50\}$ and a risk-aversion coefficient $\gamma = 1$. The minimizer of $U(\boldsymbol{w}^\star)$ is $\rho_{\min}$ given by (3.40).

3.C Optimal portfolio size for three-fund rules

In Section 3.3, we study the optimal portfolio size N for the SMV portfolio and two-fund rule. We now study the optimal N for the *three-fund rules* that add another portfolio rule to further improve the EU. Specifically, we consider in Section 3.C.1 a three-fund rule based on the GMV portfolio as in, e.g., Kan and Zhou (2007), DeMiguel et al. (2015), and Yuan and Zhou (2024), which we call the *GMV-three-fund rule*. In Section 3.C.2, we consider a three-fund rule based on the EW portfolio as in, e.g., Tu and Zhou (2011), Kan and Wang (2023), and Lassance et al. (2024b), which we call the *EW-three-fund rule*.

3.C.1 Three-fund rule with the global minimum-variance portfolio

We include the GMV portfolio as an additional robust portfolio rule to invest in,

$$\boldsymbol{w}_g = \frac{\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}. \tag{3.41}$$

We introduce the following parameters:

$$\mu_{g,N} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N},\ \sigma_{g,N}^2 = \frac{1}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N},\ \lambda_{g,N} = \frac{\mu_{g,N}}{\sigma_{g,N}^2},\ \theta_{g,N}^2 = \frac{\mu_{g,N}^2}{\sigma_{g,N}^2},\ \psi_{g,N}^2 = \theta_N^2 - \theta_{g,N}^2, \tag{3.42}$$

which stand for the mean return, variance, price of risk, squared Sharpe ratio, and inefficiency of the GMV portfolio $\boldsymbol{w}_g$ computed on N assets. Under parameter uncertainty, we define the three-fund rule that invests in the SMV portfolio, the sample GMV (SGMV) portfolio, and the risk-free asset as

$$\hat{\boldsymbol{w}}(\boldsymbol{\alpha}) = \frac{\alpha_1}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N + \frac{\alpha_2}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\mathbf{1}_N, \tag{3.43}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{R}^2$ is the vector of combination coefficients. We call $\hat{\boldsymbol{w}}(\boldsymbol{\alpha})$ the *GMV-three-fund rule*. In the next proposition, we exploit Proposition 3.2 and Kan and Lassance (2025, Proposition 8) to derive the EU of the GMV-three-fund rule in (3.43) when asset returns are elliptically distributed, as well as the resulting optimal combination coefficients $\boldsymbol{\alpha}^\star$ and which EU they deliver.

Proposition 3.5. *Let $T > N+4$ and Assumption 2 hold. Then, the expected out-of-sample utility of the GMV-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\alpha})$ in* (3.43) *is*

$$\begin{aligned} EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha})) = \frac{1}{2\gamma}\frac{T}{T-N-2}\Bigg[2k_{N,1}\left(\alpha_1\theta_N^2 + \alpha_2\lambda_{g,N}\right) - \frac{c_N T}{T-N-2} \\ \times \left(\alpha_1^2\left(k_{N,2}\theta_N^2 + k_{N,3}\frac{N}{T}\right) + \frac{\alpha_2^2 k_{N,2}}{\sigma_{g,N}^2} + 2\alpha_1\alpha_2 k_{N,2}\lambda_{g,N}\right)\Bigg]. \end{aligned} \tag{3.44}$$

Moreover, the optimal combination coefficients $\boldsymbol{\alpha}^\star = (\alpha_1^\star, \alpha_2^\star)$ maximizing (3.44) *are*

$$(\alpha_1^\star, \alpha_2^\star) = \left(\frac{T-N-2}{c_N T}\frac{k_{N,1}\psi_{g,N}^2}{k_{N,2}\psi_{g,N}^2 + k_{N,3}\frac{N}{T}}, \frac{T-N-2}{c_N T}\frac{k_{N,3}}{k_{N,2}}\frac{k_{N,1}\frac{N}{T}\mu_{g,N}}{k_{N,2}\psi_{g,N}^2 + k_{N,3}\frac{N}{T}}\right), \tag{3.45}$$

and the resulting expected out-of-sample utility is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha}^\star)) = \frac{k_{N,1}^2}{2\gamma c_N}\frac{\theta_N^2\psi_{g,N}^2 + \frac{k_{N,3}}{k_{N,2}}\frac{N}{T}\theta_{g,N}^2}{k_{N,2}\psi_{g,N}^2 + k_{N,3}\frac{N}{T}}. \tag{3.46}$$

Proposition 3.5 shows that for an investor holding the optimal GMV-three-fund rule, the optimal portfolio size N is found by maximizing the EU in (3.46). We proceed by expressing θ_N^2 and $\theta_{g,N}^2$ as explicit functions of N by assuming that the covariance matrix $\boldsymbol{\Sigma}_N$ complies with Assumption 1. This is done in Proposition 3.1 for θ_N^2, and we now derive the expression for $\theta_{g,N}^2$.

Proposition 3.6. *Under Assumption 1, the squared Sharpe ratio of the GMV portfolio is*

$$\theta_{g,N}^2 = \frac{N}{1-\rho}r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N), \quad r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N) = \frac{\left(\bar{\lambda}_N - \frac{N\rho}{1-\rho+N\rho}\bar{\theta}_{N,1}\bar{\sigma}_{N,-1}\right)^2}{\bar{\sigma}_{N,-2} - \frac{N\rho}{1-\rho+N\rho}\bar{\sigma}_{N,-1}^2}, \tag{3.47}$$

where $\bar{\theta}_{N,1}$ is defined in (3.5)*, $\bar{\boldsymbol{\sigma}}_N = (\bar{\sigma}_{N,-1}, \bar{\sigma}_{N,-2})$, and*

$$\bar{\lambda}_N = \frac{1}{N}\sum_{i=1}^N \lambda_i, \quad \bar{\sigma}_{N,k} = \frac{1}{N}\sum_{i=1}^N \sigma_i^k, \tag{3.48}$$

with $\lambda_i = \mu_i/\sigma_i^2$ the price of risk of asset i.

Using Propositions 3.1 and 3.6, the EU of the GMV-three-fund rule in (3.46) becomes, under Assumption 1,

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha}^\star)) = \frac{Nk_{N,1}^2}{2\gamma c_N(1-\rho)} \frac{\delta_N(\bar{\boldsymbol{\theta}}_N)\left(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)\right) + \frac{k_{N,3}}{k_{N,2}}\frac{1-\rho}{T} r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)}{k_{N,2}\left(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)\right) + k_{N,3}\frac{1-\rho}{T}} \tag{3.49}$$

$$=: EU_{3f,g}(N, T, \rho, \tau, \bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N). \tag{3.50}$$

Then, as in Section 3.3, we replace $(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)$ by $(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$ and find the optimal N as

$$N_{3f,g}^\star = \underset{N\in\{1,\dots,\min(M,T-5)\}}{\operatorname{argmax}} EU_{3f,g}(N, T, \rho, \tau, \bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M). \tag{3.51}$$

Note that, similar to the two-fund rule, the objective function $EU_{3f,g}$ is proportional to N/c_N, which as shown in (3.23)–(3.24) is maximized by N slightly below $T/2$.[36] The simulations in Section 3.4 show that $N_{3f,g}^\star$ is similar to $N_{2f}^\star$ and is slightly below $T/2$ when ρ is not too small, as it is the case with equity data.

3.C.2 Three-fund rule with the equally weighted portfolio

We now consider the EW portfolio as an additional portfolio rule to invest in,

$$\boldsymbol{w}_{ew} = \frac{\mathbf{1}_N}{N}, \tag{3.52}$$

which benefits from no parameter uncertainty. We introduce the following parameters similar to the case with the GMV-three-fund rule:

$$\mu_{ew,N} = \boldsymbol{w}_{ew}'\boldsymbol{\mu}_N,\ \sigma_{ew,N}^2 = \boldsymbol{w}_{ew}'\boldsymbol{\Sigma}_N\boldsymbol{w}_{ew},\ \lambda_{ew,N} = \frac{\mu_{ew,N}}{\sigma_{ew,N}^2},\ \theta_{ew,N}^2 = \frac{\mu_{ew,N}^2}{\sigma_{ew,N}^2},\ \psi_{ew,N}^2 = \theta_N^2 - \theta_{ew,N}^2. \tag{3.53}$$

Under parameter uncertainty, we define the three-fund rule that invests in the SMV portfolio, the EW portfolio, and the risk-free asset as

$$\hat{\boldsymbol{w}}(\boldsymbol{\beta}) = \frac{\beta_1}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N + \frac{\beta_2}{\gamma}\boldsymbol{w}_{ew}, \tag{3.54}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbb{R}^2$ is the vector of combination coefficients. We call $\hat{\boldsymbol{w}}(\boldsymbol{\beta})$ the *EW-three-fund rule*. In the next proposition, we exploit Proposition 3.2 to derive the EU of the EW-three-fund rule (3.43) when asset returns are elliptically distributed, as well as the resulting optimal combination coefficients $\boldsymbol{\alpha}^\star$ and which EU they deliver.

[36] That the EU of both the optimal two-fund rule and the GMV-three-fund rule is proportional to N/c_N can be explained as follows: the proportionality to $1/c_N$ is because both the SMV portfolio and the SGMV portfolio are proportional to $\hat{\boldsymbol{\Sigma}}_N^{-1}$, and the proportionality to N is because both the squared Sharpe ratio of the MV portfolio (θ_N^2) and the GMV portfolio ($\theta_{g,N}^2$) are proportional to N under Assumption 1.

Proposition 3.7. *Let $T > N+4$ and Assumption 2 hold. Then, the expected out-of-sample utility of the EW-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\beta})$ in (3.54) is*

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\beta})) = \frac{1}{2\gamma} \times \left[\frac{2\beta_1 k_{N,1}\theta_N^2 T}{T-N-2} + 2\beta_2 \mu_{ew,N} - \frac{\beta_1^2 c_N T^2}{(T-N-2)^2}\left(k_{N,2}\theta_N^2 + k_{N,3}\frac{N}{T}\right) - \beta_2^2 \sigma_{ew,N}^2 - \frac{2\beta_1\beta_2 k_{N,1}\mu_{ew,N}T}{T-N-2}\right]. \tag{3.55}$$

Moreover, the optimal combination coefficients $\boldsymbol{\beta}^\star = (\beta_1^\star, \beta_2^\star)$ maximizing (3.55) are

$$(\beta_1^\star, \beta_2^\star) = \left(\frac{T-N-2}{T} \frac{k_{N,1}\psi_{ew,N}^2}{k_{N,1}^2\psi_{ew,N}^2 + d_N}, \frac{d_N \lambda_{ew,N}}{k_{N,1}^2\psi_{ew,N}^2 + d_N} \right), \tag{3.56}$$

where $d_N = c_N k_{N,3}\frac{N}{T} + (c_N k_{N,2} - k_{N,1}^2)\theta_N^2$, and the resulting expected out-of-sample utility is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\beta}^\star)) = \frac{1}{2\gamma} \frac{k_{N,1}^2\theta_N^2\psi_{ew,N}^2 + d_N\theta_{ew,N}^2}{k_{N,1}^2\psi_{ew,N}^2 + d_N}. \tag{3.57}$$

The EU of the optimal EW-three-fund rule depends on θ_N^2 and $\theta_{ew,N}^2$. As usual, we express θ_N^2 and $\theta_{ew,N}^2$ as explicit functions of N by assuming that the covariance matrix $\boldsymbol{\Sigma}_N$ complies with Assumption 1. We derive the expression for $\theta_{ew,N}^2$ in the next proposition.

Proposition 3.8. *Under Assumption 1, the squared Sharpe ratio of the EW portfolio is*

$$\theta_{ew,N}^2 = \frac{N}{1-\rho} r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}), \quad r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) = \frac{(1-\rho)\bar{\mu}_N^2}{(1-\rho)\bar{\sigma}_{N,2} + \rho N \bar{\sigma}_{N,1}^2}, \tag{3.58}$$

where $\bar{\sigma}_{N,1}$ and $\bar{\sigma}_{N,2}$ are defined in (3.48) and

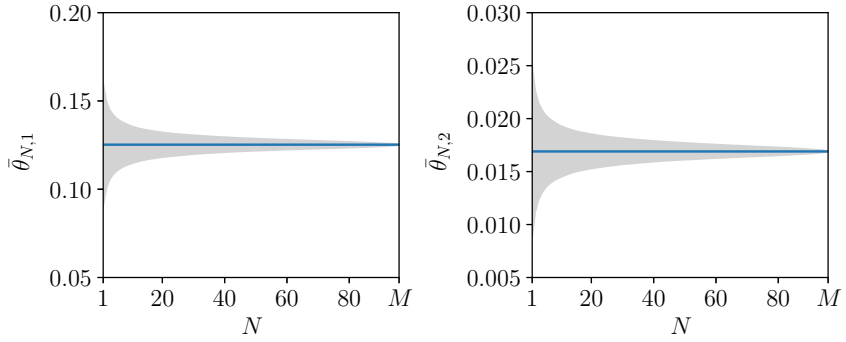
$$\bar{\mu}_N = \frac{1}{N}\sum_{i=1}^N \mu_i. \tag{3.59}$$

Using Propositions 3.1 and 3.8, the EU of the EW-three-fund rule in (3.57) becomes, under Assumption 1,

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha}^\star)) = \frac{N}{2\gamma(1-\rho)} \times \frac{k_{N,1}^2\delta_N(\bar{\boldsymbol{\theta}}_N)\Big(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})\Big) + \Big(\frac{(1-\rho)c_N k_{N,3}}{T} + (c_N k_{N,2} - k_{N,1}^2)\delta_N(\bar{\boldsymbol{\theta}}_N)\Big) r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})}{k_{N,1}^2\Big(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})\Big) + \Big(\frac{(1-\rho)c_N k_{N,3}}{T} + (c_N k_{N,2} - k_{N,1}^2)\delta_N(\bar{\boldsymbol{\theta}}_N)\Big)} \tag{3.60}$$

$$=: EU_{3f,ew}(N, T, \rho, \tau, \bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}). \tag{3.61}$$

Figure 3.9: Impact of the sequence of assets on $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$



Notes. This figure depicts the values of $\bar{\theta}_{N,1}$ (left panel) and $\bar{\theta}_{N,2}$ (right panel) as a function of the portfolio size N. The solid blue line depicts $\bar{\theta}_{M,1}$ and $\bar{\theta}_{M,2}$, and the shaded gray area depicts the one standard deviation interval of the values of $\bar{\theta}_{N,1}$ and $\bar{\theta}_{N,2}$ obtained under 10,000 random sequences of assets. We calibrate this experiment to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.

Finally, we replace $(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})$ by $(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ and find the optimal N as

$$N^{\star}_{3f,ew} = \underset{N \in \{1,\ldots,\min(M,T-5)\}}{\operatorname{argmax}} EU_{3f,ew}(N, T, \rho, \tau, \bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2}). \tag{3.62}$$

Although it does not appear as clearly as for the two-fund rule and the GMV-three-fund rule via N/c_N as a proportionality factor in the objective function, the simulations in Section 3.4 also show that $N^{\star}_{3f,ew}$ typically hovers slightly below $T/2$.

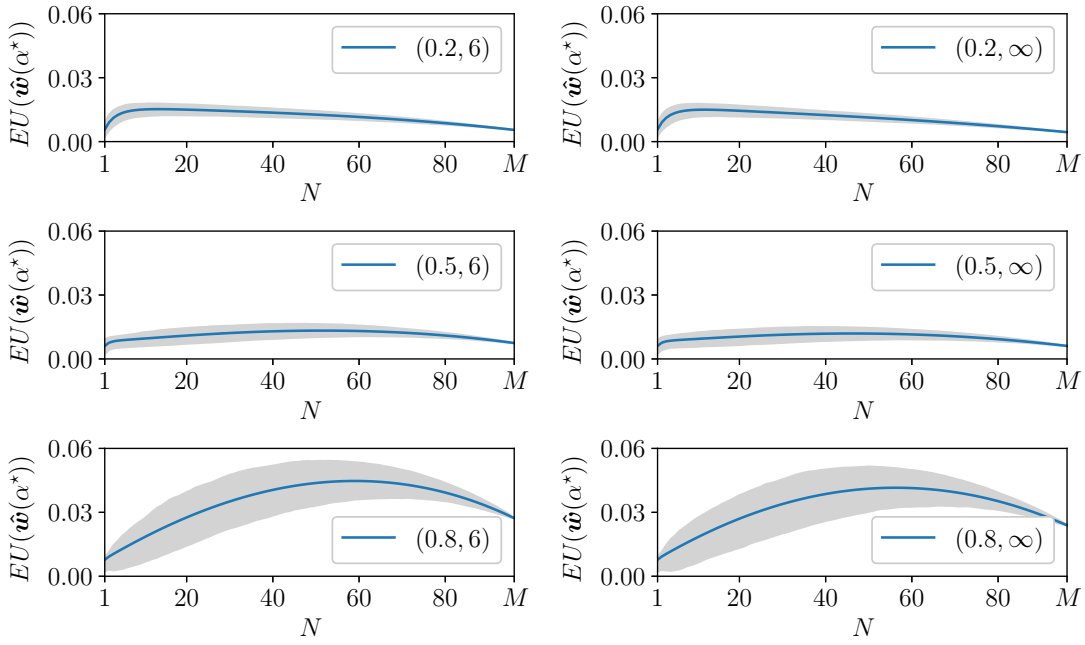
3.D Impact of the sequence of assets on expected utility

In this section, we study the impact of the sequence in which the assets are picked on $\bar{\boldsymbol{\theta}}_N$, which is defined in Proposition 3.1, and on the EU of the two-fund rule, $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ in (3.20). Specifically, we study how close $\bar{\boldsymbol{\theta}}_N$ is to $\bar{\boldsymbol{\theta}}_M$ and how close our EU approximation $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ in (3.22) is to $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ as a function of N.

To do so, we consider the following experiment in which $\bar{\boldsymbol{\theta}}_N$ is calibrated to the 96S-BM dataset. First, we randomly select $N = 1$ asset out of the M available ones and evaluate $\bar{\boldsymbol{\theta}}_N$ and $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ assuming t-distributed returns, i.e., $\tau^2 \sim (\nu - 2)/\chi^2_\nu$. Then, we randomly select a new asset on top of the first one and compute $\bar{\boldsymbol{\theta}}_N$ and the EU on the $N = 2$ assets. We follow this approach up until $N = M$, which allows us to depict $\bar{\boldsymbol{\theta}}_N$ and the EU as a function of N for a specific sequence of assets. We repeat this exercise $10,000$ times. To compute the EU, we take $\rho \in \{0.2, 0.5, 0.8\}$ and $\nu \in \{6, \infty\}$.

The results from this experiment are depicted in two figures. First, in Figure 3.9, we depict $\bar{\boldsymbol{\theta}}_M$ in solid blue and a one standard deviation interval of the 10,000 values of $\bar{\boldsymbol{\theta}}_N$ in shaded gray, as a function of N. We observe that the interval quickly shrinks as N increases. Second, in Figure 3.10, we depict, for a sample size $T = 120$, $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ in solid blue and a one standard deviation interval of the 10,000 values of $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ in shaded gray, as a function of N. We can observe that the gray area closely follows our curve for determining the optimal $N^{\star}_{2f}$, $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$.

Figure 3.10: Impact of the sequence of assets on the EU of the optimal two-fund rule



Notes. This figure depicts our approximation of the EU of the optimal two-fund rule $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ in (3.22) as a function of the portfolio size N (solid blue line). We also depict in shaded gray the one standard deviation interval of the 10,000 values of $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ obtained with random sequences of assets. We calibrate $(k_{N,1}, k_{N,2}, k_{N,3})$ to the t-distribution with $\nu \in \{6, \infty\}$ degrees of freedom, and consider $\rho \in \{0.2, 0.5, 0.8\}$. Each panel corresponds to a choice of (ρ, ν). We set $\gamma = 1$. We calibrate this experiment to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.

3.E Estimation of parameters

In this section, we explain how we estimate the different parameters that are needed as inputs in our different portfolio rules to determine the optimal portfolio size (i.e., $N^\star_{smv}$, $N^\star_{2f}$, $N^\star_{3f,g}$, $N^\star_{3f,ew}$) and the optimal combination coefficients (i.e., $\alpha^\star$, $\alpha^\star_1$, $\alpha^\star_2$, $\beta^\star_1$, $\beta^\star_2$).

3.E.1 Estimation of elliptical fat tails

The first set of parameters are those that determine the impact of the fat tails of the elliptical distribution, i.e., $k_{N,1}$, $k_{N,2}$, and $k_{N,3}$ in (3.13)–(3.15). To speed up the computation, we use the high-dimensional approximation of these parameters. Specifically, from El Karoui (2010, 2013), we have that as $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, $(k_{N,1}, k_{N,2}, k_{N,3}) \to (\eta_N, \varphi_N, \eta_N)$, where $\eta_N \geq 1$ is the unique positive solution to

$$\mathbb{E}\left[(1 - \phi + \phi\eta_N\tau)^{-1}\right] = 1, \tag{3.63}$$

and $\varphi_N \geq \eta_N^2$ is given by

$$\varphi_N = (1 - \phi)\left(\eta_N^{-2} - \mathbb{E}\left[\frac{\phi\tau^2}{(1 - \phi + \phi\eta_N\tau)^2}\right]\right)^{-1}. \tag{3.64}$$

Kan and Lassance (2025) show that this high-dimensional approximation is accurate for typical values of N and T.

Given this result, we need to estimate η_N and φ_N. We estimate them in two different ways as in Kan and Lassance (2025). First, we assume that the asset returns follow a t-distribution, i.e., $\tau^2 \sim (\nu-2)/\chi^2_\nu$, and we estimate the number of degrees of freedom ν by maximum likelihood from a sample of T historical returns $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$, giving us $\hat{\nu}$. Then, we can use the closed-form expression for η_N and φ_N when asset returns are t-distributed in Kan and Lassance (2025, Proposition 6), which yields that the estimate of η_N, denoted η_N^ν, is the unique positive solution to

$$ye^y E_{\hat{\nu}/2}(y) = \phi_N \quad \text{with} \quad y = \frac{(\hat{\nu}-2)\phi_N \eta_N^\nu}{2(1-\phi_N)} \quad \text{and} \quad \phi_N = \frac{N}{T}, \tag{3.65}$$

where $E_n(x) = \int_1^\infty t^{-n} e^{-xt} \mathrm{d}t$ is the exponential integral, and the estimate of φ_N is

$$\varphi_N^\nu = \frac{2(\eta_N^\nu)^2(1-\phi_N)}{\hat{\nu} - \eta_N^\nu(\hat{\nu}-2)}. \tag{3.66}$$

We use this first estimation method in Section 3.4 given that we simulate returns from a t-distribution.

The second method we use to estimate η_N and φ_N does not specify a particular parametric distribution for τ, but instead, relies on the following sample estimate of the distribution of τ proposed by El Karoui (2010, 2013):

$$\hat{\tau}_{N,t} = \frac{(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)'(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)}{\frac{1}{T}\sum_{i=1}^T (\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_N)'(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_N)}, \quad t = 1, \ldots, T, \tag{3.67}$$

which is consistent as $N \to \infty$. Using $\hat{\tau}_{N,t}$, we estimate η_N and φ_N with their sample counterparts, i.e., the estimate of η_N, denoted η_N^s, is the unique positive solution to

$$\frac{1}{T}\sum_{t=1}^T \left(1 - \phi_N + \phi_N \eta_N^s \hat{\tau}_{N,t}\right)^{-1} = 1, \tag{3.68}$$

and the estimate of φ_N is

$$\varphi_N^s = (1-\phi_N)\left((\eta_N^s)^{-2} - \frac{1}{T}\sum_{t=1}^T \frac{\phi_N \hat{\tau}_{N,t}^2}{(1-\phi_N+\phi_N\eta_N^s\hat{\tau}_{N,t})^2}\right)^{-1}. \tag{3.69}$$

We use this second estimation method in Section 3.5 to have more freedom in describing the tails of empirical data that may not be t-distributed.

3.E.2 Estimation of portfolios' performance

The second set of parameters are those that determine the performance of the MV, GMV, and EW portfolios and on which the optimal combination coefficients depend: θ_N^2 in (3.2), $\psi_{g,N}^2$ in (3.42), $\psi_{ew,N}^2$ in (3.53), $\mu_{g,N}$ in (3.42), and $\lambda_{ew,N}$ in (3.53).

For $\mu_{g,N}$ and $\lambda_{ew,N}$, we rely on the estimates that are unbiased when asset returns are normally distributed[37], i.e.,

$$\hat{\mu}_{g,N} = \frac{\mathbf{1}_N' \hat{\mathbf{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N}{\mathbf{1}_N' \hat{\mathbf{\Sigma}}_N^{-1} \mathbf{1}_N}, \tag{3.70}$$

$$\hat{\lambda}_{ew,N} = \frac{T-3}{T} \frac{\boldsymbol{w}_{ew}' \hat{\boldsymbol{\mu}}_N}{\boldsymbol{w}_{ew}' \hat{\mathbf{\Sigma}}_N \boldsymbol{w}_{ew}}. \tag{3.71}$$

For θ_N^2, $\psi_{g,N}^2$, and $\psi_{ew,N}^2$, we have that the sample estimates, obtained by plugging $(\hat{\boldsymbol{\mu}}_N, \hat{\mathbf{\Sigma}}_N)$, are severely biased. Therefore, we estimate them using the adjusted estimates in Kan and Zhou (2007) and Kan and Wang (2023) that correct the unbiased estimates to ensure they are positive. Specifically, let $\hat{\theta}_N^2 = \hat{\boldsymbol{\mu}}_N' \hat{\mathbf{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N$, $\hat{\psi}_{g,N}^2 = \hat{\theta}_N^2 - \hat{\theta}_{g,N}^2$, and $\hat{\psi}_{ew,N}^2 = \hat{\theta}_N^2 - \hat{\theta}_{ew,N}^2$ be the sample estimates of θ_N^2, $\psi_{g,N}^2$, and $\psi_{ew,N}^2$, where $\hat{\theta}_{g,N}^2 = (\mathbf{1}_N' \hat{\mathbf{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N)^2 / (\mathbf{1}_N' \hat{\mathbf{\Sigma}}_N^{-1} \mathbf{1}_N)$ and $\hat{\theta}_{ew,N}^2 = (\boldsymbol{w}_{ew}' \hat{\boldsymbol{\mu}}_N)^2 / (\boldsymbol{w}_{ew}' \hat{\mathbf{\Sigma}} \boldsymbol{w}_{ew})$. Then, the adjusted estimates are

$$\hat{\theta}_{N,a}^2 = \frac{(T-N-2)\hat{\theta}_N^2 - N}{T} + \frac{2(\hat{\theta}_N^2)^{\frac{N}{2}}(1+\hat{\theta}_N^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\theta}_N^2/(1+\hat{\theta}_N^2)}\left(\frac{N}{2}, \frac{T-N}{2}\right)}, \tag{3.72}$$

$$\hat{\psi}_{g,N,a}^2 = \frac{(T-N-1)\hat{\psi}_{g,N}^2 - (N-1)}{T} + \frac{2(\hat{\psi}_{g,N}^2)^{\frac{N-1}{2}}(1+\hat{\psi}_{g,N}^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\psi}_{g,N}^2/(1+\hat{\psi}_{g,N}^2)}\left(\frac{N-1}{2}, \frac{T-N+1}{2}\right)}, \tag{3.73}$$

$$\hat{\psi}_{ew,N,a}^2 = \frac{(T-N-2)\hat{\psi}_{ew,N}^2 - (N-1)(1+\hat{\theta}_{ew,N}^2)}{T} + \frac{2(1+\hat{\theta}_{ew,N}^2)^{\frac{T-N}{2}}(\hat{\psi}_{ew,N}^2)^{\frac{N-1}{2}}(1+\hat{\theta}_N^2)^{\frac{3-T}{2}}}{T \times B_{\hat{\psi}_{ew,N}^2/(1+\hat{\theta}_N^2)}\left(\frac{N-1}{2}, \frac{T-N}{2}\right)}, \tag{3.74}$$

where $B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t$ is the incomplete beta function.

3.E.3 Estimation of assets' correlation and marginal performance

The third and last set of parameters are those that control the dependence between the assets, i.e., ρ under Assumption 1, and the marginal performance of the assets, i.e., the three functions $\delta_N(\bar{\boldsymbol{\theta}}_M)$ in (3.4), $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$ in (3.47), and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ in (3.58). Recall that the parameters in these functions are computed on all M assets to find the optimal portfolio size N. Following the advice of Adams et al. (2017), we use an estimator $\hat{\rho}$ given by the average of all sample pairwise correlations $\hat{\rho}_{ij}$,

$$\hat{\rho} = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \hat{\rho}_{ij}. \tag{3.75}$$

Regarding the functions $\delta_N(\bar{\boldsymbol{\theta}}_M)$, $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$, and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$, we rely on the following proposition.

[37] We can show that when asset returns are elliptically distributed as in Assumption 2, $\hat{\mu}_{g,N}$ and $\hat{\lambda}_{ew,N}$ are also unbiased in the high-dimensional asymptotic regime as $N, T \to \infty$ and $N/T \to \phi \in (0,1)$.

Proposition 3.9. *Let the covariance matrix $\mathbf{\Sigma}_M$ be known and fulfill Assumption 1. Then, the biases of the estimators of $\delta_N(\bar{\boldsymbol{\theta}}_M)$, $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$, and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ obtained by plugging the sample mean $\hat{\boldsymbol{\mu}}_M$ are*

$$\mathbb{E}[\hat{\delta}_N] - \delta_N = \frac{1-\rho}{T} \times \frac{1 - \frac{N}{M}\rho + N\rho}{1-\rho+N\rho}, \tag{3.76}$$

$$\mathbb{E}[\hat{r}_{g,N}] - r_{g,N} = \frac{1-\rho}{MT} \times \frac{\bar{\sigma}_{M,-2} - \frac{N\rho}{1-\rho+N\rho}\left(1 + \frac{(1-\rho)(1-\frac{M}{N})}{1-\rho+N\rho}\right)\bar{\sigma}^2_{M,-1}}{\bar{\sigma}_{M,-2} - \frac{N\rho}{1-\rho+N\rho}\bar{\sigma}^2_{M,-1}}, \tag{3.77}$$

$$\mathbb{E}[\hat{r}_{ew,N}] - r_{ew,N} = \frac{1-\rho}{MT} \times \frac{\bar{\sigma}_{M,2} + \frac{\rho}{1-\rho}M\bar{\sigma}^2_{M,1}}{\bar{\sigma}_{M,2} + \frac{\rho}{1-\rho}N\bar{\sigma}^2_{M,1}}. \tag{3.78}$$

Building on Proposition 3.9, we estimate the function $\delta_N(\bar{\boldsymbol{\theta}}_M)$, $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$, and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ by plugging the sample mean $\hat{\boldsymbol{\mu}}_M$ and the sample covariance matrix $\hat{\mathbf{\Sigma}}_M$ and by removing the bias, which we estimate from $\hat{\boldsymbol{\mu}}_M$ and $\hat{\mathbf{\Sigma}}_M$ too.

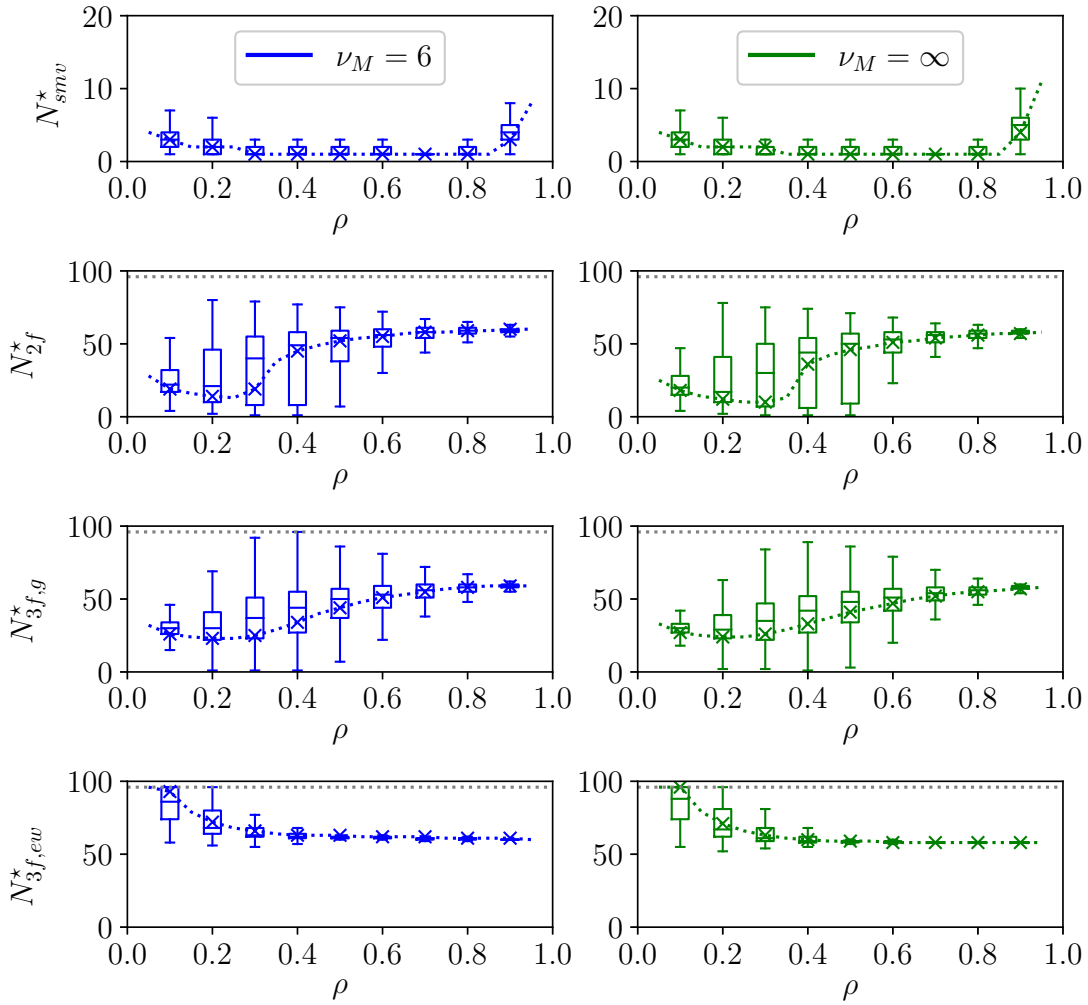
3.F Additional simulation and empirical results

We now report simulation and empirical results that complement those in the main body of the chapter. In the simulation setting, Section 3.F.1 looks at the impact of the correlation on the optimal portfolio size and Section 3.F.2 considers a block-correlation matrix. In the empirical setting, Section 3.F.3 looks at the equicorrelation and elliptical distribution assumptions, Section 3.F.4 considers a nonlinear shrinkage estimator of the covariance matrix, Section 3.F.5 applies our different asset selection rules to the SMV portfolio, Section 3.F.6 reports the out-of-sample Sharpe ratio of the considered portfolios, Section 3.F.7 evaluates the portfolio turnover and how much of it is driven by changes to the selected assets over time, and Section 3.F.8 investigates the overlap between the selection rules implemented in the main text.

3.F.1 Impact of correlation on the optimal portfolio size

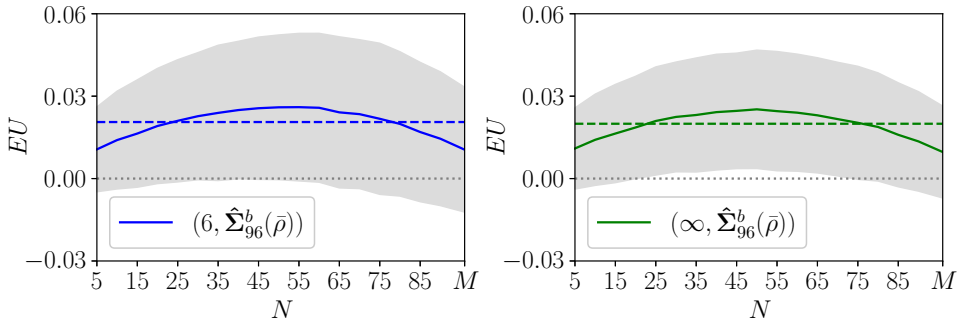
In this section, we run the same experiment as that in Section 3.4.1 except that we vary the correlation ρ. Specifically, we let $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}_{96}(\rho)$ and consider values of ρ between 0.1 and 0.9 with a step size of 0.1. Because $\mathbf{\Sigma}_M$ is of the form $\mathbf{\Sigma}_M(\rho)$, we can compare the estimated optimal N with the oracle optimal N. We depict the results in Figure 3.11. We observe that the oracle optimal N for the two-fund and three-fund rules is slightly below $T/2$ when ρ is close enough to one, as observed in Figure 3.2 where $\rho = \bar{\rho} = 0.74$. In this case of high correlation, typical for equity data, the estimated optimal N is remarkably robust to variability in the parameters as the boxplots are quite thin. However, as ρ decreases away from one, the optimal N moves away from $T/2$ and the boxplots widen as N becomes more sensitive to the parameters. Finally, we also observe that overall the boxplots for the estimated optimal N are centered close to the oracle N.

Figure 3.11: Impact of correlation ρ on optimal portfolio size in simulated data



Notes. This figure depicts boxplots of the estimated optimal portfolio size N for the SMV portfolio ($\hat{N}^\star_{smv}$), two-fund rule ($\hat{N}^\star_{2f}$), GMV-three-fund rule ($\hat{N}^\star_{3f,g}$), and EW-three-fund rule ($\hat{N}^\star_{3f,ew}$). These are the estimated counterparts of $N^\star_{smv}$ in (3.21), $N^\star_{2f}$ in (3.22), $N^\star_{3f,g}$ in (3.51), and $N^\star_{3f,ew}$ in (3.62) following the estimation methodology in Section 3.E of the supplementary material. The boxplots are obtained by simulating 10,000 times $T = 120$ t-distributed returns. Using the 96S-BM dataset, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\rho)$, where ρ varies between 0.1 and 0.9 with a step size of 0.1., and consider $\nu \in \{6, \infty\}$ degrees of freedom. Each boxplot corresponds to a different choice of (ν, ρ). We depict with dotted lines and '$\times$' markers the oracle value of the optimal N known under Assumption 1 because $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$.

Figure 3.12: EU of the two-fund rule in simulated data with a block correlation matrix



Notes. This figure depicts the expected out-of-sample utility (EU) of the two-fund rule as a function of the portfolio size N. We simulate 10,000 samples of t-distributed returns with $\nu \in \{6, \infty\}$ degrees of freedom. For each N, we conduct a rolling window exercise as described in Section 3.4.2. Using the 96S-BM dataset, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}^b_{96}(\bar{\rho})$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\mathbf{\Sigma}}_{96}$ and the construction of the block-correlation matrix is described in Section 3.F.2. Each panel corresponds to a different choice of ν. In each panel, the solid line depicts the EU in (3.25), the shaded gray area depicts the one standard deviation interval around the EU across all simulations, and the dashed horizontal line depicts the EU obtained when using the estimated optimal N for the two-fund rule, i.e., the estimate of $N^\star_{2f}$ in (3.22) according to the estimation method in Section 3.E of the supplementary material. The dotted horizontal gray line depicts the zero EU level. We set a risk-aversion coefficient of $\gamma = 1$.

3.F.2 Block-correlation matrix

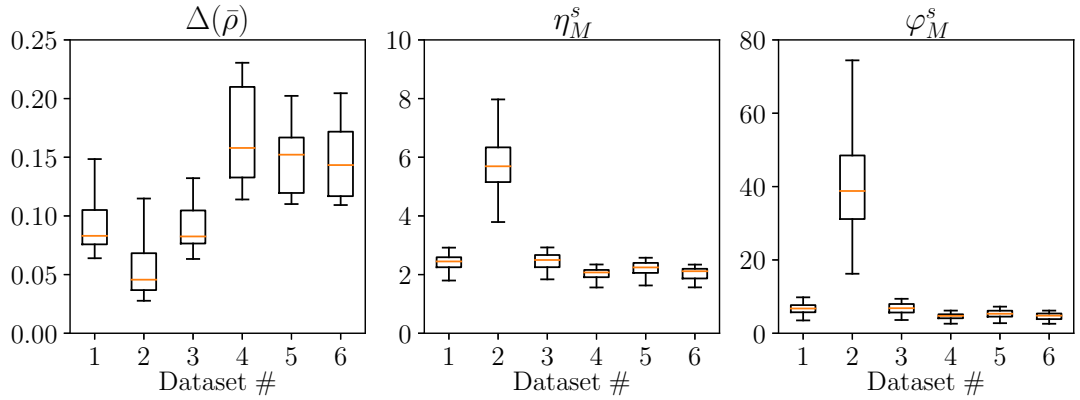
In this section, we run the same experiment as in Section 3.4.2 except that we use a different specification for the correlation matrix of asset returns. In Section 3.4.2, we consider two correlation structures: equicorrelation, i.e., we set $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}_{96}(\bar{\rho})$, and the empirical correlations we observe in the data, i.e., we set $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}_{96}$. To test the robustness of our approach, we now consider a third specification in which correlations form a block structure. Specifically, we set the population covariance matrix $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}^b_{96}(\bar{\rho}) = \hat{\boldsymbol{D}}_{96}\boldsymbol{P}^b_{96}(\bar{\rho})\hat{\boldsymbol{D}}'_{96}$ where $\boldsymbol{P}^b_{96}(\bar{\rho})$ is a block-correlation matrix. We consider two blocks of size $M/2$ for which the intra-correlation, i.e., correlation between assets of the same block, is $\bar{\rho} = 0.74$, and the inter-correlation, i.e., correlation between assets of different blocks, is zero. The two blocks are selected from the first and last 48 assets in the 96S-BM dataset. We depict the results in Figure 3.12. We observe that in this extreme setup, our estimated optimal N no longer delivers the best out-of-sample performance, but still largely outperforms investing in all or too few assets.

3.F.3 Equicorrelation and elliptical distribution in empirical data

We rely on two assumptions in our methodology to derive optimal portfolio combinations and optimal portfolio sizes: Assumption 1 states that asset returns are equicorrelated and Assumption 2 states that asset returns are elliptically distributed. We now look at these two assumptions in the six datasets we use in the empirical analysis of Section 3.5.

Regarding Assumption 1, we now define a measure of distance between the correlation matrix observed in the data and the equicorrelation matrix. Specifically, we estimate ρ as

Figure 3.13: Assumptions of equicorrelation and elliptical distribution in empirical data



Notes. This figure depicts boxplots of the estimated values of $\Delta(\bar{\rho})$ in (3.80), which measures the distance between the sample correlation matrix and the equicorrelation matrix, and η_M^s and φ_M^s in (3.68)–(3.69), which measure the impact of elliptical fat tails on the expected out-of-sample utility, for each dataset listed in Table 3.1 across all rolling estimation windows of size $T = 120$ months.

the average of sample correlations,

$$\bar{\rho} = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \hat{\rho}_{ij}, \tag{3.79}$$

where $\hat{\rho}_{ij}$ is the sample correlation between assets i and j, and we define $\Delta(\bar{\rho})$ as the root mean squared error between the sample correlation matrix and the equicorrelation matrix obtained by setting $\rho = \bar{\rho}$, i.e.,

$$\Delta(\bar{\rho}) := \sqrt{\frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} (\hat{\rho}_{ij} - \bar{\rho})^2}. \tag{3.80}$$

In the left panel of Figure 3.13, we depict, for each of the six datasets, boxplots of $\Delta(\bar{\rho})$ across all estimation windows of size $T = 120$ months. For three out of the six datasets (96S-BM, 100S-OP, 108CHA), $\Delta(\bar{\rho})$ is quite low and hovers around 5-10%. However, for the remaining datasets (107IN-CHA, 94IN-NV, 98IN-CHA-NV), $\Delta(\bar{\rho})$ is larger and around 10-20%. Thus, there are varying degrees to which the equicorrelation assumption is representative of the true correlation structure in the data, which makes the consistent outperformance delivered by our method to determine the optimal portfolio size particularly appreciable.

Regarding Assumption 2, we look at the three parameters $(k_{M,1}, k_{M,2}, k_{M,3})$ defined in (3.13)–(3.15) that control the impact of elliptical fat tails on the EU and that appear in the formulas for the optimal combination coefficients in the two-fund and three-fund rules. As explained in Section 3.E.1, in the empirical analysis we estimate them following El Karoui (2010, 2013), i.e., we estimate $k_{M,1}$ and $k_{M,3}$ with η_M^s in (3.68) and $k_{M,2}$ with φ_M^s in (3.69). The middle and right panels of Figure 3.13 depict boxplots of η_M^s and φ_M^s across all estimation windows. Recall that the closer they are to one, the closer the data is

to being normally distributed. We observe that η_M^s and φ_M^s substantially depart from one, particularly for the 108CHA dataset, but for the other datasets too. These results indicate that it is crucial to account for the fat tails of returns in determining optimal combination rules.

3.F.4 Nonlinear shrinkage covariance matrix

In Table 3.5, we replicate the results in Table 3.4 using the nonlinear shrinkage estimator of the covariance matrix of Ledoit and Wolf (2020). We observe similar results to those obtained under the linear shrinkage covariance matrix in Table 3.4, except for the 108CHA dataset where the nonlinear shrinkage estimator does not behave well.

3.F.5 Performance of optimally restricted SMV portfolio

In Table 3.6, we report the annualized net out-of-sample utility of the restricted SMV portfolio for which the portfolio size is determined using our theory in Section 3.3, i.e., using $\hat{N}^{\star}_{smv}$, and the assets are selected using the different selection rules in Section 3.5.3. We find for all asset selection rules that reducing the portfolio size substantially improves the performance of the SMV portfolio. For instance, for the 96S-BM dataset with the linear shrinkage covariance matrix, the SMV portfolio implemented on all assets delivers a net utility of -1443, whereas BWSR, which is the *worst* selection rule for that dataset, delivers a net utility of -33.3. However, because the SMV portfolio is too exposed to estimation risk, we still obtain negative utilities in most cases even with a reduced portfolio size. It is preferable instead to apply our method for determining the optimal portfolio size to a better reference portfolio like the two-fund and three-fund rules.

We now compare the obtained portfolio sizes for the three different ways of restricting the size of the SMV portfolio we consider in Table 3.6. First, using our estimated portfolio size $\hat{N}^{\star}_{smv}$ that we estimate following the methodology in Section 3.E of the supplementary material. Second, the implicit portfolio size obtained by implementing L_1-norm regularization as detailed in Section 3.5.2. Specifically, at each month t, we solve (3.29) by finding the optimal δ with cross-validation, which yields the vector of weights $\hat{\boldsymbol{w}}_{L_1,t}$, from which we can compute the implicit portfolio size $\hat{N}_{L_1,t}$,

$$\hat{N}_{L_1,t} = \sum_{i=1}^{M} \mathbb{1}_{\{|\hat{w}_{L_1,i,t}|>0\}}. \tag{3.81}$$

Third, we consider the portfolio size $\hat{N}_{\bar{w}}$ obtained implicitly by finding the optimal $\bar{w}$ with cross-validation, as detailed in Section 3.5.2. We depict the results in Figure 3.14 and observe that the portfolio sizes for the SMV portfolio are small on average compared to M, and that $\hat{N}_{L_1}$ displays a much higher variance than $\hat{N}^{\star}_{smv}$ and $\hat{N}_{\bar{w}}$.

Table 3.5: Net out-of-sample utility with a nonlinear shrinkage covariance matrix

		Sample covariance matrix						Nonlinear shrinkage covariance matrix					
Port-	Asset	Dataset						Dataset					
folio	select.	96	108	100	94	107	98IN-	96	108	100	94	107	98IN-
strat.	rule	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV
2F	All	−0.31	−59.6	−15.3	99.4	−18.8	56.8	4.64	−3E+31	1.05	44.4	2.19	16.4
2F	MaxSR	1.38	11.6●	−7.48	23.6○	−10.5	−27.1◑	1.86○	−3E+04●	−5.41	28.7◑	14.2●	10.5
2F	MinSR	25.4●	73.2●	7.10●	146	24.9●	139◑	19.1●	−2E+05●	5.54◑	104●	32.7●	68.5●
2F	BWSR	7.39	84.4●	−4.42	133	11.7○	242●	23.7●	−6E+05●	7.02	150●	39.1●	102●
2F	MaxVar	25.0◑	124●	17.9●	114	21.7●	83.8	24.9●	−2E+05●	9.04◑	126●	41.6●	72.1●
2F	MinVar	2.47	12.5●	−13.1	219◑	−1.88	225●	4.93	−3E+04●	−3.89	150●	21.2●	80.9●
2F	BWVar	13.0	89.3●	−6.76	182○	0.14	180●	22.6●	−5E+05●	5.16	141●	30.1●	93.3●
2F	MaxPC	8.80	18.3●	3.67◑	169○	9.62●	237●	15.0●	−9E+04●	5.66	125●	26.7●	89.1●
2F	MinPC	27.4●	112●	−6.23	115	17.9◑	94.1	26.3●	−5E+05●	−2.38	118●	40.5●	50.3●
2F	Best θ_N^2	−1.22	14.7●	−8.77	182○	4.73	137◑	19.1●	−5E+05●	−0.27	143●	30.5●	80.6●
2F	MaxW	−127●	−305●	−179●	−581●	−129●	−366●	33.4●	−2E+05●	9.79	205●	42.3●	132●
3FGMV	All	0.47	−60.5	−12.5	99.1	−19.7	55.4	5.45	−2E+31	1.64	44.7	3.13	16.5
3FGMV	MaxSR	−4.06	10.8●	−1.93○	15.6◑	−6.02	−27.4◑	2.99	−3E+04●	3.98	30.0◑	24.3●	11.3
3FGMV	MinSR	36.0●	71.6●	12.9●	150	31.5●	141◑	25.7●	−2E+05●	8.83◑	106●	37.9●	71.0●
3FGMV	BWSR	10.9	82.4●	1.02	143	19.9◑	253●	30.3●	−6E+05●	16.5●	152●	39.5●	104●
3FGMV	MaxVar	25.1◑	126●	24.9●	117	21.7●	85.0	27.3●	−2E+05●	12.9◑	128●	34.2●	74.2●
3FGMV	MinVar	6.13	13.3●	−14.9	222◑	0.35○	219●	7.25	−4E+04●	1.63	150●	25.2●	80.4●
3FGMV	BWVar	27.4◑	89.0●	2.74	183○	11.2◑	190●	28.4●	−5E+05●	13.1◑	145●	31.9●	94.6●
3FGMV	MaxPC	12.2	16.7●	10.8●	176○	7.50◑	225●	18.6●	−1E+05●	13.6●	127●	29.1●	90.7●
3FGMV	MinPC	33.3●	113●	6.47◑	115	25.9●	91.3	28.8●	−6E+05●	9.59○	120●	36.4●	52.7●
3FGMV	Best θ_N^2	12.3	7.39●	−0.08	122	1.33○	165●	22.9●	−6E+05●	10.2◑	147●	27.7●	81.0●
3FGMV	MaxW	−136●	−300●	−147●	−650●	−110●	−325●	38.3●	−2E+05●	20.0●	210●	43.7●	132●
3FEW	All	1.83	−62.8	−10.8	102	−16.8	60.8	6.82	−3E+31	2.95	43.2	3.06	17.2
3FEW	MaxSR	3.68	12.6●	−0.61	23.6◑	−12.1	−20.4◑	3.60	−3E+04●	2.85	28.1◑	6.34	11.6
3FEW	MinSR	24.8●	70.8●	11.3●	145	24.8●	146◑	18.6●	−2E+05●	7.41	102●	25.9●	68.0●
3FEW	BWSR	8.43	84.4●	3.41	135	17.9◑	237●	24.8●	−5E+05●	11.4◑	147●	34.8●	101●
3FEW	MaxVar	26.0◑	122●	25.3●	124	16.0●	98.9	23.5●	−2E+05●	12.2◑	126●	31.1●	71.4●
3FEW	MinVar	4.93	12.3●	−5.87	220◑	0.40	230●	7.97	−3E+04●	4.06	150●	15.6●	82.2●
3FEW	BWVar	14.2	88.6●	−3.30	195○	−0.15	187●	21.9●	−5E+05●	9.71◑	141●	21.3●	93.6●
3FEW	MaxPC	9.42	18.1●	11.9●	162	5.87◑	247●	17.4●	−9E+04●	10.7◑	120●	20.7●	89.1●
3FEW	MinPC	26.0◑	114●	2.26	120	15.9◑	103	23.9●	−5E+05●	6.26	120●	31.4●	52.8●
3FEW	Best θ_N^2	12.5	4.58●	−4.45	136	−8.83	166●	19.1●	−6E+05●	6.75	148●	21.8●	78.8●
3FEW	MaxW	−124●	−305●	−157●	−650●	−139●	−361●	23.5●	−2E+05●	8.38○	211●	12.2◑	131●
EW	All	8.11	2.80	8.00	3.98	7.13	5.61	8.11	2.80	8.00	3.98	7.13	5.61
EW	MaxSR	8.77○	4.72●	8.49○	5.97●	7.40	6.25◑	8.77○	4.72●	8.49○	5.97●	7.40	6.25◑
EW	MinSR	7.39◑	0.83●	7.55○	2.75●	6.77	5.18	7.39◑	0.83●	7.55○	2.75●	6.77	5.18
EW	BWSR	7.94	2.42●	7.72	2.86●	7.35	5.28	7.94	2.42●	7.72	2.86●	7.35	5.28
EW	MaxVar	7.95	1.38●	7.97	3.52	6.68	5.31	7.95	1.38●	7.97	3.52	6.68	5.31
EW	MinVar	8.60	4.23●	8.18	4.36	7.44	5.86	8.60	4.23●	8.18	4.36	7.44	5.86
EW	BWVar	7.60●	2.34●	7.52◑	3.99	6.88	5.58	7.60●	2.34●	7.52◑	3.99	6.88	5.58
EW	MaxPC	8.32	2.97	8.83◑	2.19●	7.37	5.08◑	8.32	2.97	8.83◑	2.19●	7.37	5.08◑
EW	MinPC	7.87	2.48●	7.20●	5.00◑	6.97	6.22◑	7.87	2.48●	7.20●	5.00◑	6.97	6.22◑
EW	Best θ_N^2	7.88○	2.72	7.92	3.82	7.02	6.03◑	8.08○	2.79	8.16	3.89	6.79	5.55◑
EWRF		1.42	−4.35	1.26	−3.06	0.80	−0.29	1.42	−4.35	1.26	−3.06	0.80	−0.29
SGMV		1.76	−49.7	4.25	−0.34	−2.28	−4.67	10.1	2.38	6.52	3.88	7.66	6.04
GMVRF		1.81	−26.3	−0.98	−11.4	−12.3	−5.34	3.81	−3E+34	1.25	0.19	2.65	0.82
SMV		−4E+05	−3E+11	−1E+06	−6E+07	−8E+08	−8E+07	−994	−1E+44	−974	−2115	−2744	−747
SMV-L_1		15.8●	−108●	11.7●	−579●	−2.91●	0.88●	21.6●	−349●	7.03●	−243●	−32.2●	9.12●
SMV-TW		−35.3●	−302●	−16.8●	−105●	−25.9●	−143●	−28.0●	−4E+32●	−29.6●	−421●	−288●	−84.8●

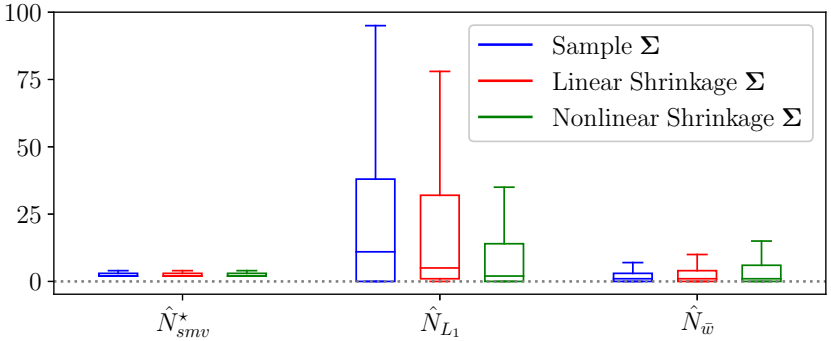
Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 10 portfolio strategies in Table 3.2, the 11 asset selection rules in Table 3.3, across the six datasets in Table 3.1. The table is constructed following the empirical methodology described in Section 3.5.4. We construct the portfolios either with the sample or the nonlinear shrinkage covariance matrix of Ledoit and Wolf (2020). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under the 'All' selection rule versus the 10 other selection rules, using the block bootstrap methodology described in Section 3.5.4. We also report p-values for SMV-TW and SMV-L_1 versus SMV. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

Table 3.6: Annualized net out-of-sample utility of the restricted SMV portfolio

		Sample covariance matrix						Linear shrinkage covariance matrix					
		Dataset						Dataset					
Port-folio strat.	Asset select. rule	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV
SMV	All	−4E+05	−3E+11	−1E+06	−6E+07	−8E+08	−8E+07	−1443	−107	−1399	−2360	−1457	−1143
SMV	MaxSR	−17.3●	−25.3●	−2011●	−38.5●	−27.2●	−21.5●	−15.8●	0.93	−34.6●	−37.6●	−25.7●	−19.8●
SMV	MinSR	−16.0●	−2.04●	−2007●	−43.4●	−12.8●	−17.9●	−14.5●	18.2	−31.7●	−13.2●	−13.1●	−17.8●
SMV	BWSR	−50.9●	18.5●	−2050●	−62.2●	−54.0●	−54.6●	−33.3●	95.1○	−63.4●	−8.63●	−44.0●	−34.3●
SMV	MaxVar	−5.94●	−57.3●	−1977●	−26.4●	−8.54●	−18.4●	−3.90●	47.1	−9.97●	−23.6●	−9.36●	−18.6●
SMV	MinVar	−22.1●	−35.1●	−1998●	−52.4●	0.22●	14.7●	−16.8●	4.38	−13.0●	30.3●	7.30●	25.7●
SMV	BWVar	0.73●	−83.2●	−1986●	−50.3●	−14.3●	−7.33●	2.36●	−1.50	−8.42●	−13.2●	−8.47●	−0.28●
SMV	MaxPC	−14.9●	26.8●	−2009●	−68.5●	−4.96●	−41.6●	−8.01●	30.8	−20.0●	11.6●	1.06●	−1.16●
SMV	MinPC	−19.3●	−49.7●	−2004●	−12.7●	−10.9●	−2.23●	−16.2●	−11.9	−30.8●	−11.1●	−10.4●	−2.07●
SMV	Best θ_N^2	−0.98●	34.2●	−1996●	22.1●	−44.8●	11.4●	9.97●	34.7	−30.6●	39.9●	−16.1●	3.14●
SMV	MaxW	−15.3●	−82.5●	−2021●	−76.2●	−16.3●	−78.4●	−17.5●	50.9	−45.4●	22.5●	−0.62●	−14.5●
SMV-L_1		15.8●	−108●	11.7●	−579●	−2.91●	0.88●	23.9●	67.8	5.13●	−195●	10.4●	16.4●
SMV-TW		−35.3●	−302●	−16.8●	−105●	−25.9●	−143●	−29.9●	−52.1	−17.8●	−376●	−33.6●	−349●

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the sample mean-variance portfolio across the six datasets in Table 3.1 and the 11 asset selection rules in Table 3.3, as well as two additional ways to implement the SMV portfolio described in Section 3.5.2: the norm constraint (L_1) and the threshold on weights (TW). The table is constructed following the empirical methodology described in Section 3.5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

Figure 3.14: Estimated portfolio sizes for the SMV portfolio



Notes. This figure depicts boxplots of the portfolio sizes obtained in empirical data for the SMV portfolio with three different strategies detailed in Section 3.5.2. First, our estimated portfolio size for the SMV portfolio, $\hat{N}^{\star}_{smv}$. Second, the implicit portfolio size obtained with L_1 norm regularization as given by (3.29), $\hat{N}_{L_1}$. Third, the portfolio size obtained using a cross-validated threshold on the weights as given by (3.30), $\hat{N}_{\bar{w}}$. We consider three different estimators of the covariance matrix, the sample covariance matrix given by (3.6) in blue and the linear and nonlinear shrinkage estimators of Ledoit and Wolf (2004, 2017), in red and green respectively. We average values obtained across the six datasets considered in the Empirical Analysis of Section 3.5.

3.F.6 Out-of-sample Sharpe ratio

Table 3.4 in the main body of the chapter reports the out-of-sample utility as a performance measure because it is the metric optimized by the portfolio combination rules in our theory. Another important mean-variance portfolio performance measure is the Sharpe ratio. Therefore, in Table 3.7, we report the annualized out-of-sample Sharpe ratio net of proportional transaction costs of the portfolio strategies considered in Table 3.4,

$$SR_k = \sqrt{12} \times \frac{\hat{\mu}_k}{\hat{\sigma}_k}, \tag{3.82}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the sample mean and standard deviation of the out-of-sample net portfolio returns, $r_{net,k,t}$ in (3.31). The main conclusions drawn from the out-of-sample utility are robust to using the out-of-sample Sharpe ratio. In particular, it is in most cases preferable to reduce the portfolio size in the 2F, 3FGMV, and 3FEW portfolio combination rules.

3.F.7 Portfolio turnover with asset selection rules

In the empirical analysis of Section 3.5, we consider different asset selection rules to select in which N assets to invest when reducing the portfolio size from M to $N \leq M$. The portfolio turnover is impacted by the changes to the investment universe over time implied by these selection rules. The larger the number of changes in selected assets in a given period, the lower the *stability* of the selection rule, and the higher the portfolio turnover.

The impact of a selection rule on portfolio turnover also depends the *selection frequency*, which is the frequency at which we decide to change the optimal N and the selected assets. In the empirical analysis of Section 3.5, we choose to rebalance the portfolio weights every month but to change the optimal N and the selected assets once a year. The rationale is to avoid excessive changes to the investment universe and make our method less costly and more practically implementable. However, we could choose a different selection frequency, and thus, it is important to test the robustness of our results to this choice.

Therefore, we have two objectives in this section. First, we investigate the stability of the different asset selection rules we consider and how they impact the portfolio turnover. Second, we evaluate the robustness of our results to using a selection frequency of once every month, six months, and 24 months instead of 12 months.

We begin by defining a measure of selection rule stability. To do so, note that each selection rule listed in Table 3.3, which is applied to the 2F, 3FGMV, 3FEW, and EW portfolio strategies, is implemented in a similar way following three steps. First, for each selection rule k, we compute at time t the $M \times 1$ vector of parameters on which the selection rule is based, which we denote as $\boldsymbol{\delta}_{k,t} = (\delta_{1,k,t}, \delta_{2,k,t}, \ldots, \delta_{M,k,t})$. For instance, if we consider a selection rule based on the in-sample Sharpe ratio (MaxSR, MinSR, or

Table 3.7: Annualized net out-of-sample Sharpe ratio (in percentage points)

		Sample covariance matrix						Linear shrinkage covariance matrix					
Port-	Asset	Dataset						Dataset					
folio	select.	96	108	100	94	107	98IN-	96	108	100	94	107	98IN-
strat.	rule	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV
2F	All	24.7	−150	−26.8	149	−6.46	113	89.4	214	40.1	244	72.6	209
2F	MaxSR	19.4	59.5●	−1.83	97.7◑	−0.85	44.7●	19.8●	162●	−2.01◑	81.5●	12.5●	55.4●
2F	MinSR	72.3●	121●	38.3●	171	70.6●	168◑	86.0	215	44.4	195●	77.5	180○
2F	BWSR	52.5◑	130●	33.4●	177	65.9●	220●	69.6○	215	32.1	223◑	67.1	200
2F	MaxVar	71.4●	163●	60.7●	157	72.7●	135	83.8	238●	50.7	197●	74.1	159●
2F	MinVar	32.0	65.4●	−13.6	211●	23.7○	212●	38.1●	179●	−2.61◑	225○	39.1◑	206
2F	BWVar	56.3◑	134●	26.7●	200◑	37.2◑	198●	75.8	217	33.0	214●	52.7	190○
2F	MaxPC	46.7	68.8●	33.2●	184○	46.9●	218●	71.8	194○	31.0	228○	61.0	236◑
2F	MinPC	74.2●	150●	12.3◑	167	71.9●	140	89.9	228○	1.89◑	176●	79.4	137●
2F	Best θ_N^2	39.7	75.4●	13.6◑	193○	66.3◑	169◑	81.7	208	9.13	228●	79.3	198◑
2F	MaxW	51.5●	3.14●	0.47◑	171	29.4●	177●	87.5	233●	28.7	238	63.2	210
3FGMV	All	26.9	−151	−14.2	149	−3.63	112	88.5	215	57.0	246	80.5	208
3FGMV	MaxSR	9.59	59.1●	8.19	94.7●	15.6	46.8●	18.2●	161●	27.7○	83.3●	35.7●	53.0●
3FGMV	MinSR	89.9●	120●	50.8●	173	79.3●	169◑	104	215	55.8	200●	91.4	183
3FGMV	BWSR	59.8●	128●	39.1●	181○	74.5●	225●	77.9	215	57.3	226○	80.1	198
3FGMV	MaxVar	71.1●	164●	70.7●	159	74.0●	136	89.8	238●	62.2	197●	76.8	156●
3FGMV	MinVar	41.4	66.2●	−18.3	213●	32.6◑	209●	42.5●	179●	5.01●	225○	48.6◑	202
3FGMV	BWVar	74.8●	133●	37.9●	201◑	53.3●	202●	87.8	216	59.0	218◑	65.2	186○
3FGMV	MaxPC	52.1○	67.4●	47.8●	188○	44.5●	212●	74.7	192◑	64.9	230	61.4	234○
3FGMV	MinPC	81.6●	151●	40.5●	167	80.9●	139	88.2	229	45.7	177●	88.3	136●
3FGMV	Best θ_N^2	53.7○	71.0●	29.6●	172	42.7●	182●	61.1◑	196◑	59.5	227○	73.7	162●
3FGMV	MaxW	47.3◑	3.59●	13.5◑	165	35.7●	182●	93.8	233●	56.7	242	80.8	208
3FEW	All	38.9	−117	12.1	151	15.9	116	40.2	5.52	32.7	135	31.8	66.6
3FEW	MaxSR	41.8	62.3●	33.9●	98.6●	27.0	51.2●	40.2	85.9●	36.9	67.3●	38.3○	49.6○
3FEW	MinSR	70.5●	119●	49.2●	170	70.9●	171◑	61.4●	175●	38.4	187◑	57.3◑	167●
3FEW	BWSR	58.4◑	130●	45.7●	178	73.7●	218●	64.4●	190●	44.5○	214●	66.8●	181●
3FEW	MaxVar	73.3●	161●	71.6●	162	69.8●	144	69.1●	212●	50.7●	191◑	60.5●	148●
3FEW	MinVar	45.6	66.5●	23.1	211●	44.2◑	215●	45.7	118●	34.7	217●	44.9●	172●
3FEW	BWVar	60.8○	133●	37.3◑	204◑	43.2○	201●	62.0●	191●	44.6○	210●	44.8○	175●
3FEW	MaxPC	53.8	69.3●	51.9●	180	47.9◑	223●	58.5●	135●	45.4◑	211●	48.2●	203●
3FEW	MinPC	75.3●	152●	38.2◑	170	72.4●	146	66.8●	198●	36.6	176○	65.7●	120●
3FEW	Best θ_N^2	56.9○	69.4●	29.6○	178	37.5○	182●	50.8○	157●	45.5◑	218●	51.7●	143●
3FEW	MaxW	50.6	2.91●	10.4	168	24.9	175●	70.7●	154●	39.8	237●	34.1	205●
EW	All	53.0	24.1	52.7	31.9	50.4	41.5	53.0	24.1	52.7	31.9	50.4	41.5
EW	MaxSR	57.6◑	35.0●	55.6○	44.6●	52.9	45.6◑	57.6◑	35.0●	55.6○	44.6●	52.9	45.6◑
EW	MinSR	48.1●	14.4●	49.7◑	24.5●	47.2○	38.5○	48.1●	14.4●	49.7◑	24.5●	47.2○	38.5○
EW	BWSR	52.0	22.2●	50.9○	25.1●	51.4	39.4○	52.0	22.2●	50.9○	25.1●	51.4	39.4○
EW	MaxVar	49.5◑	17.2●	49.9◑	28.3◑	45.0●	38.1◑	49.5◑	17.2●	49.9◑	28.3◑	45.0●	38.1◑
EW	MinVar	58.9◑	32.9●	56.7◑	35.4○	55.6◑	45.0○	58.9◑	32.9●	56.7◑	35.4○	55.6◑	45.0○
EW	BWVar	50.0●	21.8●	49.7●	32.1	49.0	41.4	50.0●	21.8●	49.7●	32.1	49.0	41.4
EW	MaxPC	52.8	24.9	56.0○	21.1●	50.9	37.8●	52.8	24.9	56.0○	21.1●	50.9	37.8●
EW	MinPC	52.4	22.5●	48.9◑	38.9●	50.0	45.5●	52.4	22.5●	48.9◑	38.9●	50.0	45.5●
EW	Best θ_N^2	51.4◑	23.7	52.2	31.1	49.7	44.1◑	53.9◑	23.8	51.6	31.7	49.2	43.5◑
EWRF		31.4	1.11	31.0	8.93	30.1	21.1	31.4	1.11	31.0	8.93	30.1	21.1
SGMV		19.3	−193	29.6	7.82	6.33	−6.63	82.0	31.1	68.0	42.4	77.5	59.6
GMVRF		23.8	−177	9.52	−27.6	−10.4	0.81	65.6	−1.26	53.0	5.35	50.9	14.0
SMV		−50.2	−29.7	−14.7	−91.5	−16.5	−76.9	91.2	240	40.0	243	72.1	206
SMV-L_1		77.5●	167●	64.0●	163●	78.9●	177●	77.0	175●	43.8	157●	65.5	147●
SMV-TW		36.0●	−97.3●	29.6◑	24.7●	20.7	46.4●	21.2●	103●	18.8	128●	15.6●	108●

Notes. This table reports the annualized net out-of-sample Sharpe ratio, in percentage points, for the 10 portfolio strategies in Table 3.2, the 11 asset selection rules in Table 3.3, across the six datasets in Table 3.1. The table is constructed following the empirical methodology described in Section 3.5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under the 'All' selection rule versus the 10 other selection rules, using the block bootstrap methodology described in Section 3.5.4. We also report p-values for SMV-TW and SMV-L_1 versus SMV. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

BWSR), then $\boldsymbol{\delta}_{k,t} = \hat{\boldsymbol{\theta}}_t$ where $\hat{\boldsymbol{\theta}}_t = (\hat{\theta}_{1,t}, \hat{\theta}_{2,t}, \ldots, \hat{\theta}_{M,t})$ is the $M \times 1$ vector of in-sample Sharpe ratios.

Second, we rank assets from 1 to M depending on their parameter values $\boldsymbol{\delta}_{k,t}$, where 1 represents the highest value and M the lowest. Specifically, we define the $M \times 1$ vector of ranks $\boldsymbol{\phi}_{k,t} = (\phi_{1,k,t}, \phi_{2,k,t}, \ldots, \phi_{M,k,t})$, where $\phi_{i,k,t}$ denotes the rank of asset i based on $\boldsymbol{\delta}_{k,t}$,

$$\phi_{i,k,t} = \sum_{j=1}^{M} \mathbb{1}_{\{\delta_{i,k,t} \leq \delta_{j,k,t}\}}. \tag{3.83}$$

Third, we select assets depending on their rank $\phi_{i,k,t}$ and the type of selection rule. For that purpose, we denote $y_{i,k,t}$ the binary variable indicating whether, for selection rule k, asset i is selected at time t: $y_{i,k,t} = 1$ if asset i is selected, and 0 otherwise. We consider three types of asset selection rules in Table 3.3. For the MaxSR, MaxVar, MaxPC, and MaxW rules that select the N assets with the highest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} \leq N\}}$. For the MinSR, MinVar and MinPC rules that select the N assets with the lowest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} > M-N\}}$. Finally, for the BWSR and BWVar rules that select the $\lceil N/2 \rceil$ assets with the highest $\delta_{i,k,t}$ values and the $\lfloor N/2 \rfloor$ assets with the lowest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} \leq \lceil N/2 \rceil \text{ or } \phi_{i,k,t} > M - \lfloor N/2 \rfloor\}}$.[38]

Given this three-step process with which the asset selection rules are implemented, we want to evaluate the impact on portfolio turnover of changes in the set of selected assets. We first introduce *selection turnover*, which measures, for a portfolio k, the number of changes in selected assets from month $t-1$ to t:

$$\text{selectionturnover}_{k,t} = \sum_{i=1}^{M} |y_{i,k,t} - y_{i,k,t-1}|, \quad t = T+1, \ldots, T_{tot}. \tag{3.84}$$

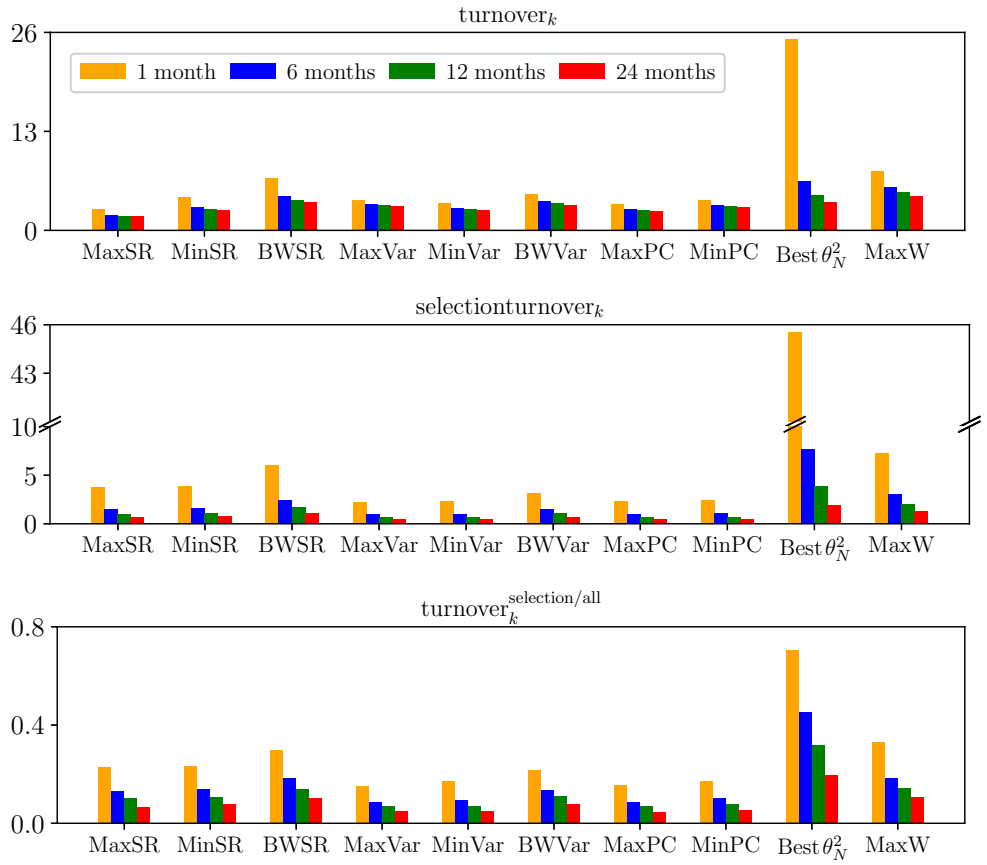
The less stable the asset selection rule and the higher the selection frequency, the higher the selection turnover, and the more we expect an increase in portfolio turnover due to changes in selected assets. To precisely measure the impact of selection turnover on the overall portfolio turnover, we define $\text{turnover}_{k,t}^{\text{selection}}$ as the turnover of portfolio k coming from those assets that were selected at time t but not at time $t-1$, and vice-versa:

$$\text{turnover}_{k,t}^{\text{selection}} = \sum_{i:\, y_{i,k,t} \neq y_{i,k,t-1}} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, T_{tot}, \tag{3.85}$$

where $w_{i,k,t}$ is the weight of asset i at month t and $w_{i,k,(t-1)^+}$ is the prior-time weight before rebalancing at month t. By construction, $\text{turnover}_{k,t}^{\text{selection}} \leq \text{turnover}_{k,t}$, where $\text{turnover}_{k,t}$ in (3.32) is the total turnover computed on all assets. Finally, we also define the proportion

[38] The Best θ_N^2 selection rule is an exception as it is not based on a ranking of assets $\boldsymbol{\phi}_{k,t}$. Instead, in that case, $y_{i,k,t} = 1$ if at time t, asset i is included in the randomly drawn selection of size N which corresponds to the 95% quantile of estimated values of θ_N^2, as detailed in Section 3.5.3. Otherwise, $y_{i,k,t} = 0$.

Figure 3.15: Portfolio turnover with asset selection rules and different selection frequencies



Notes. This figure depicts different turnover metrics for the two-fund portfolio rule across the 10 asset selection rules in Table 3.3 (we exclude the 'All' rule). The top panel depicts the average of the monthly portfolio turnover, defined in (3.32). The middle panel depicts the average of the monthly selection turnover, defined in (3.84), i.e., the average of the number of changes in selected assets each month. The bottom panel depicts the average of the proportion of the total turnover coming from changes in selected assets, defined in (3.85). These metrics are averaged across the six datasets in Table 3.1. We report them for four different selection frequencies: every 1, 6, 12, and 24 months. The selection frequency is the frequency with which we change the optimal portfolio size N and the selected assets. We construct the two-fund rule with the linear shrinkage covariance matrix of Ledoit and Wolf (2004) and a risk-aversion coefficient of $\gamma = 1$. The figure is constructed following the out-of-sample methodology described in Section 3.5.4. The middle panel includes '//' markers indicating an axis break for better visibility.

of the total turnover that originates from changes in the selected assets:[39]

$$\text{turnover}_{k,t}^{\text{selection/all}} = \frac{\text{turnover}_{k,t}^{\text{selection}}}{\text{turnover}_{k,t}} \leq 1, \tag{3.86}$$

which we expect to increase with $\text{selectionturnover}_{k,t}$.

In Figure 3.15, we depict the portfolio turnover in (3.32), the selection turnover in (3.84), and the proportion of portfolio turnover explained by changes in the selected assets in (3.86). We average these measures across all months t and across the six datasets described in Table 3.1. We report results for each of the four considered selection frequencies, i.e., every 1, 6, 12, and 24 months. We only consider the 2F portfolio combination rule for conciseness; the results for the 3FGMV and 3FEW portfolios are similar. Moreover, we

[39]The measures $\text{selectionturnover}_{k,t}$ and $\text{turnover}_{k,t}^{\text{selection}}$ are non-zero only when the investment universe changes from $t-1$ to t, which happens according to the chosen selection frequency (1, 6, 12, or 24 months).

only report the results for the case in which we use the linear shrinkage covariance matrix of Ledoit and Wolf (2004) because, as discussed in Section 3.5.5, it leads to better out-of-sample performance, and also lower portfolio turnover. We make several observations from Figure 3.15.

First, we look at the stability of the different asset selection rules. The middle panel of Figure 3.15 shows that the Best θ_N^2 selection rule has the largest selection turnover, especially for a monthly selection frequency. This is expected as this rule is based on random selections. Then, the ex-post selection rule MaxW has the second largest selection turnover because it ranks assets according to portfolio weights and these weights can vary substantially from one period to another. Then, selection rules that rank assets based on the marginal Sharpe ratio are also quite unstable because they depend on estimates of expected returns that are notoriously noisy.[40] In contrast, selection rules built on the variance and PCA are remarkably stable, with the set of selected assets changing by only one or two assets on average each month when we use a monthly selection frequency. Finally, we also observe that the best-worst asset selection rules, BWSR and BWVar, are more unstable because they select the assets with the most extreme rankings, which are also more subject to more estimation errors.

Second, the bottom panel of Figure 3.15 shows that a higher selection turnover is indeed linked with a higher proportion of portfolio turnover originating from assets that are selected in the portfolio at time t but not at time $t-1$, and vice-versa.

Third, the bottom panel of Figure 3.15 shows that, except for Best θ_N^2, the proportion of the total portfolio turnover explained by assets that enter and exit the set of selected assets each month is around 10% to 30% on average when we use a monthly selection frequency but, as expected, decreases substantially as we decrease the selection frequency. This proportion of 10% to 30% with a monthly selection frequency is not negligible, particularly given that the selection turnover in the middle panel of Figure 3.15 is generally well below 10% of the total number of selected assets. Still, it means that the majority of the portfolio turnover comes from changes in weights allocated to assets that remain selected from one period to another, particularly as we decrease the selection frequency.

We now turn to our second objective, which is to evaluate the robustness of our out-of-sample performance results to the choice of the selection frequency. Figure 3.15 confirms that decreasing the selection frequency, and thus changing the investment universe less frequently, significantly reduces selection turnover and thus the overall portfolio turnover. However, this does not automatically mean that decreasing the selection frequency improves the net out-of-sample performance because the monthly frequency could deliver better

[40]The number of changes in selected assets can be substantial. For instance, on average across the six datasets, the selection turnover of BWSR in Figure 3.15 is equal to 5.2 when using a monthly selection frequency. This means that, on average, 5.2 assets enter or exit the set of selected assets from one month to another. Given that the number of selected assets N dictated by our theory is close to $T/2 = 60$, this represents a 8.67% change of the investment universe each month on average.

gross performance. To test the impact of the selection frequency on the net out-of-sample portfolio performance, we report in Table 3.8 the annualized net out-of-sample utility of the 2F portfolio implemented with the 11 asset selection rules in Table 3.3 and a selection frequency of 1, 6, 12, and 24 months. The results for the 3FGMV and 3FEW portfolio combination rules are similar. We make three main observations from Table 3.8.

First, in the vast majority of cases, using a lower selection frequency than monthly is beneficial, and the performance gain can be substantial. For instance, consider the MinSR rule for the 96S-BM dataset and the sample covariance matrix. In this case, a monthly selection frequency delivers an annualized net utility of 19.3 whereas the 6, 12, and 24-month frequencies deliver a net utility of 25.5, 25.4 and 25.4, respectively. A similar observation holds for the linear shrinkage covariance matrix, but with lower differences in performance. This is because portfolios implemented with the linear shrinkage covariance matrix yield a lower turnover than those under the sample covariance matrix, and thus, benefit relatively less from an extra reduction in turnover.

Second, the 6, 12, and 24-month selection frequencies deliver a similar net out-of-sample performance overall, which means that our use of a 12-month frequency in the main body of the chapter is robust to other possible choices.

Third, asset selection rules that outperform investing in all assets do so under all considered selection frequencies, including monthly. This shows that our optimal diversification method is valuable regardless of the frequency at which the selection rules are implemented.

3.F.8 Overlap between selection rules

We show in Section 3.5 of the main text that reducing the portfolio size from M to N is beneficial in most cases by implementing different selection rules listed in Table 3.3. These selection rules are different by construction (marginal information vs. information on dependence, ex-ante vs. ex-post approach) to test whether our results hold for different types of selection rule. Thus, we raise the following question: do these rules indeed select dissimilar assets in practice?

To answer this question, we must measure the *overlap* between two selection rules, i.e., the number of assets jointly selected by both selection rules. Following Section 3.F.7, we denote by $\boldsymbol{y}_{k,t} = (y_{1,k,t}, y_{2,k,t}, \ldots, y_{M,k,t})$ the $(M \times 1)$ selection vector for any selection rule k: $y_{i,k,t} = 1$ if asset i is selected in rule k at time t, and 0 otherwise. The overlap between rules k and l is then given by $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \sum_{i=1}^{M} y_{i,k,t}y_{i,l,t}$.

We then compare the overlap between the selection rules listed in Table 3.3 to the overlap between two independent random selection rules. If the average overlap between our selection rules is close to or lower than that of a pair of two independent random selections, then it means that the selected assets are indeed dissimilar.

Table 3.8: Annualized net out-of-sample utility in empirical data with different selection frequencies

		Sample covariance matrix						Linear shrinkage covariance matrix					
Asset		Dataset						Dataset					
select.	Selection	96	108	100	94	107	98IN-	96	108	100	94	107	98IN-
rule	Frequency	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV
All	n.a.	−0.31	−59.6	−15.3	99.4	−18.8	56.8	5.11	1.20	1.09	41.6	1.09	17.7
MaxSR	1 month	−5.80	−30.0○	−7.72	2.25◑	−21.7	−35.2●	−1.98●	19.8●	−1.94	21.4●	0.66	7.66◑
MaxSR	6 months	0.77	3.62●	−5.15	31.7	−9.86	−31.1●	0.66●	22.3●	−2.71	26.5◑	3.72	10.3○
MaxSR	12 months	1.38	11.6●	−7.48	23.6○	−10.5	−27.1◑	1.93○	23.1●	−6.10	23.1●	0.69	10.8○
MaxSR	24 months	2.02	16.2●	−4.14○	34.8○	−13.3	−18.8◑	2.56○	22.8●	−1.54	22.4●	−0.05	8.59◑
MinSR	1 month	19.3◑	25.9●	6.84●	123	19.2●	112○	16.0●	50.7●	5.21○	97.9●	12.0●	67.3●
MinSR	6 months	25.5●	70.1●	8.49●	145	26.8●	130◑	17.4●	53.4●	6.54●	98.3●	14.8●	70.3●
MinSR	12 months	25.4●	73.2●	7.10●	146	24.9●	139◑	18.0●	53.2●	5.21○	101●	16.0●	72.1●
MinSR	24 months	25.4●	85.3●	9.09●	167	19.3●	150●	19.6●	53.9●	6.28◑	107●	15.2●	74.7●
BWSR	1 month	−2.88	14.3●	−14.5	161	−8.83	168●	16.5○	61.8●	3.02	140●	18.8●	88.8●
BWSR	6 months	0.55	66.1●	−2.22	154	−4.83	232●	18.1◑	63.1●	8.21	143●	20.1●	95.8●
BWSR	12 months	7.39	84.4●	−4.42	133	11.7○	242●	19.2◑	63.8●	4.74	143●	18.8◑	99.5●
BWSR	24 months	16.0	80.7●	−0.23	96.1	10.8○	219●	21.7●	64.0●	8.18	133●	13.0	103●
MaxVar	1 month	27.7●	112●	10.0●	112	20.9●	76.9	26.4●	69.1●	9.46◑	127●	20.5●	77.6●
MaxVar	6 months	26.5◑	127●	17.0●	118	21.9●	73.3	24.1●	70.0●	10.6◑	125●	22.6●	78.0●
MaxVar	12 months	25.0◑	124●	17.9●	114	21.7●	83.8	23.2●	70.5●	9.99◑	122●	21.1●	77.7●
MaxVar	24 months	29.8●	129●	18.6●	113	19.5●	91.9	25.1●	69.8●	12.0●	120●	18.0●	78.5●
MinVar	1 month	1.97	−18.5◑	−14.5	215◑	−0.88○	199●	5.17	32.6●	−2.38	136●	7.27◑	80.1●
MinVar	6 months	5.32	7.83●	−10.5	213◑	0.57○	222●	6.30	32.7●	−1.07	141●	6.66◑	78.8●
MinVar	12 months	2.47	12.5●	−13.1	219◑	−1.88	225●	6.08	33.1●	−5.25○	140●	5.40○	81.4●
MinVar	24 months	2.04	20.0●	−12.6	227●	−3.33	226●	5.68	33.8●	−4.89○	140●	4.11	81.8●
BWVar	1 month	19.4◑	75.0●	−0.95○	189◑	−4.76	151◑	21.2●	62.9●	9.34○	127●	11.1●	85.7●
BWVar	6 months	17.5○	90.4●	3.84◑	206◑	−0.03	168●	21.2●	63.6●	7.90	132●	12.3●	88.9●
BWVar	12 months	13.0	89.3●	−6.76	182○	0.14	180●	19.8●	63.4●	5.39	133●	10.6◑	89.9●
BWVar	24 months	20.8◑	101●	1.00○	190○	−0.89	190●	21.7●	62.8●	7.48	133●	9.12○	89.0●
MaxPC	1 month	16.7◑	8.82●	7.52●	170○	0.52○	222●	16.4●	35.5●	9.04●	119●	9.72●	91.0●
MaxPC	6 months	8.48	24.5●	6.84●	173○	7.58◑	229●	14.8●	37.0●	6.74○	119●	10.7●	90.1●
MaxPC	12 months	8.80	18.3●	3.67◑	169○	9.62●	237●	14.7●	36.9●	4.67	119●	9.81●	91.7●
MaxPC	24 months	14.6○	14.8●	9.20●	168○	9.07◑	238●	17.6●	36.8●	7.01○	120●	9.75●	94.9●
MinPC	1 month	25.7●	92.8●	−8.43	86.8	17.0●	88.9	24.2●	66.2●	−1.14	108●	22.1●	49.2●
MinPC	6 months	30.9●	105●	−7.68	89.2	27.8●	98.3	25.4●	67.8●	−2.59	105●	22.9●	52.6●
MinPC	12 months	27.4●	112●	−6.23	115	17.9◑	94.1	24.9●	67.7●	−2.55	109●	22.8●	50.5●
MinPC	24 months	20.8○	106●	−4.26	101	22.3●	106○	22.8●	67.9●	2.32	107●	24.4●	50.7●
Best θ_N^2	1 month	−40.2●	−541●	−50.8●	−99.4●	−73.1●	−104●	4.40	29.1●	−9.28●	79.2●	−3.92	33.3◑
Best θ_N^2	6 months	4.06	−47.3	−3.26	159	−1.21	102	19.6●	52.5●	5.82	120●	16.6●	72.9●
Best θ_N^2	12 months	3.77	14.7●	−8.77	182○	−7.87	137◑	20.7●	55.1●	0.49	119●	17.6●	73.2●
Best θ_N^2	24 months	14.6	50.1●	−2.44	206◑	−2.04	143◑	21.7●	56.0●	4.13	134●	20.7●	87.4●
MaxW	1 month	−307●	−798●	−350●	−740●	−300●	−543●	30.3●	79.2●	9.19	199●	17.9○	130●
MaxW	6 months	−171●	−421●	−234●	−749●	−151●	−456●	29.2●	79.4●	3.28	203●	19.4○	131●
MaxW	12 months	−127●	−305●	−179●	−581●	−129●	−366●	32.3●	79.0●	2.19	197●	19.8○	129●
MaxW	24 months	−97.6●	−74.8	−115●	−416●	−56.3○	−244●	32.2●	77.4●	3.59	195●	16.0	126●

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 2F portfolio implemented with the 11 asset selection rules in Table 3.3, across the six datasets in Table 3.1. For each selection rule, we report results for four different selection frequencies: every 1, 6, 12, and 24 months. The selection frequency is the frequency with which we change the optimal portfolio size N and the selected assets. The table is constructed following the empirical methodology described in Section 3.5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004) and rebalance the portfolio weights every month. The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F portfolio under the 'All' selection rule versus the 10 other selection rules, for all selection frequencies, using the block bootstrap methodology described in Section 3.5.4. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

Consider two selection rules k and l that randomly select N_t out of M assets at time t. In the next proposition, we derive the expected overlap if k and l are independent.

Proposition 3.10. *Consider two independent selection rules that randomly select N_t assets out of M available ones. Then, the expected overlap between the two rules is*

$$\bar{N}_{t,M} = M \frac{\binom{M-1}{N_t-1}^2}{\binom{M}{N_t}^2}. \tag{3.87}$$

For instance, if two independent selection rules randomly select $N_t = 50$ assets out of $M = 100$ available ones, we expect them to jointly select $\bar{N}_{t,M} = 25$ assets.

Recall that our objective is to identify how a pair of selection rules (k, l) selects assets relative to two independent random selection rules. Therefore, we use the expected overlap in (3.87) to build a measure of the *excess overlap* of any (k, l) pair relative to that of a pair of independent random selection rules. We denote this measure $\text{eo}_{k,l,t}$ and define it as

$$\text{eo}_{k,l,t} = (\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} - \bar{N}_{t,M}) \times \left(\frac{\mathbb{1}_{\{\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} - \bar{N}_{t,M} \geq 0\}}}{(N_t - \bar{N}_{t,M})} - \frac{\mathbb{1}_{\{\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} - \bar{N}_{t,M} < 0\}}}{(\max(0, 2N_t - M) - \bar{N}_{t,M})} \right). \tag{3.88}$$

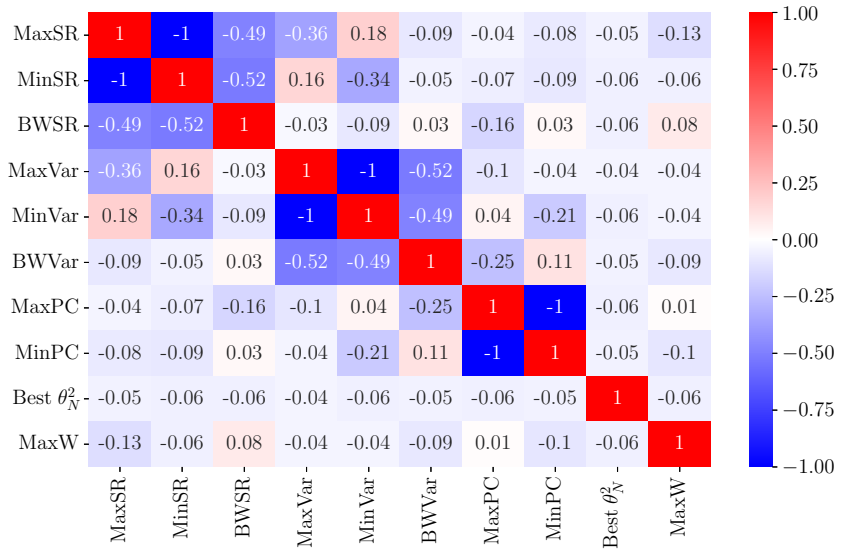
Several comments are in order. First, if $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \bar{N}_{t,M}$, then $\text{eo}_{k,l,t} = 0$. This means that there is no excess overlap if the overlap between k and l is equal to the expected overlap between two independent random selection rules. Second, the excess overlap is positive (resp. negative) if there is more (resp. less) overlap between k and l compared to independent random rules, i.e., $\text{eo}_{k,l,t} > 0$ if $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} > \bar{N}_{t,M}$ and vice-versa. Third, the measure of excess overlap is bounded between -1 and 1. A value of 1 indicates the *maximum* overlap and is obtained if and only if k and l are *identical*, i.e., $\boldsymbol{y}_{k,t} = \boldsymbol{y}_{l,t}$ such that $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = N_t$. A value of -1 indicates the *minimum* overlap, i.e., $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \max(0, 2N_t - M)$.[41]

In Figure 3.16, we depict a heatmap of the excess overlap in (3.88) for all selection rules pairs (k, l) in Table 3.3. We average it across all months t and across the six datasets described in Table 3.1. We only consider the optimal portfolio size of the two-fund rule for conciseness, but the results are similar for the three-fund rules. Moreover, because selection rules are based on sample estimates, we report results under the sample covariance matrix.

Two main observations can be made from Figure 3.16. First, the results match our expectations. As expected, opposite rules (MaxSR and MinSR, MaxVar and MinVar, MaxPC and MinPC) have an excess overlap of -1. Also, the sign of the excess overlap is coherent. For instance, the average excess overlap is positive between MaxVar and MinSR (0.16) because both rules tend to select assets with high marginal variance, while it is negative between MaxVar and MaxSR (-0.36) because they tend to select different assets.

[41] If $M \geq 2N_t$, then it is possible for two rules to have zero overlap as each rule can select N_t different assets. However, if $M < 2N_t$, then the minimum possible overlap will be strictly positive and will be of $2N_t - M$ assets. For instance, if $M = 90$ and $N_t = 50$, two rules must have an overlap of at least $2N_t - M = 10$ assets.

Figure 3.16: Average excess overlap between selection rules



Notes. This figure depicts the average excess overlap, defined in (3.88), for each pair of selection rule listed in Table 3.3 (we exclude the 'All' rule). The excess overlap is averaged over the six datasets in Table 3.1 and all estimation windows. We report the results for the optimal portfolio size of the two-fund rule and the sample covariance matrix.

Second, we observe that for essentially all pairs of selection rules we consider, there is either around the same or less overlap than in the random selection case. This answers our initial question, as it shows that the selection rules in Table 3.3 do select assets differently.

3.G Proofs

Proof of Proposition 3.1

Denoting $\boldsymbol{D}_N = \text{diag}(\sigma_1, \ldots, \sigma_N)$, $\boldsymbol{\Sigma}_N = \boldsymbol{D}_N \boldsymbol{P}_N(\rho) \boldsymbol{D}_N$ and its inverse is given by

$$\boldsymbol{\Sigma}_N^{-1} = \boldsymbol{D}_N^{-1} \boldsymbol{P}_N(\rho)^{-1} \boldsymbol{D}_N^{-1}, \tag{3.89}$$

where $\boldsymbol{D}_N^{-1} = \text{diag}(1/\sigma_1, \ldots, 1/\sigma_N)$ and $\boldsymbol{P}_N(\rho)^{-1}$ exists if and only if $-1/(N-1) < \rho < 1$ and is equal to

$$\boldsymbol{P}_N(\rho)^{-1} = \frac{1}{1-\rho} \left(\boldsymbol{I}_N - \frac{\rho}{1-\rho+N\rho} \mathbf{1}_N \mathbf{1}_N' \right). \tag{3.90}$$

Combining (3.89) and (3.90) yields

$$\boldsymbol{\Sigma}_N^{-1} = \frac{1}{1-\rho} \left(\boldsymbol{D}_N^{-2} - \frac{\rho}{1-\rho+N\rho} \boldsymbol{D}_N^{-1} \mathbf{1}_N \mathbf{1}_N' \boldsymbol{D}_N^{-1} \right), \tag{3.91}$$

and thus the maximum utility becomes

$$U(\boldsymbol{w}^\star) = \frac{\boldsymbol{\mu}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N}{2\gamma} = \frac{1}{2\gamma(1-\rho)} \left[\boldsymbol{\mu}_N' \boldsymbol{D}_N^{-2} \boldsymbol{\mu}_N - \frac{\rho}{1-\rho+N\rho} (\boldsymbol{\mu}_N' \boldsymbol{D}_N^{-1} \mathbf{1}_N)^2 \right], \tag{3.92}$$

where $\boldsymbol{\mu}_N' \boldsymbol{D}_N^{-2} \boldsymbol{\mu}_N = N\bar{\theta}_{N,2}$ and $\boldsymbol{\mu}_N' \boldsymbol{D}_N^{-1} \mathbf{1}_N = N\bar{\theta}_{N,1}$, which yields the desired result in (3.4).

We then study how $U(\boldsymbol{w}^\star)$ increases with N, which amounts to studying the difference $\theta_{N+1}^2 - \theta_N^2$. We have

$$\begin{aligned}
&(1-\rho)(\theta_{N+1}^2 - \theta_N^2) \\
&= \left[\sum_{i=1}^{N+1} s_i^2 - \frac{\rho}{1-\rho+(N+1)\rho}\left(\sum_{i=1}^{N+1} s_i\right)^2\right] - \left[\sum_{i=1}^{N} s_i^2 - \frac{\rho}{1-\rho+N\rho}\left(\sum_{i=1}^{N} s_i\right)^2\right] \quad (3.93)
\end{aligned}$$

Decomposing $\left(\sum_{i=1}^{N+1} s_i\right)^2$ as $\left(\sum_{i=1}^{N} s_i\right)^2 + s_{N+1}^2 + 2s_{N+1}\sum_{i=1}^{N} s_i$, (3.93) becomes

$$\begin{aligned}
(1-\rho)(\theta_{N+1}^2 - \theta_N^2) &= \frac{1-\rho+N\rho}{1-\rho+(N+1)\rho} s_{N+1}^2 + \frac{\rho^2\left(\sum_{i=1}^{N} s_i\right)^2}{(1-\rho+N\rho)(1-\rho+(N+1)\rho)} \\
&\quad - \frac{2\rho s_{N+1}\sum_{i=1}^{N} s_i}{1-\rho+(N+1)\rho} \\
&= \frac{\left[\rho N\bar{\theta}_{N,1} - (1-\rho+N\rho)s_{N+1}\right]^2}{(1-\rho+N\rho)(1-\rho+(N+1)\rho)}, \quad (3.94)
\end{aligned}$$

which is nonnegative, and strictly positive if and only if $s_{N+1} \neq \rho N\bar{\theta}_{N,1}/(1-\rho+N\rho)$. This concludes the proof.

Proof of Proposition 3.2

Equation (3.11) is a direct extension of Kan and Lassance (2025, Proposition 7) for a general α instead of $\alpha = 1$. It is then easy to show that the α maximizing (3.11) is equal to (3.16), which also corresponds to Kan and Lassance (2025, Equation (50)). Finally, after some developments, plugging (3.16) into (3.11) yields Equation (3.17), which concludes the proof.

Proof of Corollary 3.1

Under Assumption 1, the maximum squared Sharpe ratio θ_N^2 is given by (3.4). Plugging it into (3.11) yields the EU of the SMV portfolio $\hat{\boldsymbol{w}}^\star$ in Equation (3.18), and plugging it into (3.17) yields the EU of the optimal two-fund rule $\hat{\boldsymbol{w}}(\alpha^\star)$ in (3.20).

Proof of Proposition 3.3

Under Assumption 3, the covariance matrix of $\boldsymbol{r}$ is

$$\boldsymbol{\Sigma}_N = \sigma_f^2 \boldsymbol{\beta}_N \boldsymbol{\beta}_N' + \sigma_\varepsilon^2 \boldsymbol{I}_N, \quad (3.95)$$

and its inverse is

$$\boldsymbol{\Sigma}_N^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[\boldsymbol{I}_N - \frac{\sigma_f^2}{\sigma_\varepsilon^2 + \sigma_f^2 \boldsymbol{\beta}_N' \boldsymbol{\beta}_N} \boldsymbol{\beta}_N \boldsymbol{\beta}_N'\right]. \quad (3.96)$$

We have that θ_N^2 becomes

$$\theta_N^2 = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \frac{N}{\sigma_\varepsilon^2}\left[\bar{\ell}_N^{\mu} - \frac{N\sigma_f^2}{\sigma_\varepsilon^2 + N\sigma_f^2\bar{\ell}_N^{\beta}}(\bar{\ell}_N^{\mu,\beta})^2\right], \tag{3.97}$$

which corresponds to (3.36). Plugging (3.97) in (3.11) and (3.17) yields the desired result and thus completes the proof.

Proof of Proposition 3.4

We begin by proving that $U(\boldsymbol{w}^\star)$ in (3.39) is a convex function by showing that its second derivative with respect to ρ is positive for all $\rho \in (-\frac{1}{N-1}, 1)$. The first derivative is

$$\frac{\partial}{\partial\rho}U(\boldsymbol{w}^\star) = \frac{N}{2\gamma}\left[\frac{\bar{\theta}_{N,2}(1-\rho+N\rho)^2 - N\bar{\theta}_{N,1}^2(1-\rho^2+N\rho^2)}{(1-\rho)^2(1-\rho+N\rho)^2}\right], \tag{3.98}$$

and the second derivative is

$$\frac{\partial^2}{\partial\rho^2}U(\boldsymbol{w}^\star) = \frac{N}{\gamma(1-\rho)^3}\left[(\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2) + \frac{(N-1)^2(1-\rho)^3}{(1-\rho+N\rho)^3}\bar{\theta}_{N,1}^2\right], \tag{3.99}$$

which is positive because $\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2 \geq 0$ and $(1-\rho+N\rho)^3 \geq 0$ for all $\rho \in (-\frac{1}{N-1}, 1)$.

Next, we prove that the value of $\rho \in (-\frac{1}{N-1}, 1)$ minimizing $U(\boldsymbol{w}^\star)$ in (3.39) is given by $\rho^{\min}$ in (3.40). There are two roots to the first derivative in (3.98) given by

$$\rho^{\pm} = \frac{-\bar{\theta}_{N,2} \pm \frac{N}{\sqrt{N-1}}|\bar{\theta}_{N,1}|\sqrt{\delta}}{N\delta - \bar{\theta}_{N,2}}, \tag{3.100}$$

where we define $\delta = \bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2 \geq 0$. We proceed by showing that $\rho^- \notin (-\frac{1}{N-1}, 1)$ whereas $\rho^+ = \rho^{\min} \in (-\frac{1}{N-1}, 1)$, and thus, is a minimizer of $U(\boldsymbol{w}^\star)$ that is convex in that interval.

We first treat the case $\rho^- \notin (-\frac{1}{N-1}, 1)$. We have $\lim_{N\to\bar{\theta}_{N,2}/\delta}\rho^- = \infty$ and thus we can leave out the case $N = \bar{\theta}_{N,2}/\delta$. Let us consider the case $N > \bar{\theta}_{N,2}/\delta$. Then, ρ^- is negative and we need to show that $\rho^- < -\frac{1}{N-1}$. This is equivalent to

$$\frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N\sqrt{\delta}}{N-1}\left(\sqrt{N-1}|\bar{\theta}_{N,1}| - \sqrt{\delta}\right) > 0, \tag{3.101}$$

which holds because, given $N > \bar{\theta}_{N,2}/\delta$, we have

$$\begin{aligned}
&\frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N\sqrt{\delta}}{N-1}\left(\sqrt{N-1}|\bar{\theta}_{N,1}| - \sqrt{\delta}\right) \\
&> \frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N\sqrt{\delta}}{N-1}\left(\sqrt{\frac{\bar{\theta}_{N,2}}{\delta} - 1}|\bar{\theta}_{N,1}| - \sqrt{\delta}\right) \\
&= \frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N(2\bar{\theta}_{N,1}^2 - \bar{\theta}_{N,2})}{N-1} > 0.
\end{aligned} \tag{3.102}$$

Then, we consider the case $N < \bar{\theta}_{N,2}/\delta$. In this case, ρ^- is positive and we need to show that $\rho^- > 1$. This is equivalent to

$$N\delta + \frac{N|\bar{\theta}_{N,1}|\sqrt{\delta}}{\sqrt{N-1}} > 0, \tag{3.103}$$

which holds true.

We treat next the case $\rho^+ \in (-\frac{1}{N-1}, 1)$. When $N = \bar{\theta}_{N,2}/\delta$, $\rho^+ = 0/0$ is indeterminate and we can show using L'Hospital rule that

$$\lim_{N \to \bar{\theta}_{N,2}/\delta} = 1 - \frac{\bar{\theta}_{N,2}}{2\bar{\theta}^2_{N,1}}, \tag{3.104}$$

which is strictly smaller than one and, moreover, $1 - \bar{\theta}_{N,2}/(2\bar{\theta}^2_{N,1}) > 1 - \bar{\theta}_{N,2}/\bar{\theta}^2_{N,1} = -1/(N-1)$ when $N = \bar{\theta}_{N,2}/\delta$. We then consider the case $N > \bar{\theta}_{N,2}/\delta$. In this case, $\rho^+ < 1$ is equivalent to

$$\frac{N\sqrt{\delta}}{\sqrt{N-1}} \left(\sqrt{(N-1)\delta} - |\bar{\theta}_{N,1}| \right) > 0, \tag{3.105}$$

which holds if $\delta > \bar{\theta}^2_{N,1}/(N-1)$ and this holds true because, given $N > \bar{\theta}_{N,2}/\delta$,

$$\frac{\bar{\theta}^2_{N,1}}{N-1} < \frac{\bar{\theta}^2_{N,1}}{\bar{\theta}_{N,2}/\delta - 1} = \delta. \tag{3.106}$$

Next, still in the case $N > \bar{\theta}_{N,2}/\delta$, $\rho^+ > -\frac{1}{N-1}$ is equivalent to

$$\frac{N}{N-1} \left(\delta + |\bar{\theta}_{N-1}| \sqrt{(N-1)\delta} - \bar{\theta}_{N,2} \right) > 0, \tag{3.107}$$

which holds true because

$$\delta + |\bar{\theta}_{N-1}| \sqrt{(N-1)\delta} - \bar{\theta}_{N,2} > \delta + |\bar{\theta}_{N-1}| \sqrt{\left(\frac{\bar{\theta}_{N,2}}{\delta} - 1 \right) \delta} - \bar{\theta}_{N,2} = 0. \tag{3.108}$$

We turn then to the case in which $N < \bar{\theta}_{N,2}/\delta$. In this case, $\rho^+ < 1$ is equivalent to

$$\frac{N\sqrt{\delta}}{\sqrt{N-1}} \left(|\bar{\theta}_{N,1}| - \sqrt{(N-1)\delta} \right) > 0, \tag{3.109}$$

which holds if $\delta < \bar{\theta}^2_{N,1}/(N-1)$ and this holds true because, given $N < \bar{\theta}_{N,2}/\delta$, inequality (3.106) is reversed. Finally, still in the case $N < \bar{\theta}_{N,2}/\delta$, $\rho^+ > -\frac{1}{N-1}$ is equivalent to

$$\frac{N}{N-1} \left(\delta + |\bar{\theta}_{N-1}| \sqrt{(N-1)\delta} - \bar{\theta}_{N,2} \right) < 0, \tag{3.110}$$

which holds true because, given $N < \bar{\theta}_{N,2}/\delta$, (3.108) is reversed. This concludes the proof.

Proof of Proposition 3.5

Let $\tilde{\mu}_1 = \mathbb{E}[\hat{\boldsymbol{\mu}}_N' \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\mu}_N]$, $\tilde{\mu}_2 = \mathbb{E}[\boldsymbol{\mu}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_N]$, $\tilde{\sigma}_1^2 = \mathbb{E}[\hat{\boldsymbol{\mu}}_N' \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\Sigma}_N \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N]$, $\tilde{\sigma}_2^2 = \mathbb{E}[\mathbf{1}_N' \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\Sigma}_N \hat{\boldsymbol{\Sigma}}_N^{-1} \mathbf{1}_N]$, and $\tilde{\sigma}_{12} = \mathbb{E}[\hat{\boldsymbol{\mu}}_N' \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\Sigma}_N \hat{\boldsymbol{\Sigma}}_N^{-1} \mathbf{1}_N]$. Then, the EU of the GMV-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\alpha})$ in (3.43) is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha})) = \frac{1}{2\gamma} \left(2\alpha_1 \tilde{\mu}_1 + 2\alpha_2 \tilde{\mu}_2 - \alpha_1^2 \tilde{\sigma}_1^2 - \alpha_2^2 \tilde{\sigma}_2^2 - 2\alpha_1 \alpha_2 \tilde{\sigma}_{12} \right). \tag{3.111}$$

Kan and Lassance (2025, Equations (IA80)–(IA84)) show that

$$\tilde{\mu}_1 = \frac{k_{N,1} \theta_N^2 T}{T - N - 2}, \tag{3.112}$$

$$\tilde{\mu}_2 = \frac{k_{N,1} \lambda_{g,N} T}{T - N - 2}, \tag{3.113}$$

$$\tilde{\sigma}_1^2 = \frac{c_N T^2}{(T - N - 2)^2} \left(k_{N,2} \theta_N^2 + k_{N,3} \frac{N}{T} \right), \tag{3.114}$$

$$\tilde{\sigma}_2^2 = \frac{c_N k_{N,2} T^2}{\sigma_{g,N}^2 (T - N - 2)^2}, \tag{3.115}$$

$$\tilde{\sigma}_{12} = \frac{c_N k_{N,2} \lambda_{g,N} T^2}{(T - N - 2)^2}, \tag{3.116}$$

and plugging them into (3.111) yields the desired result in Equation (3.44). It is then easy to show that the $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ maximizing (3.44) is equal to (3.45), which also corresponds to Kan and Lassance (2025, Equations (51)–(52)). Finally, after some developments, plugging (3.45) into (3.44) yields Equation (3.46), which concludes the proof.

Proof of Proposition 3.6

The squared Sharpe ratio of the GMV portfolio depends on $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ as

$$\theta_{g,N}^2 = \frac{(\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N)^2}{\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N}. \tag{3.117}$$

Now, under Assumption 1, $\boldsymbol{\Sigma}_N^{-1}$ is given by (3.91), and thus,

$$\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N = \frac{N}{1 - \rho} \left(\bar{\lambda}_N - \frac{N\rho}{1 - \rho + N\rho} \bar{\theta}_{N,1} \bar{\sigma}_{N,-1} \right), \tag{3.118}$$

$$\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N = \frac{N}{1 - \rho} \left(\bar{\sigma}_{N,-2} - \frac{N\rho}{1 - \rho + N\rho} \bar{\sigma}_{N,-1}^2 \right). \tag{3.119}$$

Plugging (3.118)–(3.119) into (3.117) yields the desired result in Equation (3.47), which concludes the proof.

Proof of Proposition 3.7

The EU of the EW-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\beta})$ in (3.54) is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\beta})) = \frac{1}{2\gamma} \left(2\beta_1 \tilde{\mu}_1 + 2\beta_2 \mu_{ew,N} - \beta_1^2 \tilde{\sigma}_1^2 - \beta_2^2 \sigma_{ew,N}^2 - 2\mathbb{E}[\hat{\boldsymbol{\mu}}' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \boldsymbol{w}_{ew}] \right), \tag{3.120}$$

where $\tilde{\mu}_1$ is given by (3.112), $\tilde{\sigma}_1^2$ by (3.114), and $\mathbb{E}[\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\boldsymbol{w}_{ew}] = \frac{k_{N,1}\mu_{ew,N}T}{T-N-2}$, which yields the desired result in Equation (3.55). It is then easy to show that the $\boldsymbol{\beta} = (\beta_1, \beta_2)$ maximizing (3.55) is equal to (3.56). Finally, after some developments, plugging (3.56) into (3.55) yields Equation (3.57), which concludes the proof.

Proof of Proposition 3.8

We have that $\theta_{ew,N}^2 = \mu_{ew,N}^2/\sigma_{ew,N}^2$, where $\mu_{ew,N} = \bar{\mu}_N$ and, under Assumption 1,

$$\sigma_{ew,N}^2 = \frac{1}{N^2}\mathbf{1}_N'\boldsymbol{\Sigma}\mathbf{1}_N = \frac{1}{N}\left(\bar{\sigma}_{N,2} + \frac{\rho}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_i\sigma_j\right). \tag{3.121}$$

This yields the desired result in Equation (3.58) after noticing that

$$\frac{\rho}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_i\sigma_j = \rho N\bar{\sigma}_{N,1}^2 - \rho\bar{\sigma}_{N,2}, \tag{3.122}$$

which concludes the proof.

Proof of Proposition 3.9

Throughout the proof, we denote

$$f(N,\rho) := \frac{N\rho}{1-\rho+N\rho}. \tag{3.123}$$

Bias of $\hat{\delta}_N$. The expectation of $\hat{\delta}_N$ is

$$\mathbb{E}[\hat{\delta}_N] = \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[\hat{s}_i^2] - \frac{f(N,\rho)}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^{M}\hat{s}_i\right)^2\right]. \tag{3.124}$$

Given that $\mathbb{E}[\hat{s}_i^2] = s_i^2 + \frac{1}{T}$, we have

$$\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[\hat{s}_i^2] = \bar{\theta}_{M,2} + \frac{1}{T}. \tag{3.125}$$

Moreover,

$$\frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^{M}\hat{s}_i\right)^2\right] = \bar{\theta}_{M,1}^2 + \frac{1}{M^2}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{M}\hat{s}_i\right], \tag{3.126}$$

where

$$\frac{1}{M^2}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{M}\hat{s}_i\right] = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{C}\mathrm{ov}[\hat{s}_i,\hat{s}_j] = \frac{1}{T}\left(\frac{1}{M} + \frac{M-1}{M}\rho\right). \tag{3.127}$$

Plugging (3.125)–(3.127) into (3.124) yields

$$\mathbb{E}[\hat{\delta}_N] = \delta_N + \frac{1}{T}\left(1 - \frac{f(N,\rho)(1-\rho+M\rho)}{M}\right), \tag{3.128}$$

which corresponds to (3.76), as desired.

Bias of $\hat{r}_{g,N}$. The expectation of $\hat{r}_{g,N}$ is

$$\mathbb{E}[\hat{r}_{g,N}] = \frac{\mathbb{E}\left[\left(\frac{1}{M}\sum_{i=1}^{M}\hat{\lambda}_i - \frac{f(N,\rho)}{M}\bar{\sigma}_{M,-1}\sum_{i=1}^{M}\hat{s}_i\right)^2\right]}{\bar{\sigma}_{M,-2} - f(N,\rho)\bar{\sigma}^2_{M,-1}}. \tag{3.129}$$

Given that $\mathbb{E}[\hat{\lambda}_i] = \lambda_i$ and $\mathbb{E}[\hat{s}_i] = s_i$, we have

$$\mathbb{E}[\hat{r}_{g,N}] = r_{g,N} + \frac{1}{M^2}\frac{\mathbb{V}\text{ar}\left[\sum_{i=1}^{M}\hat{\lambda}_i\right] + f(N,\rho)^2\bar{\sigma}^2_{M,-1}\mathbb{V}\text{ar}\left[\sum_{i=1}^{M}\hat{s}_i\right] - 2f(N,\rho)\bar{\sigma}_{M,-1}\mathbb{C}\text{ov}\left[\sum_{i=1}^{M}\hat{\lambda}_i, \sum_{i=1}^{M}\hat{s}_i\right]}{\bar{\sigma}_{M,-2} - f(N,\rho)\bar{\sigma}^2_{M,-1}}, \tag{3.130}$$

where $\frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^{M}\hat{s}_i\right]$ is given by (3.127) and

$$\frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^{M}\hat{\lambda}_i\right] = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{C}\text{ov}[\hat{\lambda}_i, \hat{\lambda}_j] = \frac{1}{T}\left(\frac{1-\rho}{M}\bar{\sigma}_{M,-2} + \rho\bar{\sigma}^2_{M,-1}\right), \tag{3.131}$$

$$\frac{1}{M^2}\mathbb{C}\text{ov}\left[\sum_{i=1}^{M}\hat{\lambda}_i, \sum_{i=1}^{M}\hat{s}_i\right] = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{C}\text{ov}[\hat{\lambda}_i, \hat{s}_j] = \frac{1-\rho+M\rho}{MT}\bar{\sigma}_{M,-1}. \tag{3.132}$$

Plugging (3.127) and (3.131)–(3.132) into (3.130) yields

$$\mathbb{E}[\hat{r}_{g,N}] = r_{g,N} + \frac{1}{T}\frac{\frac{1-\rho}{M}\bar{\sigma}_{M,-2} + \left(\rho + \frac{f(N,\rho)^2(1-\rho+M\rho)}{M} - \frac{2f(N,\rho)(1-\rho+M\rho)}{M}\right)\bar{\sigma}^2_{M,-1}}{\bar{\sigma}_{M,-2} - f(N,\rho)\bar{\sigma}^2_{M,-1}}, \tag{3.133}$$

which corresponds to (3.77), as desired.

Bias of $\hat{r}_{ew,N}$. The expectation of $\hat{r}_{ew,N}$ is

$$\mathbb{E}[\hat{r}_{ew,N}] = \frac{(1-\rho)}{(1-\rho)\bar{\sigma}_{M,2} + \rho N\bar{\sigma}^2_{M,1}}\frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^{M}\hat{\mu}_i\right)^2\right]. \tag{3.134}$$

Given that $\mathbb{E}[\hat{\mu}_i] = \mu_i$, we have

$$\frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^{M}\hat{\mu}_i\right)^2\right] = \bar{\mu}^2_M + \frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^{M}\hat{\mu}_i\right]. \tag{3.135}$$

Moreover,

$$\frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^{M}\hat{\mu}_i\right] = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{C}\text{ov}[\hat{\mu}_i, \hat{\mu}_j] = \frac{(1-\rho)\bar{\sigma}_{M,2} + \rho M\bar{\sigma}^2_{M,1}}{MT}. \tag{3.136}$$

Plugging (3.135)–(3.136) into (3.134) yields

$$\mathbb{E}[\hat{r}_{ew,N}] = r_{ew,N} + \frac{1-\rho}{MT} \times \frac{(1-\rho)\bar{\sigma}_{M,2} + \rho M \bar{\sigma}^2_{M,1}}{(1-\rho)\bar{\sigma}_{M,2} + \rho N \bar{\sigma}^2_{M,1}}, \tag{3.137}$$

which corresponds to (3.78) and concludes the proof.

Proof of Proposition 3.10

Consider a pair of selection rules (k, l) that both select N_t assets out of M available ones. They randomly select assets from a Bernoulli distribution without replacement and are mutually independent. The selections are stored in the $(M \times 1)$ selection vectors $\boldsymbol{y}_{k,t}$ and $\boldsymbol{y}_{l,t}$.

The number of assets selected in rule $\boldsymbol{y}_{k,t}$ is $\sum_{i=1}^{M} y_{i,k,t} = \boldsymbol{y}'_{k,t}\mathbf{1}_M$ and the overlap between rules k and l is $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \sum_{i=1}^{M} y_{i,k,t}y_{i,l,t}$. We are interested in the expected overlap between these two rules, denoted $\bar{N}_{t,M}$. Using the independence properties of the two rules and that $y_{i,k,t} \in \{0,1\}\ \forall i,k,t$, we can write $\bar{N}_{t,M}$ as

$$\begin{aligned}\bar{N}_{t,M} &= \mathbb{E}\left[\sum_{i=1}^{M} y_{i,k,t}y_{i,l,t} \middle| \sum_{i=1}^{M} y_{i,k,t} = \sum_{i=1}^{M} y_{i,l,t} = N_t\right] \\ &= \sum_{i=1}^{M} \mathbb{P}\left[y_{i,k,t} = 1 \middle| \boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right] \times \mathbb{P}[y_{i,l,t} = 1 | \boldsymbol{y}'_{l,t}\mathbf{1}_M = N_t].\end{aligned} \tag{3.138}$$

Consider the probability of getting $\boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}}$ where $\tilde{\boldsymbol{y}}$ is a binary vector such that $\sum_{i=1}^{M} \tilde{y}_i = n$. Given that the sum of i.i.d. Bernoulli random variables follows a binomial distribution, we can write, using Bayes theorem,

$$\begin{aligned}\mathbb{P}\left[\boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}} \middle| \boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right] &= \frac{\mathbb{P}\left[\boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t \middle| \boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}}\right] \times \mathbb{P}\left[\boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}}\right]}{\mathbb{P}\left[\boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right]} \\ &= \frac{\mathbb{1}_{\{n=N_t\}} \times p^n(1-p)^{M-n}}{\binom{M}{N_t}p^{N_t}(1-p)^{M-N_t}} = \frac{\mathbb{1}_{\{n=N_t\}}}{\binom{M}{N_t}}\end{aligned} \tag{3.139}$$

Marginalizing this joint distribution over the entries $j \neq i$ yields the marginal distribution of $y_{i,k,t}$ conditional upon $\sum_{i=1}^{M} y_{i,k,t} = N_t$,

$$\begin{aligned}\mathbb{P}\left[y_{i,k,t} = 1 \middle| \boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right] &= \mathbb{P}\left[y_{i,k,t} = 1 \middle| \sum_{j\neq i} y_{j,k,t} = N_t - 1\right] \\ &= \sum_{h\neq i} \frac{\mathbb{1}_{\{\sum_{j\neq h} b_j = N_t - 1\}}}{\binom{M}{N_t}} \\ &= \frac{\binom{M-1}{N_t-1}}{\binom{M}{N_t}}.\end{aligned} \tag{3.140}$$

Plugging (3.140) into (3.138) yields the desired result and concludes the proof.

Chapter 4

Optimal Shrinkage of the Covariance Matrix for Portfolio Selection

Abstract

We introduce analytical linear and nonlinear shrinkage estimators of the sample covariance matrix that are optimal for mean-variance portfolio choice. Unlike classical methods based on statistical loss functions like the mean squared error, our shrinkage covariance matrices optimize the expected out-of-sample portfolio utility and account for estimation errors stemming from the mean vector. Our estimators shrink the sample eigenvalues more intensively than in conventional methods, and they especially shrink the contribution of principal components with small squared Sharpe ratios. Overall, our methodology estimates the covariance matrix and the optimal mean-variance portfolio jointly, which we show delivers significant empirical gains relative to the conventional two-step approach and recently proposed regularized mean-variance portfolios.

Reference

This chapter is based on:

Lassance, N., Vanderveken, R. and Vrins, F. (2025b). Optimal shrinkage of the covariance matrix for portfolio selection. *Work in progress*.

4.1 Introduction

We introduce shrinkage estimators of the sample covariance matrix that are optimal for mean-variance portfolio choice. The use of shrinkage covariance matrices has become pervasive in portfolio selection after Ledoit and Wolf (2003a) warned investors that "nobody should be using the sample covariance matrix for the purpose of portfolio optimization." Following a first era of linear shrinkage estimators (Ledoit and Wolf, 2003a,b, 2004; Chen et al., 2010; Bodnar et al., 2014), where the sample covariance matrix is shrunk toward a structured target estimator (e.g., the scaled identity matrix in Ledoit and Wolf (2004)), a second era of nonlinear shrinkage followed (Ledoit and Wolf, 2012, 2015, 2017, 2020, 2022a; Hediger et al., 2023), where each eigenvalue of the sample covariance matrix is shrunk with its own intensity.

The common theme among these established linear and nonlinear shrinkage estimators, which we aim to challenge, is that they minimize a statistical loss relative to the true covariance matrix, typically measured with the mean squared error (MSE), i.e., the Frobenius norm.[1] That is, these methods provide generic estimators that can then be used for various applications; see Ledoit and Wolf (2022b) for a review of applications across different fields.

Elmachtoub and Grigas (2022) call for switching from "predict-then-optimize" to "predict-and-optimize" in solving decision problems, i.e., to design estimators using loss functions related to the application of interest.[2] We answer their call and introduce analytical finite-sample linear and nonlinear shrinkage estimators of the sample covariance matrix designed to optimize the mean-variance investor's expected out-of-sample utility (EU).[3] Since Kan and Zhou (2007), the EU has been used extensively in the literature on portfolio choice with parameter uncertainty; see Lassance et al. (2024a) for a review. Our analytical finite-sample approach builds on the multivariate elliptical distribution for the asset returns[4] and contrasts with recent high-dimensional data-driven approaches to robust estimation of mean-variance portfolios like MAXSER (Ao et al., 2019), the robust SDF of Kozak et al. (2020), and UPSA (Kelly et al., 2023). Empirically, our mean-variance portfolios constructed with EU-optimal shrinkage covariance matrices compare positively

[1]The nonlinear shrinkage method in Ledoit and Wolf (2012, 2017, 2020) has been shown to be optimal not only under the MSE but also under mean-variance portfolio loss functions. However, this equivalence assumes that mean returns are known, which is not a reasonable assumption.

[2]See Vanderschueren et al. (2022) for a review and comparison of the two learning approaches. Cai et al. (2024) compare the two paradigms for constructing shrinkage estimators of the portfolio weights as in Kan and Zhou (2007) and Kan et al. (2021).

[3]Some papers also shrink the inverse covariance matrix; see Frahm and Memmel (2010), Kourtis et al. (2012), Bodnar et al. (2016), Shi et al. (2020), and Mattera (2025). However, Ledoit and Wolf (2017) show that the loss function a mean-variance investor should use is the MSE of the covariance matrix, not its inverse.

[4]We accommodate the multivariate elliptical distribution and not just the normal distribution by leveraging Kan and Lassance (2025), which extends Kan and Zhou (2007) to the multivariate elliptical distribution.

to these benchmarks, among others.

We start our theoretical analysis with the linear shrinkage of the sample covariance matrix toward the scaled identity matrix, as in Ledoit and Wolf (2004), which amounts to shrinking the sample eigenvalues toward their mean. Whereas Ledoit and Wolf (2004) consider the high-dimensional asymptotic regime where the number of assets and the sample size go to infinity, we derive the finite-sample MSE-optimal shrinkage intensity, $\delta^\star_{mse}$, which extends results in Chen et al. (2010) and Ollila and Raninen (2019). We also derive the EU-optimal shrinkage intensity, $\delta^\star_{eu}$, and show that unlike $\delta^\star_{mse}$, it depends not only on the eigenvalues but also on the squared Sharpe ratios of principal components (PCs). In calibrations, we find that $\delta^\star_{eu}$ is substantially larger than $\delta^\star_{mse}$.[5] This more intensive shrinkage is due to the use of a different loss function and the fact that the EU accounts for the impact of estimation errors stemming from the mean vector. Thus, there is a substantial difference between how intensively the sample covariance matrix should be shrunk to best fit the true covariance matrix and to deliver the best mean-variance portfolio out of sample.

We then turn to nonlinear shrinkage of sample eigenvalues as introduced by Ledoit and Wolf (2012), which we formulate so that each inverse eigenvalue is shrunk toward zero with its own intensity. We derive the finite-sample MSE-optimal and EU-optimal shrinkage intensities, $\delta^\star_{mse,i}$ and $\delta^\star_{eu,i}$, respectively, unlike the usual high-dimensional asymptotic approach. We show that $\delta^\star_{mse,i}$ shrinks the contribution of lower-variance PCs more intensively, whereas $\delta^\star_{eu,i}$ does so for PCs with lower squared Sharpe ratios. Like in linear shrinkage, the shrinkage intensities $\delta^\star_{eu,i}$ are substantially larger than $\delta^\star_{mse,i}$. In particular, the EU shrinks all inverse eigenvalues closer to zero, whereas the MSE increases small inverse eigenvalues. Although the EU shrinks the contribution of low-variance PCs substantially, their contribution does not vanish, unlike common PC-sparse mean-variance portfolio methodologies (Chen and Yuan, 2016; Kozak et al., 2020). This result is consistent with Kelly et al. (2023) who find that "PC-sparsity is just an artifact of inefficient shrinkage." We also show that the oracle EU-optimal shrinkage intensities deliver a mean-variance portfolio whose EU is always positive and substantially larger than that under the MSE-optimal intensities. In particular, the EU-based nonlinear shrinkage method is the only shrinkage method that always delivers a positive EU, among those considered.

Finally, after introducing bona fide estimators of our MSE and EU-optimal linear and nonlinear shrinkage intensities, we evaluate the empirical out-of-sample performance of our proposed mean-variance portfolios. We compare them to numerous benchmarks including mean-variance and global-minimum-variance (GMV) portfolios under MSE-optimal, PC-sparse, and cross-validated shrinkage covariance matrices, equally weighted

[5]For example, when the sample size is 120 months, the returns are multivariate t-distributed with six degrees of freedom, and the parameters are calibrated to a dataset of 25 portfolios sorted on size and book-to-market spanning July 1926 to July 2024, we have $\delta^\star_{eu} = 0.24$ whereas $\delta^\star_{mse} = 0.05$.

(EW) portfolios, and four recent mean-variance portfolio estimators using methods based on regularization, PCA, high-dimensional asymptotics, and cross-validation, namely, MAXSER (Ao et al., 2019), the robust SDF of Kozak et al. (2020), the universal portfolio shrinkage approximator (UPSA) (Kelly et al., 2023), and a factor-plus-alpha strategy based on the arbitrage portfolio of Da et al. (2024).[6] We compare the strategies in terms of EU net of transaction costs, across six characteristic-sorted datasets, for various sample sizes and risk-aversion coefficients.

The empirical results give support to our theoretical findings as our EU-optimal shrinkage covariance matrices deliver substantial and statistically significant performance gains relative to conventional MSE-optimal shrinkage. For instance, on average across the six datasets and given a risk-aversion coefficient equal to three, our EU-optimal nonlinear shrinkage covariance matrix delivers an annualized net EU of 9.60 and 9.37 for a sample size of 120 and 240 months, respectively, versus -4.74 and -17.36 under the MSE-optimal nonlinear shrinkage estimator. Overall, the comparison to the other benchmarks is also favorable.[7] Two distinguishing empirical feats achieved by our mean-variance portfolios are their low turnover due to the intensive shrinkage of low-variance eigenvalues, and their consistently strong performance. In particular, our EU-optimal linear shrinkage estimator delivers a positive net EU in all considered configurations except one, and our nonlinear estimator does so in all configurations. In comparison, the other benchmarks yield more unstable and often negative utilities. Overall, the empirical results demonstrate the practical value of designing shrinkage estimators using a loss function tailored to the portfolio selection problem.

The rest of the chapter is structured as follows. Section 4.2 introduces the problem and the notation. In Sections 4.3 and 4.4, we derive the MSE-optimal and EU-optimal linear and nonlinear shrinkage covariance matrix estimators, respectively. In Section 4.5, we introduce bona fide estimators of the oracle shrinkage intensities. Section 4.6 contains our empirical analysis. Section 4.7 concludes. The supplementary material contains additional theoretical and empirical tests, as well as the proofs of all results.

4.2 Presentation of the problem

In this section, we introduce the necessary background surrounding the problem we aim to solve in this chapter, as well as the notation.

[6] In total, we consider 21 portfolio strategies; see Table 4.2 for a description of each of them.

[7] In particular, across 24 configurations, our EU-optimal nonlinear shrinkage covariance matrix outperforms MAXSER (Ao et al., 2019), the robust SDF of Kozak et al. (2020), UPSA (Kelly et al., 2023), and the factor-plus-alpha strategy based on Da et al. (2024) in 24, 16, 18, and 22 cases, respectively.

4.2.1 Mean-variance portfolio and plug-in estimator

Let $\boldsymbol{r} \in \mathbb{R}^N$ be the vector of excess returns on $N \geq 2$ assets, which has a mean $\boldsymbol{\mu}$ and a positive-definite covariance matrix $\boldsymbol{\Sigma}$. The eigendecomposition of $\boldsymbol{\Sigma}$ is $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, where $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N]$ is the matrix of eigenvectors and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ is the matrix of eigenvalues sorted in descending order.

Given a risk-aversion coefficient $\gamma > 0$, the mean-variance utility is

$$U(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}. \tag{4.1}$$

The corresponding optimal mean-variance portfolio is

$$\boldsymbol{w}^\star = \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmax}}\, U(\boldsymbol{w}) = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{4.2}$$

which delivers the following maximum utility:

$$U(\boldsymbol{w}^\star) = \frac{\theta^2}{2\gamma} \quad \text{with} \quad \theta^2 = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{4.3}$$

The traditional plug-in estimator of $\boldsymbol{w}^\star$ is

$$\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}, \tag{4.4}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the sample unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ obtained from an i.i.d. sample of returns of size $T \geq 2$, $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$. That is,

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})', \tag{4.5}$$

where $\hat{\boldsymbol{\Sigma}}^{-1}$ almost surely exists provided $T \geq N + 1$. As is well known, $\hat{\boldsymbol{w}}^\star$ performs poorly out of sample. In particular, in the next proposition, we use results in Kan and Lassance (2025) and present the expected out-of-sample utility (EU) delivered by $\hat{\boldsymbol{w}}^\star$ under the assumption that the sample of returns $(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_T)$ is drawn from a multivariate elliptical distribution. For that purpose, we use the stochastic representation in El Karoui (2010, 2013), i.e.,

$$\boldsymbol{r}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau) \quad \Leftrightarrow \quad \boldsymbol{r}_t \overset{d}{=} \boldsymbol{\mu} + (\tau\boldsymbol{\Sigma})^{1/2}\boldsymbol{z}, \tag{4.6}$$

where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}_N, \boldsymbol{I}_N)$, τ is a positive random variable satisfying $\mathbb{E}[\tau] = 1$, and $\boldsymbol{z} \perp \tau$. For example, we recover the multivariate normal distribution when $\tau = 1$ collapses to a constant, and the multivariate t-distribution when $\tau \sim (\nu - 2)/\chi^2_\nu$, where χ^2_ν denotes a chi-square distribution with $\nu > 2$ degrees of freedom.

Proposition 4.1. *Let the i.i.d. sample of returns $(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_T)$ be drawn from a multivariate elliptical distribution in* (4.6)*. Moreover, let $T > N + 4$, $\boldsymbol{M} = \boldsymbol{I}_T - \boldsymbol{1}_T\boldsymbol{1}_T'/T$, $\boldsymbol{Z}$*

be a $T \times N$ matrix of independent standard normal variables, $\boldsymbol{T}$ be a diagonal matrix of T independent copies[8] *of $\tau^{1/2}$, and $\boldsymbol{T} \perp \boldsymbol{Z}$. Define k_1, k_2, and k_3 which are functions of only N, T, and the distribution of τ,*

$$k_1 = \frac{T-N-2}{N}\mathbb{E}\left[\mathrm{tr}\big((\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z})^{-1}\big)\right], \tag{4.7}$$

$$k_2 = \frac{(T-N-1)(T-N-2)(T-N-4)}{N(T-2)}\mathbb{E}\left[\mathrm{tr}\big((\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z})^{-2}\big)\right], \tag{4.8}$$

$$k_3 = \frac{(T-N-1)(T-N-2)(T-N-4)}{NT(T-2)}\mathbb{E}\left[\mathbf{1}_T'\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z})^{-2}\boldsymbol{Z}'\boldsymbol{T}\mathbf{1}_T\right], \tag{4.9}$$

provided the expectations exist, and $k_2 \geq k_1$. Also, $k_1 \geq 1$, $k_2 \geq 1$ and $k_3 \geq 1$ and $k_1 = k_2 = k_3 = 1$ when the sample of returns is drawn from a multivariate normal distribution. Then, the EU delivered by $\hat{\boldsymbol{w}}^\star$ in (4.4) *is*

$$\mathbb{E}[U(\hat{\boldsymbol{w}}^\star)] = \frac{T-1}{2\gamma(T-N-2)}\left[2k_1\theta^2 - \frac{(T-1)(T-2)}{(T-N-1)(T-N-4)}\left(k_2\theta^2 + k_3\frac{N}{T}\right)\right]. \tag{4.10}$$

The parameters k_1, k_2, and k_3 control the impact of fat tails on the EU of the plug-in estimator $\hat{\boldsymbol{w}}^\star$ and will also play a key role in our results. Kan and Lassance (2025) show that these parameters increase with estimation risk, i.e., N/T, and tail heaviness. In particular, as $N \to \infty$, $T \to \infty$, and $N/T \to \phi \in (0,1)$, we have that $k_1 = k_3$, $k_2 \geq k_1^2 \geq 1$, and both k_1 and k_2 increase with ϕ.

4.2.2 Conventional approach to shrinking the covariance matrix

Among existing methods to improve upon the plug-in estimator $\hat{\boldsymbol{w}}^\star$, we consider shrinking the sample covariance matrix. As is standard in the literature, we consider rotation-equivariant shrinkage estimators of $\hat{\boldsymbol{\Sigma}}$ that preserve the sample eigenvectors but correct the sample eigenvalues. Specifically, let $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{\Lambda}}$ be the eigenvector and eigenvalue matrices of $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{V}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}'$. Then, we consider shrinkage estimators of $\hat{\boldsymbol{\Sigma}}$ of the form[9]

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}', \tag{4.11}$$

where $\boldsymbol{D}$ is a diagonal matrix. In the linear shrinkage approach of Ledoit and Wolf (2004), the matrix $\boldsymbol{D}$ is parametrized as $\boldsymbol{D} = (1-\delta)\hat{\boldsymbol{\Lambda}} + \delta\bar{\lambda}\boldsymbol{I}_N$, where $\bar{\lambda} = \frac{1}{N}\mathrm{tr}(\boldsymbol{\Lambda})$ and δ is a shrinkage intensity to calibrate. In the nonlinear shrinkage approach of Ledoit and Wolf (2012, 2017, 2020), all N elements of the diagonal matrix $\boldsymbol{D}$ are parameters to calibrate. In both cases, δ or the whole matrix $\boldsymbol{D}$ are calibrated so as to minimize the mean squared error (MSE), i.e., the expected squared Frobenius norm, between $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ and the true $\boldsymbol{\Sigma}$:

$$\mathrm{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \mathbb{E}\left[||\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma}||_F^2\right] = \frac{1}{N}\mathbb{E}\left[\mathrm{tr}\big((\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma})^2\big)\right]. \tag{4.12}$$

[8] In case τ is a constant, $\boldsymbol{T}$ is then simply a matrix with the constant τ on the diagonal.

[9] Kelly et al. (2023) refer to the class of estimators (4.11) as spectral shrinkage estimators.

Remark 4.1. Ledoit and Wolf (2004, 2012, 2015, 2017, 2020) and Bodnar et al. (2014, 2016) do not need to take the expectation in (4.12) because they minimize the high-dimensional asymptotic limit of the Frobenius norm as $N \to \infty$, $T \to \infty$, and $N/T \to c < \infty$ which is nonrandom. Also, Ledoit and Wolf (2017) motivate their loss function from a mean-variance portfolio objective. However, because they assume that the mean returns $\boldsymbol{\mu}$ are known, they show that their solution for the optimal matrix $\boldsymbol{D}$ is equivalent to that minimizing the Frobenius norm $||\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma}||_F$.

4.2.3 Portfolio approach to shrinking the covariance matrix

The common theme of the shrinkage methods in Section 4.2.2 is that they aim to fit the true covariance matrix $\boldsymbol{\Sigma}$. However, for an investor, the aim is rather to obtain the estimate of $\boldsymbol{\Sigma}$ that delivers the best out-of-sample portfolio performance. Therefore, we propose to calibrate the diagonal matrix $\boldsymbol{D}$ in linear and nonlinear shrinkage estimation so as to maximize the EU of the corresponding portfolio, i.e.,

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \mathbb{E}[U(\hat{\boldsymbol{w}}_{\boldsymbol{D}})] = \mathbb{E}\left[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D}}\right] \quad \text{with} \quad \hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}. \tag{4.13}$$

There are four main differences between the proposed shrinkage approach and the conventional ones based on the MSE. First, we calibrate the shrinkage intensities using a portfolio objective, the EU, rather than the MSE statistical loss function. Second, we account for the impact that estimation errors stemming from the sample mean vector $\hat{\boldsymbol{\mu}}$ have on this portfolio performance. Third, we account for more estimation errors stemming from the sample eigenvalues. Specifically, in linear shrinkage, we account for errors from $\bar{\lambda}$ and, in nonlinear shrinkage, we set the matrix $\boldsymbol{D}$ so that it shrinks the sample eigenvalues instead of letting $\boldsymbol{D}$ be constant. Fourth, we consider the finite-sample setting in which N and T are fixed rather than an asymptotic setting.

Remark 4.2. To evaluate the different expectations appearing in our mathematical derivations in a finite-sample setting with N and T fixed, we assume that the sample of returns is drawn from a multivariate elliptical distribution, as in Equation (4.6). Most of the literature on shrinkage mean-variance portfolio estimators in a finite-sample setting employs the multivariate normal distribution assumption (Kan and Zhou, 2007; Tu and Zhou, 2011; Kan et al., 2021; Lassance et al., 2024a). Kan and Lassance (2025) extend this literature to the multivariate elliptical distribution and we leverage some of their results for our problem.

4.3 Linear shrinkage

In this section, we consider the linear shrinkage of the sample eigenvalues as pioneered by Ledoit and Wolf (2004), i.e.,

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}' \quad \text{with} \quad \boldsymbol{D} = (1-\delta)\hat{\boldsymbol{\Lambda}} + \delta\hat{\bar{\lambda}}\boldsymbol{I}_N, \ \hat{\bar{\lambda}} = \frac{1}{N}\text{tr}(\hat{\boldsymbol{\Lambda}}). \tag{4.14}$$

Note that Ledoit and Wolf (2004) derive the optimal shrinkage intensity δ minimizing the MSE for the oracle $\bar{\lambda}$, which is because $\hat{\bar{\lambda}}$ is a consistent estimator. However, because we work in a finite-sample setting, we account for estimation errors in $\bar{\lambda}$ and shrink the eigenvalues towards their sample mean $\hat{\bar{\lambda}}$ in (4.14).

4.3.1 Linear shrinkage based on mean squared error

In the next proposition, we derive the optimal shrinkage intensity δ minimizing the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.14).

Proposition 4.2. *Let the i.i.d. sample of returns $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ be drawn from a multivariate elliptical distribution, $\boldsymbol{r}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau)$. Let $\kappa = \mathbb{V}\text{ar}[\tau]/3$ be the excess kurtosis of any asset return divided by three, which we assume exists. Then, the shrinkage intensity δ minimizing the MSE in* (4.12) *of the linear shrinkage estimator $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in* (4.14) *is*

$$\delta^{\star}_{mse} = \underset{\delta\in\mathbb{R}}{\text{argmin}}\, MSE(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \frac{\frac{(N-2)s+N+2}{T-1} + \frac{\kappa[2(N-1)s+N+2]}{T}}{Ns + \frac{(N-2)s+N+2}{T-1} + \frac{\kappa[2(N-1)s+N+2]}{T}} \in [0,1], \tag{4.15}$$

where

$$s = \frac{N\frac{\text{tr}(\boldsymbol{\Lambda}^2)}{\text{tr}(\boldsymbol{\Lambda})^2} - 1}{N-1} \in [0,1) \tag{4.16}$$

is the sphericity parameter. In particular, $s = 0$ when $\boldsymbol{\Lambda} = c\boldsymbol{I}_N$ $(c > 0)$, in which case $\delta^{\star}_{mse} = 1$.

Proposition 4.2 shows that $\delta^{\star}_{mse}$ does not depend on $\boldsymbol{\mu}$, does depend on $\boldsymbol{\Sigma}$ but only via the distribution of its eigenvalues through the sphericity parameter s, and depends on the elliptical fat tails only via the excess kurtosis parameter κ.

Remark 4.3. Proposition 4.2 extends results in Ollila and Raninen (2019), who derive the MSE-optimal δ under the multivariate elliptical distribution but considering the oracle $\bar{\lambda}$ instead of $\hat{\bar{\lambda}}$ in (4.14). It also extends results in Chen et al. (2010), who also derive the MSE-optimal δ considering $\hat{\bar{\lambda}}$ in (4.14) but under the multivariate normal distribution.

4.3.2 Linear shrinkage based on expected out-of-sample utility

In the next proposition, we derive the EU of the mean-variance portfolio using the linear shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.14).

Proposition 4.3. *Let $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$, with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (4.14), be obtained from an i.i.d. sample $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ drawn from a multivariate elliptical distribution, $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$. Then, provided $T \geq N + 5$, the EU in (4.13) of the shrinkage mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ is*

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-1)}{\gamma(T-N-2)}\boldsymbol{f}(\delta)'\boldsymbol{\theta}_{PC}^2 - \frac{(T-1)^2}{2\gamma(T-N-2)(T-N-4)}\boldsymbol{f}(\delta)'\boldsymbol{A}\boldsymbol{f}(\delta), \quad (4.17)$$

where $\boldsymbol{\theta}_{PC}^2$ is the vector of squared Sharpe ratios of principal components (PCs), $(\boldsymbol{\theta}_{PC}^2)_i = \theta_{PC_i}^2 = (\boldsymbol{v}_i'\boldsymbol{\mu})^2/\lambda_i$, the vector $\boldsymbol{f}(\delta)$ is defined as

$$f_i(\delta) = \frac{\lambda_i}{(1-\delta)\lambda_i + \delta\bar{\lambda}}, \quad (4.18)$$

the (i,j)-th entry of the matrix $\boldsymbol{A}$ is given by

$$A_{ij} = \begin{cases} (k_2\theta_{PC_i}^2 + k_3/T) & \text{if } i = j, \\ \frac{1}{T-N-1}(k_2\theta_{PC_i}^2 + k_3/T) & \text{if } i \neq j, \end{cases} \quad (4.19)$$

and the parameters (k_1, k_2, k_3) are given by (4.7)–(4.9), which we assume exist.

We then define the optimal shrinkage intensity $\delta_{eu}^\star$ as the shrinkage intensity maximizing the EU in (4.17),

$$\delta_{eu}^\star = \underset{\delta \in \mathbb{R}}{\operatorname{argmax}}\, \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}). \quad (4.20)$$

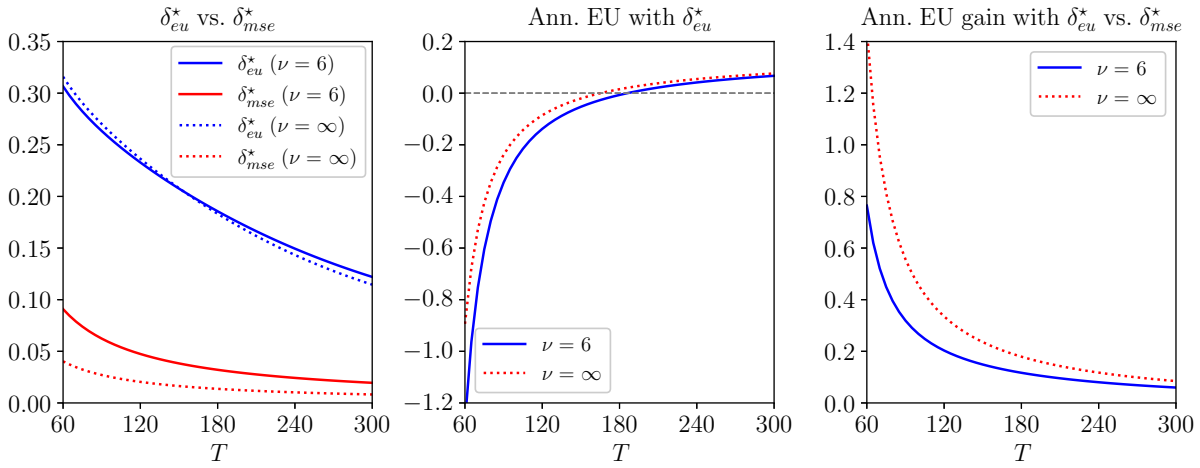
There is no closed-form expression for $\delta_{eu}^\star$, but it can easily be found numerically.[10]

Remark 4.4. By assuming a similar distribution for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = (1-\delta)\hat{\boldsymbol{\Sigma}} + \delta\hat{\bar{\lambda}}\boldsymbol{I}_N$ to that of $\hat{\boldsymbol{\Sigma}}$, we overstate the amount of randomness. Intuitively, $\hat{\bar{\lambda}}\boldsymbol{I}_N$ is less sensitive to sample noise than the sample estimate $\hat{\boldsymbol{\Sigma}}$ because of the averaging in $\hat{\bar{\lambda}}$ and the fact that the diagonal is set to a constant and all off-diagonal elements are set to zero. Because $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ shrinks $\hat{\boldsymbol{\Sigma}}$ toward $\hat{\bar{\lambda}}\boldsymbol{I}_N$, it should then be more stable than $\hat{\boldsymbol{\Sigma}}$. Therefore, our approach to determining $\delta_{eu}^\star$ is conservative, which is expected to compensate for the estimation errors stemming from the bona-fide estimator of $\delta_{eu}^\star$. We evaluate this intuition in Section 4.B of the supplementary material, where we test the approximation accuracy of $\delta_{eu}^\star$ via simulations and find that it is larger than, but close to, the optimal δ without approximation.

Compared to the MSE-optimal shrinkage intensity, $\delta_{mse}^\star$ in (4.15), that does not depend on $\boldsymbol{\mu}$ and only depends on $\boldsymbol{\Sigma}$ via the distribution of its eigenvalues through the sphericity parameter s, the EU-optimal $\delta_{eu}^\star$ in (4.20) depends on both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ via the squared Sharpe ratios of PCs $\theta_{PC_i}^2$, as well as the eigenvalues λ_i. In particular, the shrinkage intensity δ

[10] $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ is proportional to $1/\gamma$, and thus, $\delta_{eu}^\star$ does not depend on γ. This is also true for the nonlinear shrinkage intensities we consider later in Section 4.4.2.

Figure 4.1: Comparison of linear shrinkage intensities $\delta^\star_{eu}$ and $\delta^\star_{mse}$



Notes. The left panel depicts the linear shrinkage intensities $\delta^\star_{mse}$ and $\delta^\star_{eu}$ in (4.15) and (4.20), which respectively minimize the MSE of the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.14) and maximize the EU in (4.17) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$. The middle and right panels depict the annualized EU delivered by $\delta^\star_{eu}$ and its difference relative to that delivered by $\delta^\star_{mse}$. The gray dashed horizontal line in the middle panel indicates the zero-utility level. We consider a sample size $T \in [60, 300]$ and calibrate the eigenvalues $\boldsymbol{\Lambda}$ and the PCs' squared Sharpe ratios $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024. The asset returns follow a multivariate t-distribution with degrees of freedom $\nu \in \{6, \infty\}$. We set a risk-aversion coefficient $\gamma = 3$.

determines the shrinkage factor $f_i(\delta)$, which down-weights $\theta^2_{PC_i}$ if λ_i is below the average $\bar{\lambda}$ and over-weights the remaining ones. Also, $\delta^\star_{eu}$ depends on the elliptical distribution via (k_1, k_2, k_3), compared to the kurtosis parameter κ for $\delta^\star_{mse}$.

Finally, we expect the EU-optimal shrinkage intensity, $\delta^\star_{eu}$, to be larger than the MSE-optimal one, $\delta^\star_{mse}$. This is because $\delta^\star_{eu}$ accounts for the substantial impact of estimation errors stemming from the sample mean $\hat{\boldsymbol{\mu}}$ on the EU. In fact, in Section 4.A of the supplementary material, we show that even when $\boldsymbol{\Sigma}$ is known, it is optimal to shrink it quite intensively to alleviate the impact of estimation errors in $\hat{\boldsymbol{\mu}}$.

4.3.3 Comparison of linear shrinkage intensities

We now compare the oracle linear shrinkage intensities $\delta^\star_{mse}$ and $\delta^\star_{eu}$ given by (4.15) and (4.20), respectively. We have that $\delta^\star_{mse}$ depends on N, T, $\boldsymbol{\Lambda}$, and κ, whereas $\delta^\star_{eu}$ depends on N, T, $\boldsymbol{\Lambda}$, $\boldsymbol{\theta}^2_{PC}$, and (k_1, k_2, k_3). We calibrate $\boldsymbol{\Lambda}$ and $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market (25SBM) spanning July 1926 to July 2024. We consider a multivariate t-distribution for the asset returns and we consider degrees of freedom $\nu \in \{6, \infty\}$, which gives $\kappa = 2/(\nu - 4)$. The distribution is multivariate normal when $\nu = \infty$, in which case $k_1 = k_2 = k_3 = 1$ and $\kappa = 0$. In Figure 4.1, we depict the linear shrinkage intensities $\delta^\star_{eu}$ and $\delta^\star_{mse}$, the annualized EU delivered by $\delta^\star_{eu}$, and its difference relative to that delivered by $\delta^\star_{mse}$, as a function of the sample size $T \in [60, 300]$. We also set a risk-aversion coefficient $\gamma = 3$ without loss of generality.

We make three main observations. First, the left panel shows that the EU-maximizing

shrinkage intensity $\delta_{eu}^{\star}$ is higher than $\delta_{mse}^{\star}$, which is because it accounts for the impact of estimation errors stemming from both the mean and covariance matrix. The difference is substantial, e.g., when $T = 120$ and $\nu = 6$, $\delta_{eu}^{\star} = 0.24$ is almost five times larger than $\delta_{mse}^{\star} = 0.05$. Second, the middle panel shows that $\delta_{eu}^{\star}$ can deliver a positive EU for realistic sample sizes, e.g., $T > 190$ for $\nu = 6$ and $T > 170$ for $\nu = \infty$, unlike with $\delta_{mse}^{\star}$ where the required sample size is much larger. However, when setting T below these bounds, $\delta_{eu}^{\star}$ yields a negative EU, highlighting a limit of linear shrinkage that we address using nonlinear shrinkage in Section 4.4. Third, the right panel demonstrates the large increase in EU obtained with $\delta_{eu}^{\star}$ instead of $\delta_{mse}^{\star}$, particularly as T decreases. For example, when $T = 120$, the annualized EU gain is 0.20 for $\nu = 6$ and 0.34 for $\nu = \infty$.

4.4 Nonlinear shrinkage

In this section, we consider the nonlinear shrinkage of the sample eigenvalues as developed by Ledoit and Wolf (2012, 2015, 2017, 2020, 2022a). In nonlinear shrinkage, each sample eigenvalue is shrunk with its own intensity. Therefore, unlike in linear shrinkage, it does not matter which target we shrink toward, and for simplicity we shrink the sample eigenvalues toward zero. Specifically, we consider a shrinkage covariance matrix of the form

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}' \quad \text{with} \quad \boldsymbol{D} = (\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\hat{\boldsymbol{\Lambda}}, \tag{4.21}$$

where $\boldsymbol{\Delta}$ is a diagonal matrix containing the N shrinkage intensities δ_i to calibrate, i.e., $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta}) = \text{diag}(\delta_1, \ldots, \delta_N)$. We use the formulation in (4.21) so that δ_i is the shrinkage intensity of the inverse sample eigenvalue $1/\hat{\lambda}_i$ (indeed, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1} = \hat{\boldsymbol{V}}(\boldsymbol{I}_N - \boldsymbol{\Delta})\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{V}}'$), which will ease the interpretation of the results in this section. In this formulation, if δ_i is close to 1, then the inverse sample eigenvalue is heavily shrunk toward zero.

Remark 4.5. The standard approach in nonlinear shrinkage, as in Ledoit and Wolf (2012, 2015, 2017, 2020, 2022a), is to set $\boldsymbol{D}$ to be a constant diagonal matrix to calibrate. In this setting, it is well known that the matrix $\boldsymbol{D}$ minimizing the Frobenius norm $||\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma}||_F$ is $\boldsymbol{D} = \text{diag}(\hat{\boldsymbol{v}}_1'\boldsymbol{\Sigma}\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_N'\boldsymbol{\Sigma}\hat{\boldsymbol{v}}_N)$, and one uses then the high-dimensional asymptotic limit to the quantities $\hat{\boldsymbol{v}}_i'\boldsymbol{\Sigma}\hat{\boldsymbol{v}}_i$. In contrast, by specifying $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ as in (4.21), we account for estimation errors in the sample eigenvalues in determining their optimal shrinkage.

4.4.1 Nonlinear shrinkage based on mean squared error

Finding the optimal vector of shrinkage intensities $\boldsymbol{\delta}$ minimizing the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.21) in a finite-sample setting is a difficult problem because it involves the first and second moments of the sample eigenvalues $\hat{\lambda}_i$. In the next proposition, we derive the MSE-optimal $\boldsymbol{\delta}$ assuming, as in Proposition 4.3, that $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ is obtained from an i.i.d. multivariate elliptical sample with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$. In the same spirit as Ledoit and Wolf (2017), we

constrain the shrinkage intensities so that the trace of the sample covariance matrix is preserved, i.e., $\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}})$, as it is the case under linear shrinkage.

Proposition 4.4. *Let the nonlinear shrinkage estimator* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *in* (4.21) *be obtained from an i.i.d. sample* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ *drawn from a multivariate elliptical distribution,* $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$*. Then, the shrinkage intensities* δ_i *minimizing the MSE in* (4.12) *of* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *subject to the constraint* $\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}})$ *are*

$$\boldsymbol{\delta}^{\star}_{mse} = \underset{\boldsymbol{\delta} \in \mathbb{R}^N}{\text{argmin}}\, MSE(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) \tag{4.22}$$

$$\text{subject to } \text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}}), \tag{4.23}$$

where each shrinkage intensity is given by

$$\delta^{\star}_{mse,i} = \frac{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\bar{\lambda} - \lambda_i)}{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)\bar{\lambda} + \lambda_i} < 1, \tag{4.24}$$

with $\delta^{\star}_{mse,i} \to 1$ *as* $\lambda_i \to 0$*. Moreover,* $\delta^{\star}_{mse,i} > 0$ *if* $\lambda_i < \bar{\lambda}$*, and* $\delta^{\star}_{mse,i} \leq 0$ *otherwise.*

To the best of our knowledge, Proposition 4.4 is the first non-asymptotic MSE-based solution to nonlinear shrinkage. Hediger et al. (2023) also consider nonlinear shrinkage in finite samples for elliptical distributions but in a different fashion, i.e., they use nonlinear shrinkage to correct Tyler's robust estimator (Tyler, 1987).

Proposition 4.4 shows that the inverse sample eigenvalues of sufficiently low-variance PCs are shrunk toward zero (i.e., $\delta_{mse,i} > 0$ if $\lambda_i < \bar{\lambda}$), and those of sufficiently high-variance PCs are enlarged (i.e., $\delta_{mse,i} \leq 0$ if $\lambda_i \geq \bar{\lambda}$). This is consistent with linear shrinkage, except that each inverse sample eigenvalue is assigned a different intensity $\delta^{\star}_{mse,i}$ that varies with λ_i. Moreover, $\delta^{\star}_{mse,i} \to 1$ as $\lambda_i \to 0$, i.e., there is little contribution from low-variance PCs.

4.4.2 Nonlinear shrinkage based on expected out-of-sample utility

In the next proposition, we derive the EU of the mean-variance portfolio using the nonlinear shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.21), the optimal shrinkage intensities $\boldsymbol{\delta}$, and the EU they deliver. We use the conservative approximation discussed in Remark 4.4.

Proposition 4.5. *Let* $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$*, with* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *given by* (4.21)*, be obtained from an i.i.d. sample* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ *drawn from a multivariate elliptical distribution,* $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$*. Define*

$$R_i = \frac{k_1 \theta^2_{PC_i}}{k_2 \theta^2_{PC_i} + k_3/T} \in [0, 1] \tag{4.25}$$

and $\bar{R} = \frac{1}{N}\sum_{i=1}^{N} R_i$*. Then, provided* $T \geq N + 5$*, the following holds:*

1. *The EU in* (4.13) *of the shrinkage mean-variance portfolio* $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ *is*

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-1)}{\gamma(T-N-2)}(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\theta}^2_{PC} - \frac{(T-1)^2}{2\gamma(T-N-2)(T-N-4)}(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{A}(\mathbf{1}_N - \boldsymbol{\delta}), \tag{4.26}$$

where the vector $\boldsymbol{\theta}^2_{PC}$ *and the matrix* $\boldsymbol{A}$ *are defined as in Proposition 4.3 and the parameters* (k_1, k_2, k_3) *are given by* (4.7)–(4.9)*, which we assume exist.*

2. *The shrinkage intensities* δ_i *maximizing* $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *in* (4.26) *are*

$$\boldsymbol{\delta}^\star_{eu} = \underset{\boldsymbol{\delta}\in\mathbb{R}^N}{\operatorname{argmax}}\, \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}), \tag{4.27}$$

where each shrinkage intensity is given by

$$\delta^\star_{eu,i} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i - \frac{N}{T-2}\bar{R}\right), \tag{4.28}$$

and they satisfy

$$\delta^\star_{eu,i} \in \left[0\ , 1 + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\bar{R}\right], \tag{4.29}$$

where the lower bound is achieved as $T \to \infty$ *and the upper bound when* $\theta^2_{PC_i} = 0$*.*

3. *The EU delivered by* $\delta^\star_{eu,i}$ *is*

$$\begin{aligned}\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}^\star}) &= \frac{k_1(T-1)}{2\gamma(T-N-2)}\sum_{i=1}^{N}(1-\delta^\star_{eu,i})\theta^2_{PC_i} \\ &\geq \frac{k_1(T-1)}{2\gamma(T-N-2)}\sum_{i=1}^{N}(1-\delta^\star_{eu,cst})\theta^2_{PC_i} \geq 0,\end{aligned} \tag{4.30}$$

and the lower bound is obtained with $\delta_i = \delta^\star_{eu,cst}$ *for all* i*, where*

$$\delta^\star_{eu,cst} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)}\left(\frac{k_1\theta^2}{k_2\theta^2 + k_3 N/T}\right) \in [0,1] \tag{4.31}$$

maximizes $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *subject to the constraint* $\boldsymbol{\delta} = \delta\mathbf{1}_N$*.*

4. *Assume* $k_1/k_2 > \frac{N}{T-2}\bar{R}$*. Then,* $\delta^\star_{eu,i} < 1$ *if* $\theta^2_{PC_i} > \bar{\theta}^2_{PC}$*, and* $\delta^\star_{eu,i} \geq 1$ *otherwise, where*

$$\bar{\theta}^2_{PC} = \frac{k_3\frac{N}{T}\bar{R}}{k_1(T-2) - k_2 N\bar{R}}. \tag{4.32}$$

Proposition 4.5 shows that the EU and optimal shrinkage intensities under nonlinear shrinkage depend on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ only via the squared Sharpe ratios of PCs $\theta^2_{PC_i}$, and they also depend on the elliptical distribution via (k_1, k_2, k_3). Unlike in linear shrinkage, the optimal nonlinear shrinkage intensities $\delta^\star_{eu,i}$ can be expressed in closed form via (4.28).[11]

As in linear shrinkage, we typically expect the EU-optimal shrinkage intensities $\delta^\star_{eu,i}$ to be larger than the MSE-optimal ones $\delta^\star_{mse,i}$ because they also account for estimation errors stemming from the sample mean returns $\hat{\boldsymbol{\mu}}$. Indeed, in Section 4.A of the supplementary material, we study the setting in which $\boldsymbol{\Sigma}$ is known, in which case $\delta^\star_{mse,i} = 0$ but $\delta^\star_{eu,i} =$

[11] As in linear shrinkage, $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ is proportional to $1/\gamma$, and thus, $\delta^\star_{eu,i}$ does not depend on γ.

$(1/T)/(\theta^2_{PC_i} + 1/T)$. This is also suggested from their respective bounds, with $\delta^\star_{mse,i}$ always below one and negative for some i and $\delta^\star_{eu,i}$ always above zero and also above one for some i.

Part 3 of Proposition 4.5 shows that the EU-optimal shrinkage intensities $\delta^\star_{eu,i}$ deliver a positive EU. This contrasts with the case of linear shrinkage, in which $\delta^\star_{eu}$ can deliver a negative EU as illustrated in Figure 4.1. The reason for this is not because we use N different shrinkage intensities instead of one. Indeed, Part 3 of Proposition 4.5 shows that even if we constrain all shrinkage intensities to be equal, $\boldsymbol{\delta} = \delta \mathbf{1}_N$, the resulting optimal δ, $\delta^\star_{eu,cst}$ in (4.31), delivers a positive EU. Rather, linear shrinkage can deliver a negative EU because, by construction, it increases the first few inverse sample eigenvalues and decreases the remaining ones, whereas Proposition 4.5 shows that it is optimal to decrease all of them since $\delta^\star_{eu,i} \geq 0$.

Finally, another important difference between the MSE-optimal shrinkage intensities, $\delta^\star_{mse,i}$ in (4.24), and the EU-optimal ones, $\delta^\star_{eu,i}$ in (4.28), resides in the bounds they satisfy. We have $\delta^\star_{eu,i} \geq 0$ whereas $\delta^\star_{mse,i}$ can turn negative, which means that under the EU it is always optimal to decrease the inverse sample eigenvalues. It also holds that $\delta^\star_{mse,i} < 1$ whereas $\delta^\star_{eu,i}$ can surpass one for low-Sharpe-ratio i.e., $\theta^2_{PC_i} < \bar{\theta}^2_{PC}$ with $\bar{\theta}^2_{PC}$ defined in (4.32). In that case, the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ uses some negative eigenvalues because $\hat{\lambda}_i/(1 - \delta^\star_{eu,i}) < 0$. Intuitively, it is optimal out of sample for the mean-variance portfolio to short PCs that perform badly, i.e., have low Sharpe ratios.

Remark 4.6. One may be reluctant to using negative shrunk eigenvalues even though they are optimal under the EU objective function, because it implies losing the positive-definiteness property of the covariance matrix. Therefore, in Section 4.C of the supplementary material, we study two ways of making all shrinkage intensities below one. First, we consider the shrinkage intensities δ_i that maximize $\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ under the constraints $\delta_i \leq 1$. We find that this solution is typically close to the unconstrained one, $\delta^\star_{eu,i}$ in (4.28), because even the $\delta^\star_{eu,i}$ above one are typically close to one. Second, observing that

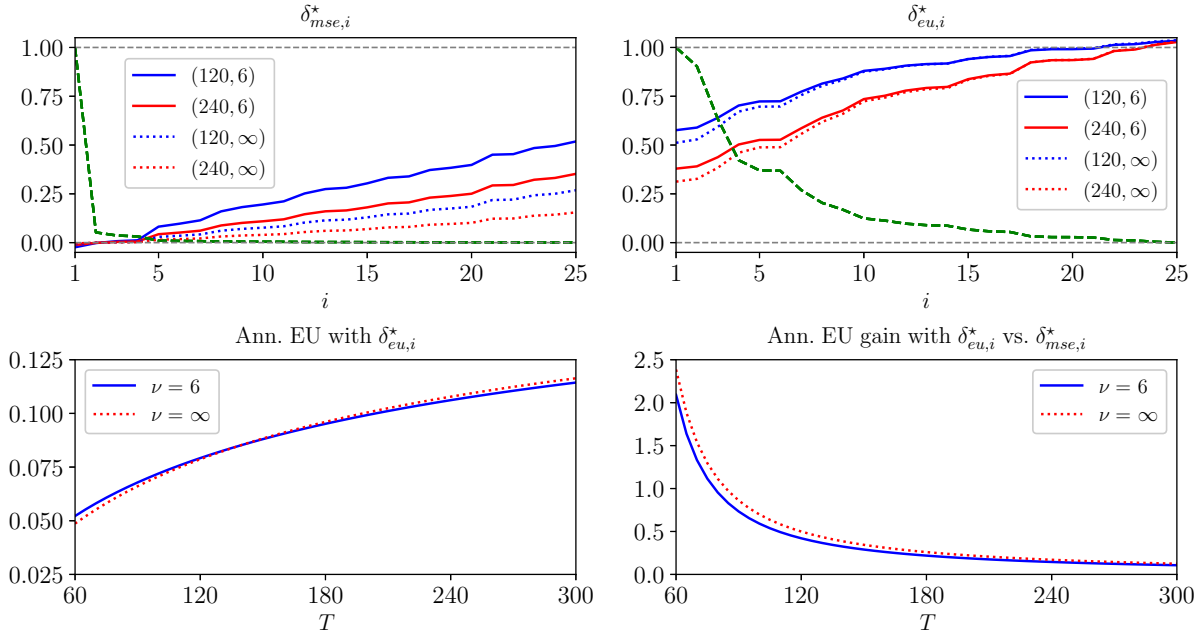
$$\sum_{i=1}^{N} \delta^\star_{eu,i} = N - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)} \sum_{i=1}^{N} R_i, \tag{4.33}$$

we consider the solution $\delta_i = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)} R_i \in [0,1]$, which satisfies $\sum_{i=1}^N \delta_i = \sum_{i=1}^N \delta^\star_{eu,i}$. We find that this solution is a dampened version of $\delta^\star_{eu,i}$, i.e., ordering the $\theta^2_{PC_i}$ in decreasing order, it shrinks more intensively for i from 1 up to some k and less intensively for i from $k+1$ up to N. The empirical performance delivered by these two alternative nonlinear shrinkage intensities is similar to that delivered by the optimal solution $\delta^\star_{eu,i}$, with the constrained intensities performing slightly better overall. Thus, our methodology can also be used to obtain a well-behaved nonlinear shrinkage covariance matrix with positive eigenvalues without sacrificing performance.

4.4.3 Comparison of nonlinear shrinkage intensities

We now compare the oracle nonlinear shrinkage intensities $\delta^\star_{mse,i}$ and $\delta^\star_{eu,i}$ given by (4.24) and (4.28), respectively. We have that $\delta^\star_{mse,i}$ depends on T, $\boldsymbol{\Lambda}$, and κ, whereas $\delta^\star_{eu,i}$ depends on N, T, $\boldsymbol{\theta}^2_{PC}$, and (k_1, k_2, k_3). We calibrate these parameters as in Figure 4.1. The top two panels of Figure 4.2 depict $\delta^\star_{mse,i}$ and $\delta^\star_{eu,i}$ for each i using $T = 120$ and 240. The $\delta^\star_{mse,i}$ are in increasing order because the eigenvalues λ_i are in decreasing order. In contrast, the $\delta^\star_{eu,i}$ are not necessarily increasing with i because the squared Sharpe ratios $\theta^2_{PC_i}$ are not monotonically decreasing with i. Therefore, for visibility, we order the $\theta^2_{PC_i}$ in decreasing order, and thus, the $\delta^\star_{eu,i}$ in increasing order. The two bottom panels of Figure 4.2 depict the annualized EU delivered by $\delta^\star_{eu,i}$ and its difference relative to that delivered by $\delta^\star_{mse,i}$.

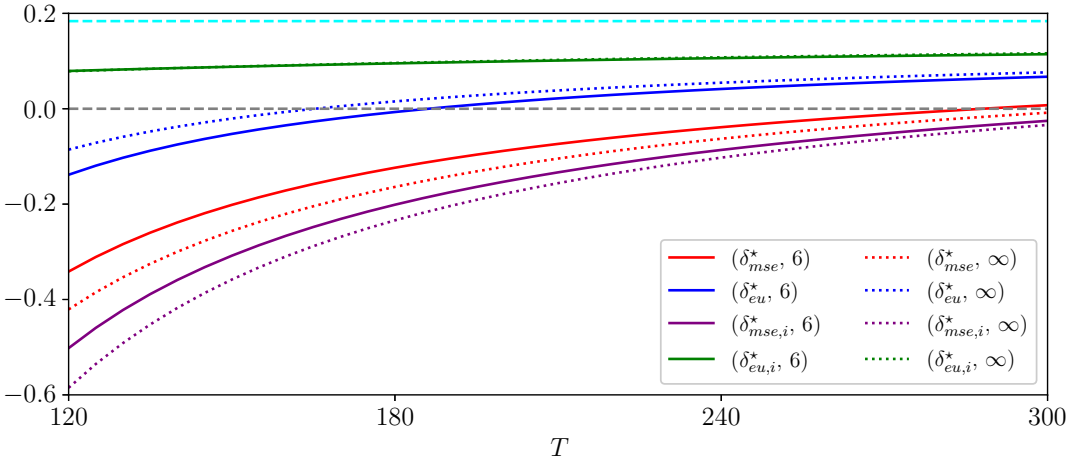
Figure 4.2: Comparison of nonlinear shrinkage intensities $\delta^\star_{eu,i}$ and $\delta^\star_{mse,i}$



Notes. The top two panels depict the nonlinear shrinkage intensities $\delta^\star_{mse,i}$ and $\delta^\star_{eu,i}$ in Equations (4.24) and (4.28), which minimize the MSE of the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.21) and maximize the EU in (4.26) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{D}}\hat{\boldsymbol{\mu}}$, respectively. We depict $\delta^\star_{mse,i}$ for the eigenvalues λ_i sorted in descending order, and $\delta^\star_{eu,i}$ for the eigenvalues λ_i sorted in descending order of their corresponding squared Sharpe ratios $\theta^2_{PC_i}$. The green dashed line depicts the ratios λ_i/λ_1 in the top left panel and $\theta^2_{PC_i}/\theta^2_{PC_1}$ in top right panel. The horizontal gray dashed lines depict a shrinkage intensity δ_i of zero and one. The asset returns follow a multivariate t-distribution with degrees of freedom $\nu = \{6, \infty\}$, and each line represents a different choice of (T, ν) with $T = \{120, 240\}$. The two bottom panels depict the annualized EU delivered by $\delta^\star_{eu,i}$ (bottom left) and its difference relative to that delivered by $\delta^\star_{mse,i}$ (bottom right), as a function of the sample size $T \in [60, 300]$. We calibrate the eigenvalues $\boldsymbol{\Lambda}$ and the PCs' squared Sharpe ratios $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024. We set a risk-aversion coefficient $\gamma = 3$.

We make three main observations. First, the top panels are in line with Propositions 4.4 and 4.5. Specifically, $\delta^\star_{mse,i}$ shrinks lower-variance PCs more intensively, $\delta^\star_{mse,i} < 0$ for the very first few eigenvalues that satisfy $\lambda_i > \bar{\lambda}$, $\delta^\star_{eu,i}$ shrinks lower-Sharpe-ratio PCs more intensively, and $\delta^\star_{eu,i} > 0$ for the PCs that satisfy $\theta^2_{PC,i} < \bar{\theta}^2_{PC}$. Second, we observe in the top panels that $\delta^\star_{eu,i}$ shrinks all inverse eigenvalues much more intensively toward

Figure 4.3: Comparison of the EUs delivered by linear and nonlinear shrinkage intensities



Notes. This figure depicts the annualized EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ under the linear shrinkage intensities $\delta^\star_{mse}$ in (4.15) (red) and $\delta^\star_{eu}$ in (4.20) (blue), for which the EU depends on δ via (4.17), and the nonlinear shrinkage intensities $\delta^\star_{mse,i}$ in (4.24) (purple) and $\delta^\star_{eu,i}$ in (4.28) (green), for which the EU depends on the δ_i's via (4.26). We depict the EU as a function of the sample $T \in [120, 300]$. The horizontal dashed cyan line corresponds to the maximum achievable annualized utility with no parameter uncertainty, $12U(\boldsymbol{w}^\star)$ in (4.3). The asset returns follow a multivariate t-distribution with degrees of freedom $\nu \in \{6, \infty\}$. Each line corresponds to a choice of shrinkage intensity and ν. We calibrate the eigenvalues $\boldsymbol{\Lambda}$ and PCs' squared Sharpe ratios $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024. We set a risk-aversion coefficient $\gamma = 3$.

zero compared to $\delta^\star_{mse,i}$, which is because the EU criterion accounts for the impact of estimation errors stemming from both the mean and covariance matrix. For example, when $T = 120$ and $\nu = 6$, we have for all i that $\delta^\star_{mse,i} < 0.52$ whereas $\delta^\star_{eu,i} > 0.59$. Third, the bottom left panel shows that $\delta^\star_{eu,i}$ delivers a positive EU for any sample size T, in line with Proposition 4.5, which is a significant difference relative to linear shrinkage in Figure 4.1. This EU delivered by $\delta^\star_{eu,i}$ is significantly larger than that delivered by $\delta^\star_{mse,i}$, as shown in the bottom right panel. For example, when $T = 120$, the annualized EU gain is 0.42 for $\nu = 6$ and 0.50 for $\nu = \infty$.

4.4.4 Comparison of linear and nonlinear shrinkage

In this section, we compare the EU delivered by the linear and nonlinear shrinkage intensities. Figure 4.3 depicts the annualized EU delivered by the linear shrinkage intensities $\delta^\star_{mse}$ in (4.15) and $\delta^\star_{eu}$ in (4.20), for which the EU depends on δ via (4.17), and by the nonlinear shrinkage intensities $\delta^\star_{mse,i}$ in (4.24) and $\delta^\star_{eu,i}$ in (4.28), for which the EU depends on the δ_i's via (4.26). We calibrate the required parameters as in Figures 4.1 and 4.2. For visibility, we start at $T = 120$.

Figure 4.3 shows that the MSE-minimizing shrinkage intensities $\delta^\star_{mse}$ and $\delta^\star_{mse,i}$ struggle to deliver a positive EU. In particular, $\delta^\star_{mse,i}$ performs worst and delivers a negative EU for all sample sizes T considered. This result might seem surprising given that $\delta^\star_{mse,i}$ delivers a lower MSE than $\delta^\star_{mse}$ as it uses N shrinkage intensities instead of a single one, which

shows that a lower MSE does not necessarily mean a higher EU. In contrast, $\delta^{\star}_{eu}$ delivers a positive EU for realistic values of T, and $\delta^{\star}_{eu,i}$ systematically delivers a positive EU, substantially improving upon all other methods and approaching the maximum achievable utility with no parameter uncertainty, $12U(\boldsymbol{w}^{\star})$ in (4.3). This highlights the importance of calibrating the shrinkage intensities to a portfolio-motivated objective function over the generic MSE.

4.5 Estimation of oracle shrinkage intensities

In this section, we explain how we estimate the MSE-optimal shrinkage intensities, $\delta^{\star}_{mse}$ in (4.15) and $\delta^{\star}_{mse,i}$ in (4.24) for linear and nonlinear shrinkage, respectively, as well as the EU-optimal shrinkage intensities, $\delta^{\star}_{eu}$ in (4.20) and $\delta^{\star}_{eu,i}$ in (4.28) for linear and nonlinear shrinkage, respectively, which are oracles because they depend on the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Starting with $\delta^{\star}_{mse}$ and $\delta^{\star}_{mse,i}$, they require estimating the kurtosis parameter κ, the sphericity parameter s, and the eigenvalues λ_i. We estimate κ with the sample unbiased estimator in Ollila and Raninen (2019, Equation (24)), i.e.,

$$\hat{\kappa} = \frac{1}{3N}\sum_{i=1}^{N}\hat{K}_i, \quad \hat{K}_i = \frac{(T-1)[(T+1)\hat{k}_i + 6]}{(T-2)(T-3)}, \tag{4.34}$$

where $\hat{K}_i$ is the sample unbiased estimator of the excess kurtosis of asset i with $\hat{k}_i = m_4(\boldsymbol{r}_i)/m_2(\boldsymbol{r}_i)^2 - 3$ and $m_k(\boldsymbol{r}_i) = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i)^k$.

Then, we estimate the sphericity parameter s using the asymptotically unbiased estimator in Ollila and Raninen (2019, Equation (27)), i.e.,

$$\hat{s} = \frac{1}{N-1}\left[\frac{(\hat{\kappa}+T)(T-1)^2}{(T-2)(3\hat{\kappa}(T-1)+T(T+1))}\left(\frac{N\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^2)}{\mathrm{tr}(\hat{\boldsymbol{\Lambda}})^2} - \frac{N(\hat{\kappa}+\frac{T}{T-1})}{\hat{\kappa}+T}\right) - 1\right], \tag{4.35}$$

which we truncate in $[0,1]$ because the population s belongs to that interval.

Finally, we estimate the eigenvalues λ_i with the MSE-optimal shrinkage estimator, i.e.,

$$\hat{\lambda}^{\star}_{i,mse} = (1-\hat{\delta}^{\star}_{mse})\hat{\lambda}_i + \hat{\delta}^{\star}_{mse}\hat{\bar{\lambda}}, \tag{4.36}$$

where $\hat{\delta}^{\star}_{mse}$ is the estimate of $\delta^{\star}_{mse}$ obtained from $\hat{\kappa}$ and $\hat{s}$ above.

Turning to $\delta^{\star}_{eu}$ and $\delta^{\star}_{eu,i}$, we need to estimate, in addition to the eigenvalues λ_i, the squared Sharpe ratios of PCs $\theta^2_{PC_i}$ and the fat-tails parameters (k_1, k_2, k_3). We estimate the the squared Sharpe ratios of PCs $\theta^2_{PC_i}$ in a structured way to alleviate their sensitivity to estimation errors stemming from the sample mean returns $\hat{\boldsymbol{\mu}}$.[12] Specifically, following the asset-pricing rationale of Kozak et al. (2020) that low-variance PCs with high Sharpe

[12]In unreported results, we find that using the sample estimates of $\theta^2_{PC_i}$ instead of the structured approach we propose results in a small performance loss, which does not affect the overall empirical conclusions.

ratios are unlikely to persist out of sample,[13] we model $\theta^2_{PC_i}$ as decreasing with the index i. Specifically, we model $\theta^2_{PC_i}$ as being geometrically decreasing with i, i.e., $\theta^2_{PC_{i+1}} = g \times \theta^2_{PC_i}$ with $g \in [0,1]$, which implies that $\theta^2_{PC_i} = g^{i-1}\theta^2_{PC_1}$. In that case, the value of g is determined the squared Sharpe ratio of the first PC, $\theta^2_{PC_1}$, and the sum of all squared Sharpe ratios, $\theta^2 = \sum_{i=1}^N \theta^2_{PC_i}$ in (4.3). Indeed, $\theta^2 = \sum_{i=1}^N \theta^2_{PC_i} = \sum_{i=1}^N g^{i-1}\theta^2_{PC_1}$ implies the relation

$$\theta^2_{PC_1}\frac{1-g^N}{1-g} = \theta^2. \tag{4.37}$$

This means that, under this model, estimating all $\theta^2_{PC_i}$'s only requires estimating θ^2 and $\theta^2_{PC_1}$. We estimate θ^2 with the popular adjusted estimator of Kan and Zhou (2007), i.e.,

$$\hat{\theta}^2_a = \frac{(T-N-2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{N/2}(1+\hat{\theta}^2)^{-(T-2)/2}}{T \times B_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2,(T-N)/2)}, \tag{4.38}$$

where $\hat{\theta}^2 = \frac{T}{T-1}\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$ and $B_x(a,b) = \int_0^x y^{a-1}(1-y)^{b-1}\mathrm{d}y$ is the incomplete beta function. We use a similar adjusted estimator for $\theta^2_{PC_1}$, giving us

$$\hat{\theta}^2_{a,PC_1} = \frac{(T-3)\hat{\theta}^2_{PC_1} - 1}{T} + \frac{2(\hat{\theta}^2_{PC_1})^{1/2}(1+\hat{\theta}^2_{PC_1})^{-(T-2)/2}}{T \times B_{\hat{\theta}^2_{PC_1}/(1+\hat{\theta}^2_{PC_1})}(1/2,(T-1)/2)}, \tag{4.39}$$

where $\hat{\theta}^2_{PC_1} = \frac{T}{T-1}(\hat{\boldsymbol{v}}_1'\hat{\boldsymbol{\mu}})^2/\hat{\lambda}_1$.

Finally, we estimate (k_1,k_2,k_3) in a fast and accurate way that avoids relying on Monte Carlo simulations to evaluate the expectations in (4.7)–(4.9) by relying on their high-dimensional asymptotic limit. Specifically, using results in El Karoui (2010, 2013) and Kan and Lassance (2025), we have that as $N \to \infty$, $T \to \infty$, and $N/T \to \phi \in (0,1)$, the asymptotic limit is $(k_1,k_2,k_3) \to (\eta,\varphi,\eta)$, where $\eta \geq 1$ is the unique positive solution to

$$\mathbb{E}\left[(1-\phi+\phi\eta\tau)^{-1}\right] = 1, \tag{4.40}$$

where the random variable τ determines the elliptical distribution in (4.6), and

$$\varphi = (1-\phi)\left(\eta^{-2} - \mathbb{E}\left[\frac{\phi\tau^2}{(1-\phi+\phi\eta\tau)^2}\right]\right)^{-1} \geq \eta^2. \tag{4.41}$$

Then, to estimate η and φ, we use the sample estimate of the distribution of τ proposed by El Karoui (2010, 2013), i.e.,

$$\hat{\tau}_t = \frac{(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})'(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})}{\frac{1}{T}\sum_{i=1}^T (\boldsymbol{r}_i - \hat{\boldsymbol{\mu}})'(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}})}, \quad t = 1,\ldots,T, \tag{4.42}$$

which is consistent as $N \to \infty$. The resulting sample estimate of $\hat{\eta}$ is the unique positive solution to

$$\frac{1}{T}\sum_{t=1}^T (1-\phi+\phi\hat{\eta}\hat{\tau}_t)^{-1} = 1, \tag{4.43}$$

[13] Kozak et al. (2020, p.277) argue that "[...] Sharpe ratios of low-eigenvalue PCs should be smaller than those of the high-eigenvalue PCs that are the major source of risk premia."

Table 4.1: List of datasets considered in the empirical analysis

#	Name	Description	N	T_{tot}	Time period
1	25SBM	25 portfolios sorted on size and book-to-market	25	1177	07/1926–07/2024
2	10MOM	10 portfolios sorted on momentum	10	1171	01/1927–07/2024
3	25OPIN	25 portfolios sorted on operating profitability and investment	25	733	07/1963–07/2024
4	13JKP	13 characteristic-managed portfolios in Jensen et al. (2023)	13	866	11/1951–12/2023
5	16ANOM	16 characteristic portfolios built on long and short legs of eight low-turnover characteristics in Novy-Marx and Velikov (2015)	16	486	07/1973–12/2013
6	46ANOM	46 characteristic portfolios built on long and short legs of 23 characteristics in Novy-Marx and Velikov (2015)	46	606	07/1963–12/2013

Notes. This table lists the datasets of monthly excess returns considered in the empirical analysis of Section 4.6. The columns report, in order, the dataset number, the dataset name, the dataset description, the number of assets N, the total number of observations T_{tot}, and the time period. The first three datasets are collected from Kenneth French's website. The fourth dataset is collected from the website https://jkpfactors.com. The last two datasets are collected from Robert Novy-Marx's website.

and the sample estimate of φ is

$$\hat{\varphi} = (1-\phi)\left(\hat{\eta}^{-2} - \frac{1}{T}\sum_{t=1}^{T}\frac{\phi\hat{\tau}_t^2}{(1-\phi+\phi\hat{\eta}\hat{\tau}_t)^2}\right)^{-1}. \tag{4.44}$$

4.6 Empirical analysis

We now evaluate empirically the portfolio performance obtained with our shrinkage covariance matrix estimators. In Section 4.6.1, we present the six datasets we use. In Section 4.6.2, we detail the different shrinkage covariance estimators we consider as well as the benchmark portfolio strategies. In Section 4.6.3, we explain the out-of-sample methodology. Finally, we report and discuss the results in Section 4.6.4.

4.6.1 Datasets

Table 4.1 lists the six datasets of monthly excess returns we consider in this empirical analysis. The first three datasets are characteristic-sorted portfolios collected from Kenneth French's website. Specifically, the first dataset is composed of 25 quintile portfolios sorted on size and book-to-market (25SBM), the second dataset is composed of decile portfolios sorted on momentum (10MOM), and the third is composed of 25 quintile portfolios sorted on operating profitability and investment (25OPIN). The fourth dataset is composed of 13 characteristic-managed portfolios (13JKP) based on the construction in Jensen et al. (2023). The last two datasets are composed of the long and short legs obtained from the eight low-turnover characteristics (16ANOM) and the 23 characteristics (46ANOM) in Novy-Marx and Velikov (2015), which are available on Robert Novy-Marx's website.

4.6.2 Portfolio strategies

We consider 21 portfolio strategies in total, which we list in Table 4.2. The first nine are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using rotation-equivariant covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}'$ for different choices of $\boldsymbol{D}$. First, the linear shrinkage estimator maximizing the EU in Proposition 4.3 (EUL). Second, the nonlinear shrinkage estimator maximizing the EU in Proposition 4.5 (EUNL). Third, the linear shrinkage estimator minimizing the MSE in Proposition 4.2 (MSEL). Fourth, the nonlinear shrinkage estimator minimizing the MSE in Proposition 4.4 (MSENL). In Section 4.5, we explain how we estimate the parameters in these first four estimators.[14] Fifth, the sample covariance matrix in (4.5) (SAMPLE). Sixth, the linear shrinkage estimator minimizing the MSE in Ledoit and Wolf (2004) (LW2004). Seventh, the nonlinear shrinkage estimator minimizing the MSE in Ledoit and Wolf (2020) (LW2020).

The eighth covariance matrix estimator is the linear shrinkage estimator minimizing the MSE in Bodnar et al. (2014) (BGP). Their shrinkage estimator considers a general target matrix $\boldsymbol{\Sigma}_0$, which when choosing $\boldsymbol{\Sigma}_0 = \hat{\bar{\lambda}}\boldsymbol{I}_N$ and applying their Equations (4.3)–(4.4) gives

$$\hat{\boldsymbol{\Sigma}}_{BGP} = \hat{\delta}_{BGP}\hat{\boldsymbol{\Sigma}} + (1-\hat{\delta}_{BGP})\hat{\bar{\lambda}}\boldsymbol{I}_N, \quad \text{with} \quad \hat{\delta}_{BGP} = 1 - \frac{N/T}{(N-1)\hat{s}_p}, \tag{4.45}$$

where $\hat{s}_p = \big(\frac{N\text{tr}(\hat{\boldsymbol{\Lambda}}^2)}{\text{tr}(\hat{\boldsymbol{\Lambda}})^2} - 1\big)/(N-1)$ is the sample plug-in estimator of the sphericity parameter s.

The ninth shrinkage covariance matrix estimator uses PCA to keep only the first $K < N$ eigenvalues of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$, as in Chen and Yuan (2016). Specifically, let $\hat{\boldsymbol{V}}_K$ be the $N \times K$ matrix whose columns are the first K eigenvectors of $\hat{\boldsymbol{\Sigma}}$ and let $\hat{\boldsymbol{\Lambda}}_K$ be the $K \times K$ diagonal matrix of the corresponding eigenvalues sorted in decreasing order. Then, we construct the PCA mean-variance portfolio as $\hat{\boldsymbol{w}} = \frac{1}{\gamma}\hat{\boldsymbol{V}}_{\hat{K}}\hat{\boldsymbol{\Lambda}}_{\hat{K}}^{-1}\hat{\boldsymbol{V}}_{\hat{K}}'\hat{\boldsymbol{\mu}}$, where the estimator $\hat{K}$ is that in Bai and Ng (2002), i.e.,

$$\hat{K} = \underset{K\in[1,N-1]}{\text{argmin}} \ \ln\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}^2\right) + K\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

where $\boldsymbol{\epsilon} = (\boldsymbol{I}_N - \hat{\boldsymbol{V}}_K\hat{\boldsymbol{V}}_K')\boldsymbol{R}$ with $\boldsymbol{R}$ the $N \times T$ matrix of return samples.[15]

The next three portfolio strategies listed in Table 4.2 are the global-minimum-variance (GMV) portfolio, $\hat{\boldsymbol{w}}_g = (\boldsymbol{1}_N'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{1}_N)^{-1}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{1}_N$, where $\hat{\boldsymbol{\Sigma}}$ is either the sample estimator, the LW2004 estimator, and the cross-validated maximum-likelihood estimator of Shi et al.

[14] In Section 4.F of the supplementary material, we compare the empirical performance of the EUL, EUNL, MSEL, and MSENL portfolio strategies when the shrinkage intensities are estimated under the multivariate normal versus elliptical distribution. We find that whereas MSEL and MSENL perform substantially worse under the multivariate normal distribution, EUL and EUNL are much more robust to the distributional assumption as they perform similarly overall under the two distributions.

[15] The PCA estimator amounts to taking $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1} = \hat{\boldsymbol{V}}\boldsymbol{D}^{-1}\hat{\boldsymbol{V}}'$ with $\boldsymbol{D}^{-1} = \text{diag}\big(1/\hat{\lambda}_1, \ldots, 1/\hat{\lambda}_{\hat{K}}, \boldsymbol{0}_{N-K}\big)$.

Table 4.2: List of portfolio strategies considered in the empirical analysis

#	Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	Description
1	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with linear shrinkage estimator maximizing the EU (Proposition 4.3)
2	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with nonlinear shrinkage estimator maximizing the EU (Proposition 4.5)
3	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with linear shrinkage estimator minimizing the MSE (Proposition 4.2)
4	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with nonlinear shrinkage estimator minimizing the MSE (Proposition 4.4)
5	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with sample estimator in (4.5)
6	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with linear shrinkage estimator of Ledoit and Wolf (2004)
7	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with nonlinear shrinkage estimator of Ledoit and Wolf (2020)
8	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with linear shrinkage estimator of Bodnar et al. (2014)
9	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) with PCA estimator of Chen and Yuan (2016)
10	GMV	SAMPLE	GMV portfolio with sample estimator in (4.5)
11	GMV	LW2004	GMV portfolio with linear shrinkage estimator of Ledoit and Wolf (2004)
12	GMV	SSHH	GMV portfolio with linear shrinkage estimator of Shi et al. (2025)
13	GMVRF	SAMPLE	GMVRF portfolio in (4.46) with sample estimator in (4.5)
14	GMVRF	LW2004	GMVRF portfolio in (4.46) with linear shrinkage estimator of Ledoit and Wolf (2004)
15	GMVRF	SSHH	GMVRF portfolio in (4.46) with linear shrinkage estimator of Shi et al. (2025)
16	EW	/	Equally-weighted portfolio
17	EWRF	/	Optimal combination of the EW portfolio with the risk-free asset in (4.47)
18	MAXSER	/	MAXSER portfolio of Ao et al. (2019) adapted to maximum-utility objective
19	UPSA	/	Universal portfolio shrinkage estimator of Kelly et al. (2023)
20	KNS	/	Norm-constrained PCA-based mean-variance portfolio of Kozak et al. (2020)
21	F+A	/	Factor-plus-alpha strategy based on PCA and arbitrage portfolio of Da et al. (2024)

Notes. This table lists the 21 portfolio strategies considered in the empirical analysis of Section 4.6. The first two portfolios, EUL and EUNL, are the strategies introduced in this chapter. The first nine portfolios are of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13) using different rotation-equivariant estimators of the covariance matrix of the form $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.11) for different choices of the diagonal matrix $\boldsymbol{D}$. More details about the 21 portfolio strategies are available in Section 4.6.2.

(2025) (SSHH) using the diagonal target matrix $\hat{\bar{\lambda}}\boldsymbol{I}_N$ and 10 folds.[16] The following three portfolio strategies consist of the optimal combination of the GMV portfolio with the risk-free asset (GMVRF) maximizing the EU in Kan and Lassance (2025, Section VI.C.2.),

$$\hat{\boldsymbol{w}}_{gmvrf} = \frac{\hat{\eta}(T-N-1)(T-N-4)}{\gamma\hat{\varphi}T(T-2)}\left(\frac{\mathbf{1}_N'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_N'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}\right)\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N, \tag{4.46}$$

where $\hat{\boldsymbol{\Sigma}}$ is either of the same three estimators as for the GMV portfolio, and $\hat{\eta}$ and $\hat{\varphi}$ are obtained following Section 4.5.[17]

Two other portfolio strategies we consider as benchmarks are the equally weighted portfolio (EW), $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$, and the optimal combination of the EW portfolio with the risk-free asset that maximizes the EU (EWRF) considered in, e.g., Lassance et al. (2024b),

$$\hat{\boldsymbol{w}}_{ewrf} = \frac{T-3}{\gamma T}\left(\frac{\boldsymbol{w}_{ew}'\hat{\boldsymbol{\mu}}}{\boldsymbol{w}_{ew}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}_{ew}}\right)\boldsymbol{w}_{ew}, \tag{4.47}$$

[16]We implement the SSHH covariance matrix estimator using R codes available here: `data.mendeley.com/datasets/vm4wj9mycr/1`.

[17]We also implement the GMV and GMVRF portfolios with the nonlinear shrinkage covariance matrix of Ledoit and Wolf (2020) and obtain similar results, which we do not report for brevity.

where $\hat{\mathbf{\Sigma}}$ is the sample covariance matrix.

Finally, we consider four recent proposals for estimating the mean-variance portfolio using regularization, PCA, high-dimensional, and cross-validation methods. First, the sparse MAXSER portfolio of Ao et al. (2019). While the original MAXSER portfolio is designed to target a given mean return or volatility, we adapt it to the mean-variance utility objective following Cai et al. (2024, Section A).[18] Second, the universal portfolio shrinkage estimator of Kelly et al. (2023) (UPSA), which shrinks the sample eigenvalues nonlinearly to maximize the leave-one-out EU.[19] We adapt UPSA to any risk-aversion coefficient γ instead of only $\gamma = 1$. Third, the mean-variance portfolio of Kozak et al. (2020) (KNS) that optimally weighs the PCs under elastic net regularization using cross-validation. We use five folds and we adapt their method to any risk-aversion coefficient γ instead of only $\gamma = 1$. Fourth, we implement a factor-plus-alpha (F+A) strategy as in Lassance et al. (2025a, Section F), which optimally combines the mean-variance portfolio of the first K PCs with its orthogonal arbitrage portfolio estimated with the empirical Bayes method of Da et al. (2024), where K is estimated as in Bai and Ng (2002).

4.6.3 Out-of-sample methodology

We measure portfolio performance using rolling windows. At the end of month t, we estimate portfolio k using the T previous months, and we compute its out-of-sample return in month $t+1$. We consider $T = 120$ and 240 months. We repeat this process iteratively, giving us a time series of $T_{tot} - T$ out-of-sample gross portfolio returns $r_{gross,k,t}$, where T_{tot} is the total number of months in the dataset. We then compute the net out-of-sample returns, $r_{net,k,t} = r_{gross,k,t}$ if $t = T + 1$ and

$$r_{net,k,t} = (1 + r_{gross,k,t})\,(1 - p \times \text{turnover}_{k,t-1}) - 1 \quad \text{if} \quad t = T+2, \ldots, T_{tot}, \tag{4.48}$$

where p is the proportional cost required to rebalance the portfolio and

$$\text{turnover}_{k,t} = \sum_{i=1}^{N} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, T_{tot}, \tag{4.49}$$

with $w_{i,k,t}$ the weight of asset i in month t and $w_{i,k,(t-1)^+}$ the prior-month weight before rebalancing in month t. We set $p = 10$ basis points in line with the average bid-ask spreads reported by Engle et al. (2012) and Frazzini et al. (2018). Finally, we compare the portfolio strategies in terms of *annualized out-of-sample utility net of transaction costs*,

$$U_k = 12 \times \left(\hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2\right), \tag{4.50}$$

[18] We thank the authors for sharing their code with us.

[19] We implement the UPSA strategy of Kelly et al. (2023) using Python codes available here: `github.com/pourmohammadimohammad/Universal_Portfolio_Shrinkage/tree/main`.

where $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the sample mean and variance of $r_{net,k,t}$. In Sections 4.D and 4.E of the supplementary material, we also report the portfolio turnover and the net out-of-sample Sharpe ratio of the different portfolio strategies, respectively.[20]

We compute two-sided p-values for the statistical test of the difference between the net utility of two pairs of strategies: EUL vs. MSEL and EUNL vs. MSENL. To compute the p-values, we generate 10,000 bootstrap samples using the stationary block bootstrap approach of Politis and Romano (1994) with an average block size of five, and then use the methodology of Ledoit and Wolf (2008, Remark 3.2) to produce the resulting p-values. We use the symbols ○, ◑, and ● to indicate that the p-value is less than 10%, 5%, and 1%, respectively.

4.6.4 Results

In Table 4.3, we report the annualized net out-of-sample utility, in percentage points, of the 21 portfolio strategies listed in Table 4.2 across the six datasets listed in Table 4.1, following the methodology detailed in Section 4.6.3. We report results for $\gamma = 3$ and 10.

Our main objective is to assess the empirical performance gain of our proposed shrinkage portfolios, EUL and EUNL, relative to their MSE-based counterparts, MSEL and MSENL. Therefore, we start with this comparison. Concerning linear shrinkage, EUL systematically outperforms MSEL, which often performs poorly. On average across datasets and given $\gamma = 3$, EUL delivers a net utility of 20.44 and 15.71 percentage points for $T = 120$ and 240, respectively, versus -19.72 and -29.68 for the MSEL strategy. Moreover, EUL delivers a negative net utility in only one configuration (25OPIN dataset with $T = 120$), whereas it happens in half of the cases for MSEL. The net EU gain delivered by EUL versus MSEL is statistically significant at the 1% (10%) level in 71% (83%) of the cases.

Turning to nonlinear shrinkage, EUNL outperforms MSENL in 75% of the tested configurations. On average across datasets and given $\gamma = 3$, EUNL delivers a net utility of 9.60 and 9.37 percentage points for $T = 120$ and 240, respectively, versus -4.74 and -17.36 for the MSENL strategy. Even when EUNL does not outperform MSENL, the difference is typically small and is never statistically significant. Another striking observation is that, consistent with the theoretical prediction in Proposition 4.5, EUNL systematically delivers a positive net utility, whereas MSENL delivers a negative net utility in 50% of the cases. Among the 21 competitors, EUNL is the only considered strategy, along with GMVRF, to achieve this feat.

[20] We do not especially expect our proposed shrinkage portfolio strategies, EUL and EUNL, to outperform in terms of Sharpe ratio, because it is not the objective function underlying their construction. And indeed, we observe in particular from Table 4.3 of the supplementary material that EUL and EUNL have a comparable, but not necessarily better, out-of-sample Sharpe ratio than MSEL and MSENL. Therefore, a relevant extension for future research is to derive the optimal linear and nonlinear shrinkage of the sample covariance matrix that maximizes the expected out-of-sample Sharpe ratio of the mean-variance portfolio.

Table 4.3: Annualized net out-of-sample utility in empirical data (in percentage points)

Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	$T = 120$						$T = 240$					
		Dataset						Dataset					
		25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13)	EUL	9.12$^{\bullet}$	10.1$^{\bullet}$	−4.58$^{\bullet}$	42.0$^{◑}$	**4.59**$^{\bullet}$	**61.4**$^{\bullet}$	11.9$^{\bullet}$	10.4$^{○}$	3.58$^{\bullet}$	55.2	**7.92**	5.28$^{\bullet}$
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	**9.36**$^{○}$	**11.0**	3.42$^{\bullet}$	20.5	2.82	10.5	**12.2**	**11.0**	6.88$^{◑}$	16.6	3.24	6.28$^{\bullet}$
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	−8.75	4.00	−49.1	16.3	−8.89	−71.9	0.19	6.66	−19.5	48.8	2.75	−217
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	−1.93	6.12	−37.9	24.3	−5.30	−13.7	4.38	7.93	−15.5	53.3	4.73	−159
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	−109	−14.9	−126	−207	−48.0	−1467	−29.7	0.35	−34.4	−26.3	−11.4	−610
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	−8.68	4.33	−48.0	24.7	−8.67	−74.0	0.58	6.83	−18.9	54.0	3.29	−215
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	−45.5	−7.32	−61.8	−106	−28.8	−338	−17.9	2.02	−21.0	−5.72	−6.03	−375
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	−40.6	−3.82	−84.6	−32.7	−22.1	−411	−15.9	3.11	−28.7	17.0	−4.56	−419
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	1.90	7.79	2.00	15.2	−0.88	3.22	0.24	8.31	3.11	6.27	−2.28	−8.64
GMV	SAMPLE	6.15	5.42	4.61	−2.74	0.32	−1.47	7.63	6.42	8.51	−2.67	4.68	6.60
GMV	LW2004	6.59	5.66	5.69	−2.42	1.20	0.81	7.25	6.51	8.54	−2.47	5.34	6.64
GMV	SSHH	6.52	5.54	**5.75**	−2.66	0.77	−0.14	7.48	6.46	**8.56**	−2.63	4.99	6.41
GMVRF	SAMPLE	7.59	4.48	2.74	51.3	0.48	2.07	7.28	2.54	7.41	72.1	1.76	1.19
GMVRF	LW2004	6.60	4.45	3.70	**65.0**	0.58	2.17	6.35	2.74	7.57	**74.7**	1.87	2.37
GMVRF	SSHH	7.49	4.47	3.79	61.7	0.60	2.58	7.11	2.70	7.61	74.1	1.88	1.95
EW	/	4.65	3.48	5.18	−1.93	−2.50	−2.55	4.73	3.98	5.60	−1.95	−1.30	−0.65
EWRF	/	2.91	2.31	1.39	17.5	−2.31	−2.38	2.12	1.20	3.84	1.03	−0.24	−0.32
MAXSER	/	4.19	5.78	−2.27	−65.8	−0.93	−75.8	5.32	7.82	−0.85	1.21	−1.08	−176
UPSA	/	3.00	−0.22	−38.1	46.9	−4.11	50.7	11.8	8.31	−1.83	61.3	0.15	**16.2**
KNS	/	2.55	4.51	−3.07	−3.76	−0.56	28.9	4.64	8.55	3.73	33.9	7.32	−59.4
F+A	/	−20.5	−5.03	−44.3	−142	−13.6	−28.1	−4.89	4.87	−11.9	−9.99	6.57	−123
Panel (b): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13)	EUL	2.73$^{\bullet}$	3.03$^{\bullet}$	−1.39$^{\bullet}$	12.5$^{\bullet}$	**1.38**$^{\bullet}$	**18.4**$^{\bullet}$	3.58$^{\bullet}$	3.12$^{○}$	1.07$^{\bullet}$	16.5	**2.38**	1.51$^{\bullet}$
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	**2.80**$^{○}$	**3.31**	1.03$^{\bullet}$	6.15	0.85	3.15	**3.66**	**3.28**	2.06$^{\bullet}$	4.99	0.97	1.89$^{\bullet}$
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	−2.72	1.18	−14.9	4.37	−2.69	−23.1	0.01	1.99	−5.90	14.3	0.81	−66.7
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	−0.64	1.82	−11.5	6.86	−1.60	−4.98	1.28	2.37	−4.70	15.7	1.41	−48.9
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	−33.5	−4.53	−38.2	−66.0	−14.6	−488	−9.10	0.09	−10.4	−8.73	−3.45	−190
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	−2.69	1.29	−14.5	6.94	−2.62	−23.7	0.13	2.04	−5.72	15.9	0.98	−66.3
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	−14.0	−2.24	−18.7	−34.0	−8.71	−107	−5.50	0.59	−6.35	−2.38	−1.84	−116
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	−12.5	−1.18	−25.6	−10.8	−6.69	−130	−4.89	0.92	−8.66	4.64	−1.39	−130
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	0.57	2.34	0.60	4.47	−0.26	0.96	0.07	2.49	0.93	1.82	−0.68	−2.61
GMV	SAMPLE	−0.66	−2.56	−2.83	−2.79	−6.52	−7.78	1.77	−0.56	1.60	−2.73	−1.14	0.61
GMV	LW2004	0.62	−1.81	−1.01	−2.48	−5.11	−3.83	1.62	−0.24	1.96	−2.54	−0.30	2.13
GMV	SSHH	0.45	−2.16	−0.92	−2.71	−5.73	−4.82	1.80	−0.42	2.04	−2.69	−0.77	1.77
GMVRF	SAMPLE	2.28	1.34	0.82	15.1	0.14	0.62	2.18	0.76	2.22	21.5	0.53	0.36
GMVRF	LW2004	1.98	1.34	1.11	**19.5**	0.17	0.65	1.90	0.82	2.27	**22.4**	0.56	0.71
GMVRF	SSHH	2.25	1.34	**1.14**	18.3	0.18	0.78	2.13	0.81	**2.28**	22.1	0.56	0.59
EW	/	−8.45	−7.18	−4.23	−2.09	−14.0	−14.2	−5.76	−5.01	−3.19	−2.12	−11.8	−12.7
EWRF	/	0.87	0.69	0.42	5.21	−0.69	−0.71	0.63	0.36	1.15	0.24	−0.07	−0.10
MAXSER	/	0.82	1.80	−1.03	−18.0	−0.22	−21.8	1.39	2.24	−0.64	0.35	−0.52	−53.0
UPSA	/	0.82	−0.09	−11.5	13.9	−1.26	15.1	3.54	2.49	−0.57	18.4	0.03	**4.76**
KNS	/	0.70	1.32	−0.94	−3.17	−0.18	7.74	1.34	2.55	1.12	9.68	2.19	−18.6
F+A	/	−6.32	−1.55	−13.4	−45.3	−4.15	−9.39	−1.50	1.45	−3.62	−3.62	1.96	−38.1

Notes. This table reports the annualized net-of-cost out-of-sample utility, in percentage points, for the 21 portfolio strategies described in Section 4.6.2 and listed in Table 4.2. The first nine of them are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using rotation-equivariant covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}'$ for different choices of the diagonal matrix $\boldsymbol{D}$. EUL and EUNL are the linear and nonlinear shrinkage estimators maximizing the EU introduced in this chapter. We report results across the six datasets described in Table 4.1 and we follow the out-of-sample methodology detailed in Section 4.6.3. The net out-of-sample utility is computed using rolling windows with sample sizes $T = 120$ and $T = 240$ months and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 3$ in Panel (a) and $\gamma = 10$ in Panel (b). We compute two-sided p-values for the statistical test of the difference between the utility of two pairs of strategies: EUL vs. MSEL and EUNL vs. MSENL. We generate 10,000 bootstrap samples using the stationary block bootstrap approach of Politis and Romano (1994) with an average block size of five, and then use the methodology of Ledoit and Wolf (2008, Remark 3.2) to produce the resulting p-values. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively. We depict the best strategy for each dataset, sample size, and γ in bold, and we underline the second-best strategy.

These results support the theoretical evidence presented in Sections 4.3 and 4.4 and indicate the substantial benefits of calibrating shrinkage covariance matrix estimators with a portfolio-specific loss function, in our case the EU, instead of the generic MSE statistical loss function.

Comparing EUL and EUNL, we find no systematic evidence in favor of linear or nonlinear shrinkage. Specifically, EUNL outperforms EUL for half of the six datasets when $T = 120$ and four datasets when $T = 240$. This is at odds with what we obtain under the oracle shrinkage intensities in Figure 4.3 where EUNL substantially improves upon EUL. This difference can be explained because nonlinear shrinkage involves N different shrinkage intensities versus only one for linear shrinkage, and thus, suffers more from estimation errors in these intensities. However, one important benefit of EUNL is that it can shrink the large inverse eigenvalues more intensively than EUL, which makes the portfolio weights less extreme and reduces turnover. In particular, Table 4.5 of the supplementary material shows that EUNL has a turnover 33.74% and 16.34% smaller than EUL for $T = 120$ and 240, respectively.

Having addressed our main comparison between EUL, EUNL, MSEL, and MSENL, we now turn to comparing EUL and EUNL to the mean-variance portfolios constructed with other covariance matrix estimators, i.e., SAMPLE, LW2004, LW2020, BGP, and PCA. As expected, the sample covariance matrix delivers by far the worst performance. The MSE-based shrinkage estimators LW2004, LW2020, and BGP offer a clear improvement relative to SAMPLE but still perform quite poorly overall in terms of EU, which often turns negative. This is particularly notable for LW2020 and BGP as LW2004 systematically outperforms both of them. As a result, our proposed EU-maximizing shrinkage estimators, EUL and EUNL, almost systematically improve upon each of LW2004, LW2020, and BGP, which is consistent with the improvement relative to MSENL and MSEL discussed above.

The PCA-based estimator delivers superior and more stable results relative to the other benchmarks, with a positive EU in most cases. This improved performance can be explained because PCA only keeps the first few eigenvalues, avoiding issues related to small eigenvalues. Consequently, as documented by Chen and Yuan (2016), PCA delivers stable portfolio weights. Indeed, Table 4.5 of the supplementary material shows that among all portfolio strategies of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$, PCA delivers the smallest turnover overall. Nonetheless, our EU-maximizing estimators EUL and EUNL outperform PCA in all configurations except one, even though the EU is net of transaction costs proportional to turnover. This suggests, as Kelly et al. (2023) put it, that "PC-sparsity is just an artifact of inefficient shrinkage."

Next, we compare our EUL and EUNL portfolios with the GMV, GMVRF, EW, and EWRF benchmarks. As expected, both GMV and GMVRF perform better under the LW2004 and SSHH covariance matrix estimators than the sample covariance matrix. The GMV portfolio under LW2004 is a notoriously difficult benchmark. Still, EUL (resp. EUNL)

outperforms this benchmark in nine (resp. eight) out of 12 cases when $\gamma = 3$ and nine (resp. 11) out of 12 cases when $\gamma = 10$. Compared to GMVRF, both EUL and EUNL outperform in eight out of 12 cases for both values of γ. Turning to EW and EWRF, which can also be difficult competitors as they are not much affected by estimation risk, the comparison is even more in favor of EUL and EUNL. Specifically, EUL (resp. EUNL) outperforms EW in 10 (resp. 11) out of 12 cases when $\gamma = 3$, and both EUL and EUNL outperform EW in all cases when $\gamma = 10$. Compared to EWRF, EUL (resp. EUNL) outperforms in 10 (resp. 12) out of 12 cases for both values of γ. These results imply that it is possible to significantly outperform naive portfolio strategies using mean-variance portfolios, provided that the impact of estimation errors stemming from mean returns is properly accounted for in the design of the shrinkage covariance matrix used in the construction of the mean-variance portfolio.

Finally, we compare our EUL and EUNL portfolios to the recently proposed MAXSER, UPSA, KNS, and F+A benchmarks. Overall, these benchmarks offer substantial improvements relative to the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ implemented with the SAMPLE estimator and improved estimators like the LW2020 nonlinear shrinkage method. However, unlike with our EUL and EUNL portfolios, these benchmarks deliver out-of-sample utilities that can turn negative and are quite variable across datasets. EUL and EUNL outperform these four benchmarks in the majority of cases. In particular, across 24 configurations, EUNL outperforms MAXSER, UPSA, KNS, and F+A in 24, 16, 18, and 22 cases, respectively. As shown in Table 4.5 of the supplementary material, these benchmarks also face a substantially larger turnover than our EUL and EUNL portfolios.

4.7 Conclusion

More than two decades ago, Ledoit and Wolf (2003a) warned investors that "nobody should be using the sample covariance matrix for the purpose of portfolio optimization", and instead, one should use a shrinkage estimator. Similarly, in this chapter, we recommend caution in using off-the-shelf shrinkage estimators of the sample covariance matrix in the construction of mean-variance portfolios. This is because existing linear and nonlinear shrinkage estimators aim to minimize the mean squared error (MSE) relative to the true covariance matrix, which is not the investor's objective. Instead, we design linear and nonlinear shrinkage estimators of the covariance matrix that optimize the mean-variance investor's expected out-of-sample utility (EU), analytically and in finite samples, hence also incorporating estimation errors in expected returns in the calibration of the shrinkage intensities. This proposal addresses the recent call of switching from "predict-then-optimize" to "predict-and-optimize" in solving decision problems (Elmachtoub and Grigas, 2022).

We show that one should shrink the eigenvalues of the sample covariance matrix much more intensively when the objective is to optimize the EU instead of the MSE. In

nonlinear shrinkage, the EU criterion shrinks all inverse eigenvalues closer to zero, and more intensively so for PCs with low squared Sharpe ratio. However, our shrinkage estimators still keep all PCs, i.e., they are not PC-sparse. Across six empirical datasets, mean-variance portfolios constructed with our linear and nonlinear EU-optimized shrinkage covariance matrices compare favorably out of sample to MSE-optimized covariance matrices (Ledoit and Wolf, 2004; Bodnar et al., 2014; Ledoit and Wolf, 2020), PC-sparse covariance matrices (Chen and Yuan, 2016), regularized mean-variance portfolios under a PCA factor model (Kozak et al., 2020; Da et al., 2024), the sparse MAXSER strategy (Ao et al., 2019), the universal portfolio shrinkage approximator (UPSA) (Kelly et al., 2023), the GMV portfolio under different covariance matrices, and the naive equally weighted portfolio. Moreover, the mean-variance portfolio constructed with our EU-optimized shrinkage covariance matrices have a smaller turnover than alternative mean-variance portfolios due to the intensive shrinkage implied by our method.

Future research can exploit the ideas developed in this chapter to shrink both the mean vector and the covariance matrix to maximize the EU, similar to Bodnar et al. (2024) who shrink the covariance matrix and the portfolio weights to minimize the out-of-sample variance, and to shrink the covariance matrix so as to maximize the expected out-of-sample Sharpe ratio of the mean-variance portfolio. In addition, mean-variance efficient portfolio weights provide the stochastic discount factor (SDF) in asset pricing. MSE-based shrinkage covariance matrices are routinely used for that purpose; see, e.g., Chernov et al. (2023) who apply the Ledoit and Wolf (2020) method to obtain a currency SDF. Our methodology can be used to obtain a shrinkage SDF delivering smaller out-of-sample pricing errors.

Supplementary Material

This supplementary material contains six sections. In Section 4.A, we study the EU-optimal linear and nonlinear shrinkage intensities when the covariance matrix is known. In Section 4.B, we test the accuracy of the proposed approximation to the optimal linear shrinkage intensity. In Section 4.C, we study two ways of constraining the EU-optimal nonlinearly shrunk eigenvalues to be positive. In Section 4.D, we investigate the empirical stability of our shrinkage intensities in empirical data and its impact on portfolio turnover. In Section 4.E, we report the out-of-sample Sharpe ratio of the considered strategies in the empirical analysis. Finally, in Section 4.G, we provide the proofs of all theoretical results.

4.A Shrinkage with known covariance matrix

In this section, we consider the case in which the covariance matrix $\boldsymbol{\Sigma}$ is known and estimation errors only stem from the sample mean returns $\hat{\boldsymbol{\mu}}$. In that case, the MSE loss function would set the shrinkage intensities of eigenvalues equal to zero, unlike the EU criterion.

4.A.1 Linear shrinkage with known covariance matrix

The linear shrinkage covariance matrix when $\boldsymbol{\Sigma}$ is known is

$$\boldsymbol{\Sigma}_{\boldsymbol{D}} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}' \quad \text{with} \quad \boldsymbol{D} = (1-\delta)\boldsymbol{\Lambda} + \delta\bar{\lambda}\boldsymbol{I}_N,\ \bar{\lambda} = \frac{1}{N}\text{tr}(\boldsymbol{\Lambda}). \tag{4.51}$$

In the next proposition, we derive the EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (4.51). Unlike the case in which we shrink the sample covariance matrix in Proposition 4.3, we don't need any distributional assumption for the returns and we can have $N > T$. To distinguish with the case in which we use the sample covariance matrix, and because assuming $\boldsymbol{\Sigma}$ is known is an approximation, we denote the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ by $\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$.

Proposition 4.6. *The EU in* (4.13) *of the shrinkage mean-variance portfolio* $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$*, with* $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ *in* (4.51)*, is*

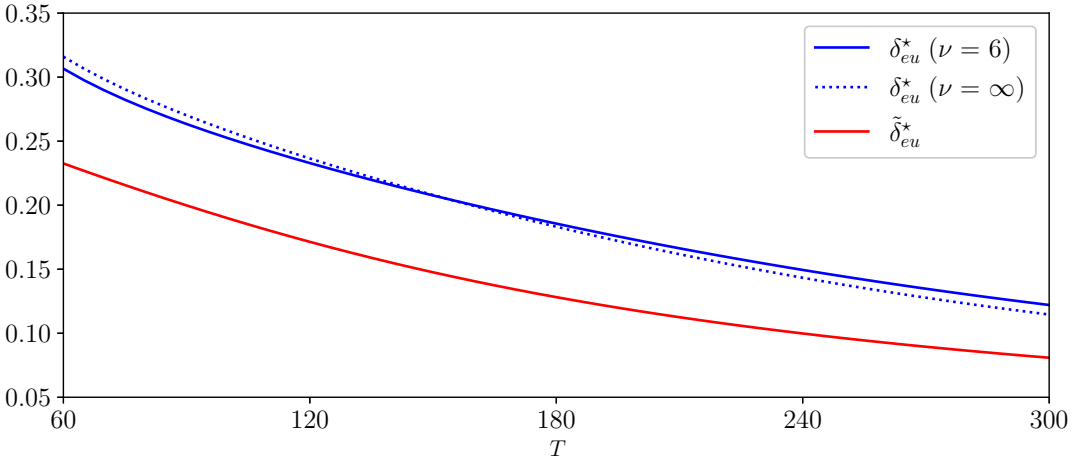
$$\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{\gamma}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2 - \frac{1}{2\gamma}\sum_{i=1}^{N} f_i(\delta)^2\left(\theta_{PC_i}^2 + \frac{1}{T}\right), \tag{4.52}$$

where $f_i(\delta)$ *is defined in* (4.18)*.*

As is the case when shrinking the sample covariance matrix, there is no closed-form expression for the shrinkage intensity maximizing (4.52),

$$\tilde{\delta}_{eu}^{\star} = \underset{\delta\in\mathbb{R}}{\text{argmax}}\,\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}), \tag{4.53}$$

Figure 4.4: EU-optimal linear shrinkage intensity with known covariance matrix



Notes. This figure depicts, in red, the linear shrinkage intensity $\tilde{\delta}^\star_{eu}$ maximizing the EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where the shrinkage covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (4.51) considers that the covariance matrix $\boldsymbol{\Sigma}$ is known. For comparison, we also depict, in blue, the optimal shrinkage intensity under the sample covariance matrix, $\delta^\star_{eu}$ in (4.20), for degrees of freedom $\nu \in \{6, \infty\}$ in the multivariate t-distribution. We depict the shrinkage intensities as a function of the sample size $T \in [60, 300]$ and we calibrate the eigenvalues $\boldsymbol{\Lambda}$ and the PCs' squared Sharpe ratios $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024.

but it can easily be found numerically. In Figure 4.4, we depict $\tilde{\delta}^\star_{eu}$ using a setup similar to that in Figure 4.1. For comparison, we also depict the optimal shrinkage intensity under the sample covariance matrix, $\delta^\star_{eu}$ in (4.20). We depict the shrinkage intensities as a function of the sample size $T \in [60, 300]$. We make two main observations. First, $\tilde{\delta}^\star_{eu}$ is much larger than zero, which contrasts with $\delta^\star_{mse} = 0$ under a known covariance matrix. Second, $\tilde{\delta}^\star_{eu}$ is lower than $\delta^\star_{eu}$, i.e., we shrink less intensively when the covariance matrix is known, as expected.

4.A.2 Nonlinear shrinkage with known covariance matrix

The nonlinear shrinkage covariance matrix when $\boldsymbol{\Sigma}$ is known is

$$\boldsymbol{\Sigma}_{\boldsymbol{D}} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}' \quad \text{with} \quad \boldsymbol{D} = (\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}, \tag{4.54}$$

where $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_N)$ is a diagonal matrix containing the N shrinkage intensities δ_i to calibrate. In the next proposition, we derive the EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (4.54). Unlike the case in which we shrink the sample covariance matrix in Proposition 4.5, we don't need any distributional assumption for the returns and we can have $N > T$. As in Section 4.A.1, we denote the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ by $\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$.

Proposition 4.7. *The following holds concerning the shrinkage mean-variance portfolio* $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$*, with* $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ *in* (4.54)*:*

1. The EU in (4.13) *of* $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ *is*

$$\widetilde{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{\gamma}\sum_{i=1}^{N}(1-\delta_i)\theta_{PC_i}^2 - \frac{1}{2\gamma}\sum_{i=1}^{N}(1-\delta_i)^2\left(\theta_{PC_i}^2 + \frac{1}{T}\right). \tag{4.55}$$

2. The shrinkage intensities δ_i *maximizing* $\widetilde{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *in* (4.55) *are*

$$\tilde{\delta}_{eu,i}^{\star} = \frac{1/T}{\theta_{PC_i}^2 + 1/T} \in [0,1], \tag{4.56}$$

which are lower than those under the sample covariance matrix, $\delta_{eu,i}^{\star}$ *in* (4.28).

3. The EU delivered by $\tilde{\delta}_{eu,i}^{\star}$ *is*

$$\widetilde{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}^\star}) = \frac{1}{2\gamma}\sum_{i=1}^{N}\left(1-\delta_{eu,i}^{\star}\right)\theta_{PC_i}^2. \tag{4.57}$$

Proposition 4.7 shows that even when the covariance matrix $\boldsymbol{\Sigma}$ is known, the eigenvalues with a small corresponding squared Sharpe ratio $\theta_{PC_i}^2$ are shrunk intensively. In particular $\tilde{\delta}_{eu,i}^{\star} = 1$ when $\theta_{PC_i}^2 = 0$. We also have $\tilde{\delta}_{eu,i}^{\star} \in [0,1]$ for all i.

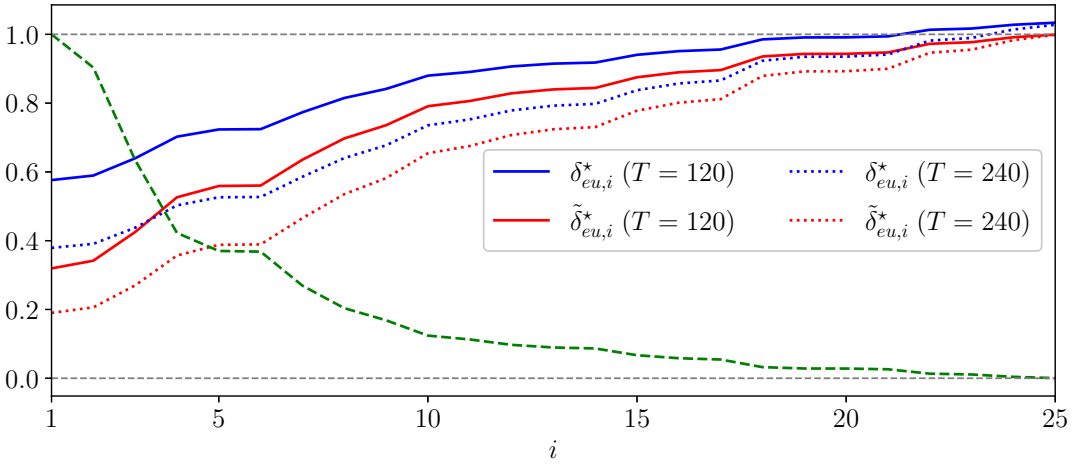
Figure 4.5 depicts the shrinkage intensities $\tilde{\delta}_{eu,i}^{\star}$ in (4.56) using a setup similar to that in Figure 4.2. For comparison, we also depict the optimal shrinkage intensities under the sample covariance matrix, $\delta_{eu,i}^{\star}$ in (4.28). We depict the shrinkage intensities for a sample size $T \in \{120, 240\}$. We make three main observations. First, $\tilde{\delta}_{eu,i}^{\star}$ is much larger than zero, which contrasts with $\delta_{mse,i}^{\star} = 0$ for all i under a known covariance matrix. Second, in line with Proposition 4.7, the shrinkage intensities $\tilde{\delta}_{eu,i}^{\star}$ are lower than $\delta_{eu,i}^{\star}$ for all i, i.e., we shrink less intensively when the covariance matrix is known. Third, $\tilde{\delta}_{eu,i}^{\star}$ does indeed approach one when $\theta_{PC_i}^2$ is close to zero.

4.B Accuracy of approximation to shrinkage intensity

In this section, we illustrate the accuracy of the approximation used in Proposition 4.3 to derive the EU-optimal linear shrinkage intensity, $\delta_{eu}^{\star}$ in (4.20). Proposition 4.5 uses a similar approximation to derive the EU-optimal nonlinear shrinkage intensities, but we focus on the linear shrinkage intensity because it is easier to test the accuracy for a single parameter.

We proceed as follows. As in the analysis of Section 4.3.3, we consider a sample size $T \in [60, 300]$ and calibrate $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024 (25SBM). This gives us the eigenvalues $\boldsymbol{\Lambda}$ and the PCs' squared Sharpe ratios $\boldsymbol{\theta}_{PC}^2$. Assuming a multivariate t-distribution for the asset returns with $\nu \in \{6, \infty\}$ degrees of freedom, we can then find (k_1, k_2, k_3) and compute the approximate $\delta_{eu}^{\star}$ by solving Problem (4.20). We can also find the EU-optimal shrinkage intensity without approximation, denoted $\delta_{eu}^{\star\star}$, via Monte Carlo simulations. Specifically,

Figure 4.5: EU-optimal nonlinear shrinkage intensities with known covariance matrix



Notes. This figure depicts, in red, the nonlinear shrinkage intensities $\tilde{\delta}^{\star}_{eu,i}$ maximizing the EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where the shrinkage covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (4.54) considers that the covariance matrix $\boldsymbol{\Sigma}$ is known. We also depict, in blue, the optimal shrinkage intensities under the sample covariance matrix, $\delta^{\star}_{eu,i}$ in (4.28), for degrees of freedom $\nu \in \{6, \infty\}$ in the multivariate t-distribution. We depict the shrinkage intensities for the eigenvalues λ_i sorted in descending order of their corresponding squared Sharpe ratios $\theta^2_{PC_i}$. The sample size is $T \in \{120, 240\}$. The green dashed line depicts the ratio $\theta^2_{PC_i}/\theta^2_{PC_1}$. We calibrate the eigenvalues $\boldsymbol{\Lambda}$ and the PCs' squared Sharpe ratios $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024. The horizontal gray dashed lines depict a shrinkage intensity δ_i of zero and one.

given $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, we simulate $M = 10{,}000$ times a sample of multivariate t-distributed asset returns of size T, $(\boldsymbol{r}_{1,m}, \ldots, \boldsymbol{r}_{T,m})$ with $m = 1, \ldots, M$. In each simulation m, we compute the sample mean $\hat{\boldsymbol{\mu}}_m$ and sample covariance matrix $\hat{\boldsymbol{\Sigma}}_m$ as in (4.5). Then, given a shrinkage intensity δ, we construct the shrinkage mean-variance portfolio in each simulation m as

$$\hat{\boldsymbol{w}}_{\boldsymbol{D},m} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}^{-1}_{\boldsymbol{D},\boldsymbol{m}}\hat{\boldsymbol{\mu}}_m \quad \text{where} \quad \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D},\boldsymbol{m}} = (1-\delta)\hat{\boldsymbol{\Sigma}}_m + \delta\frac{\operatorname{tr}(\hat{\boldsymbol{\Sigma}}_m)}{N}\boldsymbol{I}_N. \tag{4.58}$$

Averaging over all M simulations, the simulated EU for a given δ is

$$\overline{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\boldsymbol{w}}'_{\boldsymbol{D},m}\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}'_{\boldsymbol{D},m}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D},m}\right). \tag{4.59}$$

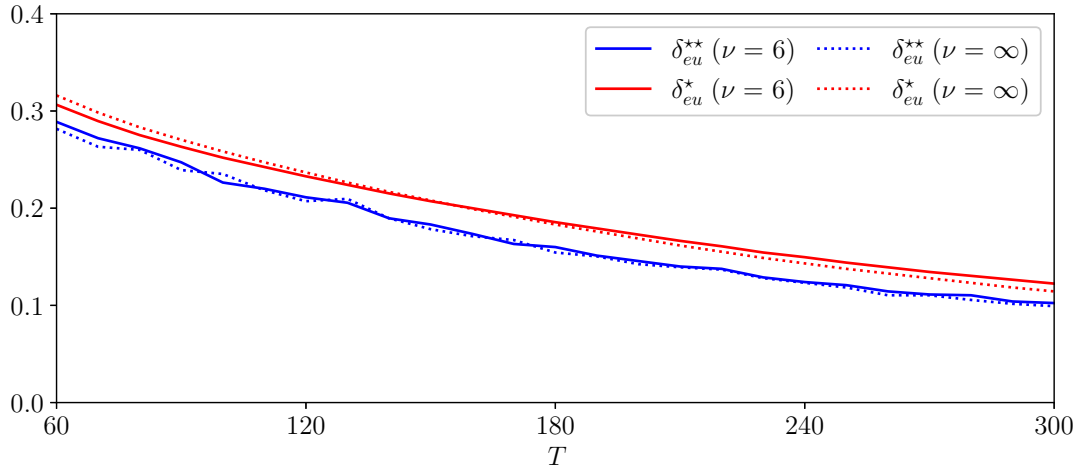
Finally, we find the EU-optimal shrinkage intensity $\delta^{\star\star}_{eu}$ maximizing (4.59),

$$\delta^{\star\star}_{eu} = \underset{\delta\in[0,1]}{\operatorname{argmax}}\,\overline{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}), \tag{4.60}$$

which we obtain by considering a grid of 1,000 values of δ.

In Figure 4.6, we compare $\delta^{\star}_{eu}$ with $\delta^{\star\star}_{eu}$ for $\nu \in \{6, \infty\}$ and $T \in [60, 300]$. We make two main observations. First, $\delta^{\star\star}_{eu}$ and $\delta^{\star}_{eu}$ are consistently close to one another, indicating that the approximation used in Proposition 4.3 is accurate. Second, $\delta^{\star}_{eu}$ is consistently larger than $\delta^{\star\star}_{eu}$, indicating that the approximation yields more conservative shrinkage intensities. This result is expected because the approximation overstates the amount of randomness in the shrinkage covariance matrix, as discussed in Remark 4.4. These results are robust to considering the other datasets used in the empirical analysis of Section 4.6.

Figure 4.6: Accuracy of approximation to the EU-optimal linear shrinkage intensity



Notes. This figure depicts linear shrinkage intensities maximizing the EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (4.14). The intensity $\delta^{\star}_{eu}$ in (4.20) (red) is based on the approximation in Proposition 4.3, whereas the intensity $\delta^{\star\star}_{eu}$ in (4.60) (blue) is without approximation and obtained via Monte Carlo simulations as explained in Section 4.B. We consider a sample size $T \in [60, 300]$ and calibrate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024. The asset returns follow a multivariate t-distribution with degrees of freedom $\nu \in \{6, \infty\}$.

4.C Positive nonlinearly shrunk eigenvalues

We demonstrate in Proposition 4.5, and illustrate in Figure 4.2, that the EU-optimal nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$ in (4.28) are larger than one provided that $\theta^2_{PC_i} < \bar{\theta}^2_{PC}$ in (4.32). As a result, the nonlinearly shrunk sample eigenvalues, $\hat{\lambda}_i/(1 - \delta^{\star}_{eu,i})$, can turn negative, which might be considered undesirable even if theoretically optimal. Therefore, in this section, we propose two ways of making the nonlinear shrinkage intensities stay below one, without sacrificing the empirical performance.

The first method we propose to make the nonlinear shrinkage intensities stay below one starts from the observation that the sum of the $\delta^{\star}_{eu,i}$ in (4.28) is equal to

$$\sum_{i=1}^{N} \delta^{\star}_{eu,i} = N - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)} \sum_{i=1}^{N} R_i, \tag{4.61}$$

where $R_i = k_1\theta^2_{PC_i}/(k_2\theta^2_{PC_i} + k_3/T)$. Therefore, we propose to use the shrinkage intensities

$$\delta^{\circ}_{eu,i} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)} R_i, \tag{4.62}$$

which satisfy $\delta^{\circ}_{eu,i} \in [0, 1]$, as desired, as well as $\sum_{i=1}^{N} \delta^{\circ}_{eu,i} = \sum_{i=1}^{N} \delta^{\star}_{eu,i}$. In the next proposition, we compare $\delta^{\circ}_{eu,i}$ with $\delta^{\star}_{eu,i}$ and we derive the EU delivered by $\delta^{\circ}_{eu,i}$.

Proposition 4.8. *The following holds concerning the nonlinear shrinkage intensities* $\delta^{\circ}_{eu,i}$ *in* (4.62)*:*

1. *The difference between* $\delta^{\circ}_{eu,i}$ *and the EU-optimal nonlinear shrinkage intensities* $\delta^{\star}_{eu,i}$ *in* (4.28) *is*

$$\delta^{\circ}_{eu,i} - \delta^{\star}_{eu,i} = \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} \left(R_i - \bar{R}\right). \tag{4.63}$$

In particular, $\delta^{\circ}_{eu,i} \geq \delta^{\star}_{eu,i}$ *if* $R_i \geq \bar{R}$*, and* $\delta^{\circ}_{eu,i} < \delta^{\star}_{eu,i}$ *otherwise.*

2. *The EU* $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *in* (4.26) *delivered by* $\delta^{\circ}_{eu,i}$ *is*

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}^{\circ}}) = \frac{k_1(T-1)(T+N-2)}{2\gamma(T-2)(T-N-2)} \sum_{i=1}^{N} (1-\delta^{\circ}_{eu,i}) \left(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2}\right) \geq 0. \tag{4.64}$$

Part 1 of Proposition 4.8 shows that $\delta^{\circ}_{eu,i}$ is a dampened version of $\delta^{\star}_{eu,i}$, i.e., ordering the $\theta^2_{PC_i}$ in decreasing order, $\delta^{\circ}_{eu,i} \geq \delta^{\star}_{eu,i}$ for i from 1 up to k satisfying $R_k \geq \bar{R}$ and $R_{k+1} \leq \bar{R}$, and then $\delta^{\circ}_{eu,i} \leq \delta^{\star}_{eu,i}$ for i from $k+1$ up to N. Part 2 of Proposition 4.8 shows, quite unexpectedly, that $\delta^{\circ}_{eu,i}$ always delivers a positive EU even though it is not EU-optimal.

The second method we propose to make the nonlinear shrinkage intensities stay below one is to solve the following constrained quadratic optimization problem:

$$\boldsymbol{\delta}^{+}_{eu} = \underset{\boldsymbol{\delta} \in \mathbb{R}^N}{\operatorname{argmax}} \, \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) \quad \text{subject to} \quad \delta_i \leq 1 \; \forall i, \tag{4.65}$$

where the + subscript refers to the shrunk eigenvalues being positive and $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ is the quadratic function of $\boldsymbol{\delta}$ in (4.26). In the next proposition, we derive an expression for $\boldsymbol{\delta}^{+}_{eu,l}$.

Proposition 4.9. *Let* $\eta_i \geq 0$ *be the Lagrange multiplier corresponding to the* i*th inequality constraint in* (4.65) *at the optimum. Define*

$$R^{+}_i = \frac{k_1(\theta^2_{PC_i} + \eta_i)}{k_2\theta^2_{PC_i} + \frac{k_3}{T}} \geq R_i, \tag{4.66}$$

and $\bar{R}^{+} = \frac{1}{N}\sum_{i=1}^{N} R^{+}_i$*. Then, the constrained nonlinear shrinkage intensities* $\delta^{+}_{eu,i}$ *in* (4.65) *are*

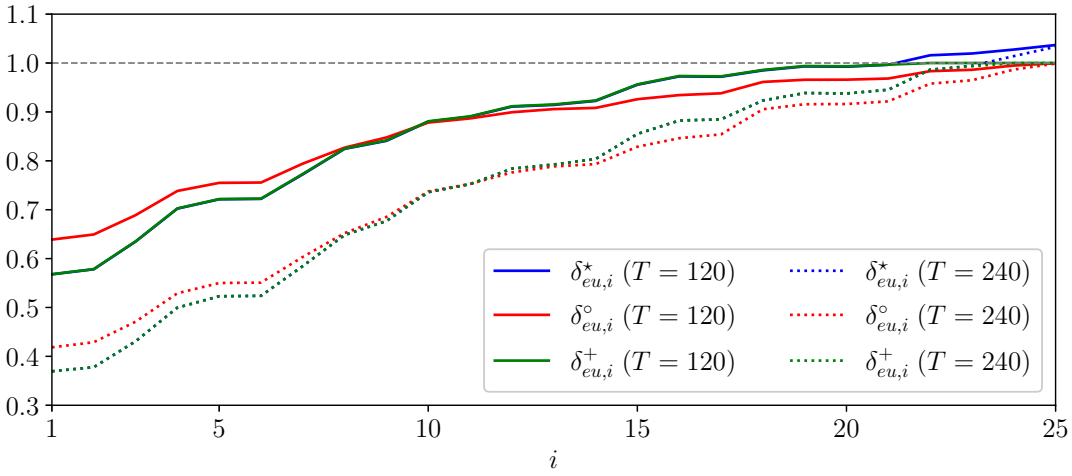
$$\delta^{+}_{eu,i} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)} \left(R^{+}_i - \frac{N}{T-2}\bar{R}^{+}\right), \tag{4.67}$$

and they satisfy

$$\delta^{+}_{eu,i} = \begin{cases} 1 & \text{if } \eta_i > 0, \\ \delta^{\star}_{eu,i} + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} \left(\bar{R}^{+} - \bar{R}\right) \geq \delta^{\star}_{eu,i} & \text{if } \eta_i = 0, \end{cases} \tag{4.68}$$

where $\delta^{\star}_{eu,i}$ *in* (4.28) *are the EU-optimal nonlinear shrinkage intensities.*

Figure 4.7: Comparison of nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$ and $\delta^{+}_{eu,i}$



Notes. This figure depicts the nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$ and $\delta^{+}_{eu,i}$, which are studied in Propositions 4.5, 4.8, and 4.9, respectively. $\delta^{\star}_{eu,i}$ maximizes the EU in (4.26) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (4.21). $\delta^{\circ}_{eu,i}$ is a dampened version of $\delta^{\star}_{eu,i}$ such that $\delta^{\circ}_{eu,i} \in [0, 1]$ and $\sum_{i=1}^{N} \delta^{\circ}_{eu,i} = \sum_{i=1}^{N} \delta^{\star}_{eu,i}$. $\delta^{+}_{eu,i}$ also maximizes the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$, but under the constraint that $\delta^{+}_{i} \leq 1$. We depict the shrinkage intensities for the eigenvalues λ_i sorted in descending order of their corresponding squared Sharpe ratios $\theta^2_{PC_i}$. We consider sample sizes $T \in \{120, 240\}$. We calibrate the eigenvalues $\boldsymbol{\Lambda}$ and the PCs' squared Sharpe ratios $\boldsymbol{\theta}^2_{PC}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024. The asset returns follow a multivariate t-distribution with degrees of freedom $\nu \in \{6, \infty\}$.

Proposition 4.9 shows that, as a compensation for constraining the nonlinear shrinkage intensities to stay below one, those for which the constraint is not binding are larger than the unconstrained EU-optimal nonlinear shrinkage intensities, $\delta^{\star}_{eu,i}$ in (4.28).

Figure 4.7 depicts the three nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$, and $\delta^{+}_{eu,i}$ using a setup similar to that in Figure 4.2 of the main text, for sample sizes $T = 120$ and 240. We observe that $\delta^{\circ}_{eu,i}$ and $\delta^{+}_{eu,i}$ are indeed smaller than one when $\delta^{\star}_{eu,i}$ exceeds one. Whereas $\delta^{\circ}_{eu,i}$ can visibly differ from $\delta^{\star}_{eu,i}$, particularly when T is smaller as expected from (4.63), $\delta^{+}_{eu,i}$ and $\delta^{\star}_{eu,i}$ are essentially equal for the indices i such that $\delta^{\star}_{eu,i} < 1$. Overall, the three different nonlinear shrinkage methods are reasonably close and, in unreported results, we show that $\delta^{\circ}_{eu,i}$ and $\delta^{+}_{eu,i}$ deliver only a small theoretical EU loss relative to $\delta^{\star}_{eu,i}$.

Next, we compare the empirical performance delivered by $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$, and $\delta^{+}_{eu,i}$. Table 4.4 reports the annualized net-of-cost out-of-sample utility delivered by each of these shrinkage intensities in the same empirical setting as in Section 4.6. We observe that the three nonlinear shrinkage covariance matrix estimators deliver a similar empirical performance. If anything, the constrained intensities $\delta^{+}_{eu,i}$ perform slightly better overall.

4.D Stability of estimated shrinkage intensities and impact on portfolio turnover

In Section 4.5 of the main text, we detail how we estimate the EU-optimal and MSE-optimal linear and nonlinear shrinkage intensities that we use in the empirical analysis

Table 4.4: Annualized net out-of-sample utility in empirical data delivered by the nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$ and $\delta^{+}_{eu,i}$ (in percentage points)

		$T = 120$						$T = 240$					
		Dataset						Dataset					
Portfolio	δ_i	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\delta^{\star}_{eu,i}$	9.36	11.0	3.42	20.5	2.82	10.5	12.2	11.0	6.88	16.6	3.24	6.28
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\delta^{\circ}_{eu,i}$	9.08	10.9	3.52	20.5	2.80	10.4	12.2	10.9	6.72	16.6	3.23	6.26
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\delta^{+}_{eu,i}$	9.99	11.0	3.64	20.5	2.81	10.6	12.3	11.0	6.95	16.6	3.24	6.30
Panel (b): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\delta^{\star}_{eu,i}$	2.80	3.31	1.03	6.15	0.85	3.15	3.66	3.28	2.06	4.99	0.97	1.89
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\delta^{\circ}_{eu,i}$	2.72	3.27	1.06	6.16	0.84	3.13	3.65	3.27	2.02	4.99	0.97	1.88
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\delta^{+}_{eu,i}$	2.95	3.30	1.09	6.15	0.84	3.18	3.68	3.29	2.09	5.01	0.97	1.90

Notes. This table reports the annualized net-of-cost out-of-sample utility, in percentage points, delivered by nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$ and $\delta^{+}_{eu,i}$, which are studied in Propositions 4.5, 4.8, and 4.9, respectively. $\delta^{\star}_{eu,i}$ maximizes the EU in (4.26) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (4.21). $\delta^{\circ}_{eu,i}$ is a dampened version of $\delta^{\star}_{eu,i}$ such that $\delta^{\circ}_{eu,i} \in [0,1]$ and $\sum_{i=1}^{N}\delta^{\circ}_{eu,i} = \sum_{i=1}^{N}\delta^{\star}_{eu,i}$. $\delta^{+}_{eu,i}$ also maximizes the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$, but under the constraint that $\delta^{+}_{i} \leq 1$. We report results across the six datasets described in Table 4.1 and we follow the out-of-sample methodology detailed in Section 4.6.3. The net out-of-sample utility is computed using rolling windows with sample sizes $T = 120$ and $T = 240$ months and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 3$ in Panel (a) and $\gamma = 10$ in Panel (b).

of Section 4.6. In this section, we assess the stability of these estimators across different estimation windows. In particular, because the estimated EU-optimal intensities depend on the sample mean returns $\hat{\boldsymbol{\mu}}$, one concern is that they might be more unstable than the estimated MSE-optimal intensities, which could increase portfolio turnover.

We start by comparing the estimated EU-optimal and MSE-optimal linear shrinkage intensities across estimation windows. We focus on the linear intensities for conciseness because they are sufficient to make our point. Figure 4.8 depicts boxplots of the estimated linear shrinkage intensities $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$ following the methodology detailed in Section 4.5, which we average across the six datasets listed in Table 4.1. We consider a sample $T = 120$ and 240. We make two main observations. First, as expected, the EU-optimal shrinkage intensities are larger than the MSE-optimal ones, and both are larger for $T = 120$ than 240. Second, $\hat{\delta}^{\star}_{eu}$ faces a considerably higher volatility than $\hat{\delta}^{\star}_{mse}$, which is because $\hat{\delta}^{\star}_{eu}$ is a function of both $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, unlike $\hat{\delta}^{\star}_{mse}$ that only depends on $\hat{\boldsymbol{\Sigma}}$.

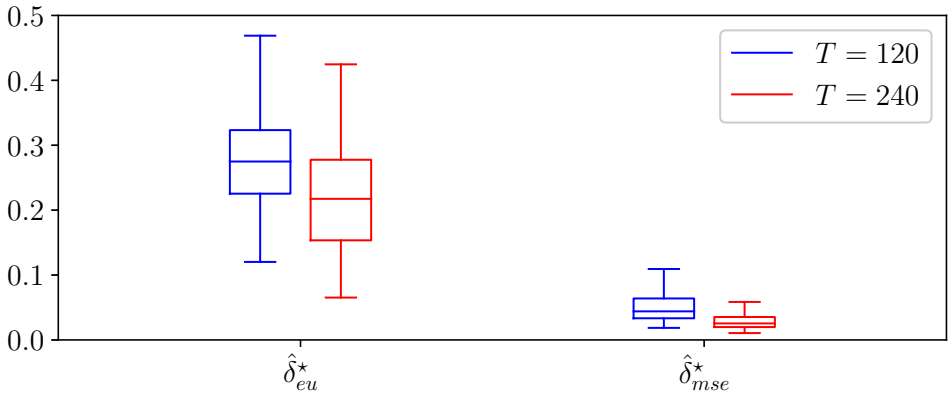
Next, we examine whether this more unstable shrinkage intensity translates into a higher turnover. To that end, Table 4.5 reports the average monthly turnover of the 16 portfolio strategies considered in the empirical analysis. We can observe that the turnover of the EUL strategy is systematically significantly lower than that of the MSEL strategy. Therefore, by shrinking the covariance matrix more intensively, EUL achieves a smaller turnover than MSEL, even if the shrinkage intensity is more volatile. Similarly, under nonlinear shrinkage, EUNL delivers a much lower turnover than MSENL. In fact, Table 4.5 further reveals that EUL (EUNL) achieves the third-lowest (second-lowest) turnover among

Table 4.5: Average monthly portfolio turnover in empirical data

Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	$T = 120$						$T = 240$					
		Dataset						Dataset					
		25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13)	EUL	1.34	0.73	1.52	3.13	0.83	2.21	1.02	0.54	0.90	1.63	0.50	1.63
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	0.89	0.55	0.68	2.83	0.47	1.05	0.83	0.46	0.54	1.73	0.50	1.15
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	4.39	1.58	4.28	5.83	2.45	10.6	2.85	1.05	2.28	3.31	1.60	7.74
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	3.65	1.38	3.74	5.33	2.08	8.36	2.48	0.95	2.11	3.05	1.40	6.54
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	10.7	2.76	7.29	15.3	5.80	50.1	4.68	1.44	2.85	6.47	2.64	15.7
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	4.36	1.56	4.23	5.36	2.45	10.7	2.81	1.04	2.26	3.01	1.57	7.74
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	6.98	2.37	4.65	12.1	4.44	18.7	3.86	1.35	2.31	5.80	2.32	10.5
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	6.90	2.14	5.80	7.93	3.73	21.1	3.91	1.29	2.64	4.62	2.16	11.8
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	0.48	0.26	0.37	4.13	0.31	0.66	0.32	0.19	0.20	1.83	0.16	0.55
GMV	SAMPLE	0.79	0.30	0.57	0.07	0.50	1.51	0.45	0.18	0.29	0.04	0.28	0.77
GMV	LW2004	0.38	0.18	0.35	0.03	0.25	0.53	0.29	0.14	0.24	0.02	0.20	0.43
GMV	SSHH	0.50	0.25	0.35	0.06	0.38	0.60	0.35	0.17	0.23	0.04	0.25	0.45
GMVRF	SAMPLE	1.12	0.40	0.65	5.61	0.53	1.73	0.72	0.25	0.38	2.76	0.28	0.73
GMVRF	LW2004	0.48	0.25	0.38	2.01	0.28	0.44	0.44	0.19	0.30	1.33	0.20	0.37
GMVRF	SSHH	0.63	0.32	0.36	4.39	0.39	0.50	0.54	0.22	0.28	2.56	0.24	0.38
EW	/	0.04	0.04	0.04	0.01	0.04	0.04	0.04	0.04	0.04	0.01	0.04	0.05
EWRF	/	0.06	0.06	0.06	1.06	0.06	0.06	0.04	0.04	0.04	0.61	0.03	0.03
MAXSER	/	4.11	1.63	2.27	12.0	2.70	11.4	2.42	1.01	1.33	5.77	1.85	5.54
UPSA	/	3.97	1.93	3.71	6.94	3.29	7.49	1.91	1.11	1.86	3.09	2.23	3.34
KNS	/	5.52	2.43	3.56	16.3	2.94	16.3	3.98	1.34	2.21	9.45	2.25	8.91
F+A	/	6.65	3.41	7.54	16.4	3.84	11.2	3.62	1.86	4.13	7.48	2.03	7.80
Panel (b): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13)	EUL	0.40	0.22	0.46	0.94	0.25	0.66	0.31	0.16	0.27	0.49	0.15	0.49
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	0.27	0.16	0.20	0.85	0.14	0.31	0.25	0.14	0.16	0.52	0.15	0.35
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	1.32	0.47	1.28	1.75	0.74	3.17	0.85	0.31	0.68	0.99	0.48	2.32
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	1.09	0.41	1.12	1.60	0.62	2.51	0.74	0.29	0.63	0.91	0.42	1.96
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	3.21	0.83	2.19	4.58	1.74	15.0	1.40	0.43	0.85	1.94	0.79	4.71
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	1.31	0.47	1.27	1.61	0.73	3.22	0.84	0.31	0.68	0.90	0.47	2.32
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	2.09	0.71	1.39	3.63	1.33	5.62	1.16	0.40	0.69	1.74	0.69	3.14
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	2.07	0.64	1.74	2.38	1.12	6.34	1.17	0.39	0.79	1.39	0.65	3.54
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	0.14	0.08	0.11	1.24	0.09	0.20	0.10	0.06	0.06	0.55	0.05	0.16
GMV	SAMPLE	0.79	0.30	0.57	0.07	0.50	1.51	0.45	0.18	0.29	0.04	0.28	0.77
GMV	LW2004	0.38	0.18	0.35	0.03	0.25	0.53	0.29	0.14	0.24	0.02	0.20	0.43
GMV	SSHH	0.50	0.25	0.35	0.06	0.38	0.60	0.35	0.17	0.23	0.04	0.25	0.45
GMVRF	SAMPLE	0.34	0.12	0.20	1.68	0.16	0.52	0.22	0.07	0.11	0.83	0.09	0.22
GMVRF	LW2004	0.14	0.08	0.12	0.60	0.08	0.13	0.13	0.06	0.09	0.40	0.06	0.11
GMVRF	SSHH	0.19	0.10	0.11	1.32	0.12	0.15	0.16	0.07	0.08	0.77	0.07	0.11
EW	/	0.04	0.04	0.04	0.01	0.04	0.04	0.04	0.04	0.04	0.01	0.04	0.05
EWRF	/	0.02	0.02	0.02	0.32	0.02	0.02	0.01	0.01	0.01	0.18	0.01	0.01
MAXSER	/	1.62	0.55	0.93	4.38	0.93	3.83	1.02	0.36	0.53	2.09	0.68	1.94
UPSA	/	1.19	0.58	1.11	2.08	0.99	2.25	0.57	0.33	0.56	0.93	0.67	1.00
KNS	/	1.65	0.73	1.07	4.90	0.88	4.89	1.20	0.40	0.66	2.83	0.67	2.67
F+A	/	2.00	1.03	2.26	4.93	1.15	3.37	1.09	0.56	1.24	2.24	0.61	2.34

Notes. This table reports the average monthly turnover for the 21 portfolio strategies described in Section 4.6.2 and listed in Table 4.2. The first nine of them are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using rotation-equivariant covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}'$ for different choices of the diagonal matrix $\boldsymbol{D}$. EUL and EUNL are the linear and nonlinear shrinkage estimators maximizing the EU introduced in this chapter. We report results across the six datasets described in Table 4.1 and we follow the out-of-sample methodology detailed in Section 4.6.3. The net out-of-sample utility is computed using rolling windows with sample sizes $T = 120$ and $T = 240$ months and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 3$ in Panel (a) and $\gamma = 10$ in Panel (b).

Figure 4.8: Comparison of estimated EU- and MSE-optimal linear shrinkage intensities



Notes. This figure depicts boxplots of the estimated EU-optimal and MSE-optimal linear shrinkage intensities, $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$. The estimation methodology is detailed in Section 4.5. The boxplots are averaged across the six datasets listed in Table 4.1. We consider sample sizes $T = 120$ in blue and $T = 240$ in red.

the nine mean-variance portfolios $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ based on nine different covariance matrix estimators. As discussed in Section 4.6.4, the PCA estimator achieves the lowest turnover among the nine estimators because it completely removes low-variance PCs.

4.E Empirical out-of-sample Sharpe ratio

Table 4.3 in the main text reports the out-of-sample utility as a performance measure because it is the portfolio performance measure used to calibrate the shrinkage intensity we propose in this chapter. Another important mean-variance portfolio performance measure is the Sharpe ratio. Therefore, in Table 4.6, we report the annualized out-of-sample Sharpe ratio net of proportional transaction costs of the 16 portfolio strategies considered in Table 4.3,

$$SR_k = \sqrt{12} \times \frac{\hat{\mu}_k}{\hat{\sigma}_k}, \tag{4.69}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the sample mean and standard deviation of the out-of-sample net portfolio returns, $r_{net,k,t}$ in (4.48), for strategy k. We note that one should not especially expect our proposed shrinkage portfolio strategies, EUL and EUNL, to outperform in terms of Sharpe ratio, because it is not the objective function underlying their construction. And indeed, we observe in particular from Table 4.3 that EUL and EUNL have a comparable, but not necessarily better, Sharpe ratio than MSEL and MSENL. Therefore, a relevant extension for future research is to derive the optimal linear and nonlinear shrinkage of the sample covariance matrix that maximize the expected out-of-sample Sharpe ratio of the mean-variance portfolio.

Table 4.6: Annualized out-of-sample Sharpe ratio in empirical data (in percentage points)

Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	$T = 120$						$T = 240$					
		Dataset						Dataset					
		25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13)	EUL	88.5	**87.3**	**64.3**◑	186●	58.7	194◑	90.0	86.8	68.5○	189●	69.1	131
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	81.7	84.0	45.4	122●	48.2	104●	86.7	81.2	65.5	113●	56.3◑	77.2●
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	89.8	84.8	51.1	205	57.9	220	91.4	86.1	56.2	207	77.0	147
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	90.5	85.7	53.0	204	58.3	**221**	92.1	**86.9**	57.6	207	77.3	149
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	79.6	77.0	43.1	201	54.0	179	85.7	81.8	52.0	199	73.6	128
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	89.3	85.0	50.8	203	57.9	219	91.3	86.2	56.2	207	77.6	147
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	90.3	81.6	49.8	204	56.6	219	89.1	83.8	57.7	201	75.6	142
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	87.9	81.5	46.8	**207**	56.6	209	88.3	83.6	53.5	202	74.8	138
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	50.7	75.1	51.0	118	23.7	62.1	37.5	74.7	47.4	118	−1.85	37.4
GMV	SAMPLE	65.0	58.5	53.5	−219	23.2	9.14	78.4	66.6	81.6	−208	55.7	70.1
GMV	LW2004	70.0	60.7	61.9	−186	29.1	24.3	76.2	67.7	82.9	−187	61.1	75.5
GMV	SSHH	69.2	59.6	62.4	−214	26.1	16.1	77.8	67.1	**83.2**	−206	58.2	73.0
GMVRF	SAMPLE	67.6	52.6	41.8	205	27.1	36.3	66.3	43.5	67.0	214	33.6	27.3
GMVRF	LW2004	65.0	52.0	47.2	198	24.3	38.0	61.7	43.9	68.4	212	33.9	39.8
GMVRF	SSHH	69.2	52.1	47.9	205	25.6	42.7	65.3	43.9	68.9	**214**	34.1	35.4
EW	/	53.1	46.1	56.2	−86.7	13.4	13.4	53.3	48.9	59.1	−83.9	18.6	24.4
EWRF	/	45.9	41.9	38.2	132	−8.45	−13.5	40.7	33.3	48.4	108	2.12	−4.06
MAXSER	/	**92.0**	78.3	54.2	184	51.5	149	85.7	81.4	43.6	181	47.6	96.6
UPSA	/	90.2	76.1	43.3	174	**68.3**	179	**93.2**	82.4	59.5	192	67.9	138
KNS	/	72.3	75.5	42.8	186	41.7	204	79.7	80.8	53.2	189	69.0	138
F+A	/	70.2	78.9	41.2	172	49.1	205	66.6	85.1	54.0	183	**81.6**	**153**
Panel (b): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (4.13)	EUL	88.4	**87.3**	**64.3**◑	185●	58.7	194◑	89.9	86.7	68.5○	188◑	69.1	131
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	81.6	84.0	45.4	122●	48.2	104●	86.7	81.2	65.5	113●	56.3◑	77.2●
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	89.8	84.8	51.1	205	57.9	220	91.4	86.1	56.2	207	77.0	147
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	**90.5**	85.7	53.1	204	58.3	**220**	92.0	**86.9**	57.6	207	77.3	149
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	79.6	77.0	43.1	201	54.1	180	85.6	81.7	52.0	198	73.6	128
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	89.3	85.0	50.9	203	57.9	219	91.2	86.2	56.2	206	77.6	147
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	90.3	81.6	49.8	203	56.7	219	89.0	83.8	57.7	201	75.6	142
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	87.9	81.5	46.8	**206**	56.6	209	88.3	83.6	53.4	202	74.8	138
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	50.7	75.1	51.0	118	23.7	62.1	37.5	74.7	47.4	118	−1.84	37.4
GMV	SAMPLE	65.0	58.5	53.5	−219	23.2	9.14	78.4	66.6	81.6	−208	55.7	70.1
GMV	LW2004	70.0	60.7	61.9	−186	29.1	24.3	76.2	67.7	82.9	−187	61.1	75.5
GMV	SSHH	69.2	59.6	62.4	−214	26.1	16.1	77.8	67.1	**83.2**	−206	58.2	73.0
GMVRF	SAMPLE	67.6	52.6	41.8	205	27.1	36.3	66.3	43.5	67.0	214	33.6	27.3
GMVRF	LW2004	65.0	52.0	47.2	198	24.3	38.0	61.7	43.9	68.4	211	33.8	39.8
GMVRF	SSHH	69.2	52.1	47.9	205	25.6	42.6	65.3	43.9	68.9	**214**	34.1	35.4
EW	/	53.1	46.1	56.2	−86.7	13.4	13.4	53.3	48.9	59.1	−83.9	18.6	24.4
EWRF	/	45.9	41.9	38.2	132	−8.45	−13.5	40.7	33.3	48.4	108	2.12	−4.06
MAXSER	/	87.4	78.6	50.3	185	51.8	153	83.8	80.6	40.3	181	45.7	97.0
UPSA	/	90.1	76.1	43.3	173	**68.4**	179	**93.2**	82.4	59.5	192	67.8	138
KNS	/	72.2	75.5	42.8	186	41.7	204	79.7	80.6	53.2	188	69.0	138
F+A	/	70.1	78.9	41.6	172	49.0	205	66.6	85.2	54.0	183	**81.6**	**153**

Notes. This table reports the annualized net out-of-sample Sharpe ratio, in percentage points, for the 21 portfolio strategies described in Section 4.6.2 and listed in Table 4.2. The first nine of them are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using rotation-equivariant covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}'$ for different choices of the diagonal matrix $\boldsymbol{D}$. EUL and EUNL are the linear and nonlinear shrinkage estimators maximizing the EU introduced in this chapter. We report results across the six datasets described in Table 4.1 and we follow the out-of-sample methodology detailed in Section 4.6.3. The net out-of-sample Sharpe ratio is computed using rolling windows with sample sizes $T = 120$ and $T = 240$ months and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 3$ in Panel (a) and $\gamma = 10$ in Panel (b). We compute two-sided p-values for the statistical test of the difference between the Sharpe ratio of two pairs of strategies: EUL vs. MSEL and EUNL vs. MSENL. We generate 10,000 bootstrap samples using the stationary block bootstrap approach of Politis and Romano (1994) with an average block size of five, and then use the methodology of Ledoit and Wolf (2008, Remark 3.2) to produce the resulting p-values. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

Table 4.7: Annualized net out-of-sample utility in empirical data (in percentage points): elliptical versus normal assumption

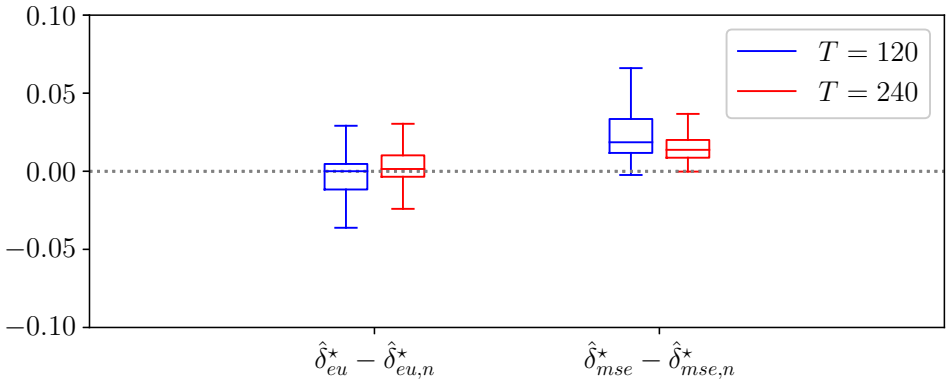
		$T = 120$						$T = 240$					
		Dataset						Dataset					
Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM	25 SBM	10 MOM	25 OPIN	13 JKP	16 ANOM	46 ANOM
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	−8.75	4.00	−49.1	16.3	−8.89	−71.9	0.19	6.66	−19.5	48.8	2.75	−217
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL-N	−23.6	−0.37	−70.1	−13.2	−16.0	−240	−10.1	4.26	−26.0	24.8	−2.13	−349
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL	9.12	10.1	−4.58	42.0	4.59	61.4	11.9	10.4	3.58	55.2	7.92	5.28
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL-N	9.35	9.99	−4.95	40.7	4.69	61.5	11.3	9.87	3.51	54.3	7.93	5.23
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	−1.93	6.12	−37.9	24.3	−5.30	−13.7	4.38	7.93	−15.5	53.3	4.73	−159
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL-N	−4.30	5.22	−43.4	20.5	−6.88	−40.2	2.16	7.24	−17.9	50.1	3.47	−194
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	9.36	11.0	3.42	20.5	2.82	10.5	12.2	11.0	6.88	16.6	3.24	6.28
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL-N	11.9	−71.5	3.26	19.2	2.99	12.4	12.7	11.3	7.26	15.6	3.35	6.58
Panel (b): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	−2.72	1.18	−14.9	4.37	−2.69	−23.1	0.01	1.99	−5.90	14.3	0.81	−66.7
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL-N	−7.24	−0.13	−21.2	−4.72	−4.83	−75.7	−3.14	1.26	−7.85	7.02	−0.66	−108
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL	2.73	3.03	−1.39	12.5	1.38	18.4	3.58	3.12	1.07	16.5	2.38	1.51
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL-N	2.79	2.99	−1.50	12.1	1.41	18.4	3.39	2.96	1.05	16.2	2.38	1.50
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	−0.64	1.82	−11.5	6.86	−1.60	−4.98	1.28	2.37	−4.70	15.7	1.41	−48.9
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL-N	−1.36	1.55	−13.1	5.68	−2.08	−13.2	0.60	2.16	−5.41	14.7	1.03	−59.7
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	2.80	3.31	1.03	6.15	0.85	3.15	3.66	3.28	2.06	4.99	0.97	1.89
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL-N	3.58	−11.5	0.98	5.74	0.90	3.73	3.81	3.39	2.18	4.69	1.00	1.98

Notes. This table reports the annualized net-of-cost out-of-sample utility, in percentage points, delivered by the linear shrinkage intensities $\delta^{\star}_{mse}$ and $\delta^{\star}_{eu}$ studied in Propositions 4.2 and 4.3, as well as the nonlinear shrinkage intensities $\delta^{\star}_{mse,i}$ and $\delta^{\star}_{eu,i}$ studied in Propositions 4.4 and 4.5. For each shrinkage intensity, we use two estimation methods. First, the method introduced in Section 4.5 of the main body of the chapter that assumes a multivariate elliptical distribution for the asset returns (MSEL, EUL, MSENL, and EUNL). Second, an alternative method in which we assume a multivariate normal distribution for the asset returns, i.e., we set $(k_1, k_2, k_3) = (1, 1, 1)$ and $\kappa = 0$ (MSEL-N, EUL-N, MSENL-N, and EUNL-N). We report results across the six datasets described in Table 4.1 and we follow the out-of-sample methodology detailed in Section 4.6.3. The net out-of-sample utility is computed using rolling windows with sample sizes $T = 120$ and $T = 240$ months and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 3$ in Panel (a) and $\gamma = 10$ in Panel (b).

4.F Empirical performance under normality

In this section, we test whether estimating the MSE-optimal and EU-optimal linear and nonlinear shrinkage intensities under the multivariate elliptical distribution, as we do in the main body of the chapter, makes a substantial difference compared to estimating them under the simpler multivariate normal distribution. Specifically, in Table 4.7, we report the annualized net out-of-sample utility delivered by the linear shrinkage intensities $\delta^{\star}_{mse}$ and $\delta^{\star}_{eu}$ studied in Propositions 4.2 and 4.3, as well as the nonlinear shrinkage intensities $\delta^{\star}_{mse,i}$ and $\delta^{\star}_{eu,i}$ studied in Propositions 4.4 and 4.5, under two estimation methods. First, the method introduced in Section 4.5 of the main body of the chapter, i.e., assuming a multivariate elliptical distribution for the asset returns (MSEL, EUL, MSENL, and EUNL). Second, an alternative method in which we assume a multivariate normal distribution for the asset returns, i.e., we set $(k_1, k_2, k_3) = (1, 1, 1)$ and $\kappa = 0$ (MSEL-N, EUL-N, MSENL-N, and EUNL-N).

Figure 4.9: Comparison of estimated linear shrinkage intensities under elliptical and normal distributions



Notes. This figure depicts boxplots of the difference between the linear shrinkage intensities estimated under the multivariate elliptical distribution, $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$, relative to those estimated under the multivariate normal distribution, $\hat{\delta}^{\star}_{eu,n}$ and $\hat{\delta}^{\star}_{mse,n}$, i.e., we set $(k_1, k_2, k_3) = (1, 1, 1)$ and $\kappa = 0$. The estimation methodology is detailed in Section 4.5 for $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$, and in Section 4.F for $\hat{\delta}^{\star}_{eu,n}$ and $\hat{\delta}^{\star}_{mse,n}$. The estimated shrinkage intensities are obtained in rolling estimation windows of either $T = 120$ months (in blue) or 240 months (in red), and the boxplots are averaged across the six datasets listed in Table 4.1. The horizontal gray dashed line corresponds to the zero level.

We observe from Table 4.7 that the MSE-optimal shrinkage intensities underperform systematically and significantly when computed under the normal distribution relative to the elliptical distribution. This is because the MSE leads to rather small shrinkage intensities, and thus, shrinking even less by assuming a normal distribution rather than elliptical worsens portfolio performance even more. However, when turning to the EU-optimal shrinkage intensities, we find that the performance they deliver is much more robust to the distributional assumption, as they perform similarly under the normal and elliptical distribution overall.

This intuition is confirmed in Figure 4.9, in which we depict boxplots of the difference between the estimated linear shrinkage intensities under the elliptical distribution, $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$, relative to those estimated under the normal distribution, $\hat{\delta}^{\star}_{eu,n}$ and $\hat{\delta}^{\star}_{mse,n}$. The estimated shrinkage intensities are obtained in rolling estimation windows of either $T = 120$ or 240 months, and the boxplots are averaged across the six datasets we consider. We observe that $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{eu,n}$ are essentially equal on average, whereas $\hat{\delta}^{\star}_{mse}$ is consistently above $\hat{\delta}^{\star}_{mse,n}$.

4.G Proofs

Proof of Proposition 4.1

This proposition corresponds to Kan and Lassance (2025, Proposition 7).

Proof of Proposition 4.2

The linear shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.14) depends on the shrinkage intensity δ as $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = (1-\delta)\hat{\boldsymbol{\Sigma}} + \delta\hat{\sigma}_m^2\boldsymbol{I}_N$, where $\hat{\boldsymbol{\Sigma}}$ in (4.5) is the sample covariance matrix and $\hat{\sigma}_m^2 = \hat{\bar{\lambda}} = \frac{1}{N}\text{tr}(\hat{\boldsymbol{\Sigma}})$. Therefore, the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$, as defined in (4.12), is

$$\begin{aligned} N \times \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) &= \mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^2)\big] + \text{tr}(\boldsymbol{\Sigma}^2) - 2\mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}\boldsymbol{\Sigma})\big] \\ &= (1-\delta)^2\mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + \delta^2 N\mathbb{E}\big[\hat{\sigma}_m^4\big] + 2\delta(1-\delta)N\mathbb{E}\big[\hat{\sigma}_m^4\big] \\ &\quad + \text{tr}(\boldsymbol{\Sigma}^2) - 2(1-\delta)\text{tr}(\boldsymbol{\Sigma}^2) - 2\delta N\sigma_m^4 \\ &= (1-\delta)^2\mathbb{B}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + \delta^2\big(\text{tr}(\boldsymbol{\Sigma}^2) - N\sigma_m^4\big) + \delta(2-\delta)N\mathbb{B}\big[\hat{\sigma}_m^4\big], \\ &= (1-\delta)^2\mathbb{B}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + \delta^2 N(N-1)s\sigma_m^4 + \delta(2-\delta)N\mathbb{B}\big[\hat{\sigma}_m^4\big], \end{aligned} \tag{4.70}$$

where $\mathbb{B}$ is the bias operator and $s \in [0,1)$ is the sphericity parameter defined in (4.16). The value $s = 1$ would be obtained if and only if $\boldsymbol{\Sigma}$ were of rank one, which cannot be the case because we assume that $\boldsymbol{\Sigma}$ is positive definite. The δ minimizing $\text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$ in (4.70) is

$$\delta_{mse}^{\star} = \underset{\delta\in\mathbb{R}}{\text{argmin}}\,\text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \frac{\mathbb{B}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] - N\mathbb{B}\big[\hat{\sigma}_m^4\big]}{\mathbb{B}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + N(N-1)s\sigma_m^4 - N\mathbb{B}\big[\hat{\sigma}_m^4\big]}. \tag{4.71}$$

Then, under the assumption that the i.i.d. sample of returns $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ is drawn from a multivariate elliptical distribution with an excess kurtosis equal to 3κ, we have from Ollila and Raninen (2019, Lemma 2) that

$$\mathbb{B}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] = N\sigma_m^4\left(\frac{N+1+(N-1)s}{T-1} + \frac{\kappa(N+2+2(N-1)s)}{T}\right), \tag{4.72}$$

$$N\mathbb{B}\big[\hat{\sigma}_m^4\big] = \sigma_m^4\left(\frac{2+2(N-1)s}{T-1} + \frac{\kappa(N+2+2(N-1)s)}{T}\right). \tag{4.73}$$

Plugging (4.72)–(4.73) into (4.71) yields Equation (4.15). It is straightforward to check that $s = 0$ when $\boldsymbol{\Sigma} = c\boldsymbol{I}_N$ for some $c > 0$, or equivalently, $\boldsymbol{\Lambda} = c\boldsymbol{I}_N$. This completes the proof.

Proof of Proposition 4.3

Under the assumptions of the proposition, we have from Kan and Lassance (2025, Proposition 7) that, for $T \geq N+3$,

$$\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}\big] = \frac{k_1(T-1)}{\gamma(T-N-2)}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\mu} = \frac{k_1(T-1)}{\gamma(T-N-2)}\boldsymbol{V}\left[(1-\delta)\boldsymbol{\Lambda} + \delta\bar{\lambda}\boldsymbol{I}_N\right]^{-1}\boldsymbol{V}'\boldsymbol{\mu}, \tag{4.74}$$

where k_1 is defined in (4.7). Therefore, the expected out-of-sample mean return of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ is

$$\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\mu}\big] = \frac{k_1(T-1)}{\gamma(T-N-2)}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2, \tag{4.75}$$

where the vector $\boldsymbol{f}(\delta)$ is defined in (4.18) and $\theta_{PC_i}^2 = (\boldsymbol{v}_i'\boldsymbol{\mu})^2/\lambda_i$. Next, from Kan and Lassance (2025, Proposition 7), we have that the expected out-of-sample variance of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ is

$$\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D}}\big] = \frac{1}{\gamma^2}\left(k_2\boldsymbol{\mu}'\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\mu} + \frac{k_3}{T}\mathrm{tr}\big(\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\Sigma}\big)\right), \tag{4.76}$$

where k_2 and k_3 are defined in (4.8)–(4.9) and $\mathbb{E}_{\mathcal{W}}$ stands for the expectation under the Wishart distribution for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$, i.e. $(T-1)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} \sim \mathcal{W}_N(T-1, \boldsymbol{\Sigma}_{\boldsymbol{D}})$. From, e.g., Haff (1979), we then have, for $T \geq N+5$,

$$\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big] = \frac{(T-1)^2\left[(T-N-2)\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\mathrm{tr}\big(\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\big)\right]}{(T-N-1)(T-N-2)(T-N-4)}, \tag{4.77}$$

where

$$\mathrm{tr}\big(\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\big) = \sum_{i=1}^{N} f_i(\delta). \tag{4.78}$$

Therefore, the two terms composing $\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D}}\big]$ in (4.76) become

$$\boldsymbol{\mu}'\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\mu} = \frac{(T-1)^2}{(T-N-1)(T-N-4)}\sum_{i=1}^{N}\theta_{PC_i}^2\left(f_i(\delta)^2 + \frac{f_i(\delta)\sum_{j=1}^{N}\boldsymbol{f}_j(\delta)}{T-N-2}\right), \tag{4.79}$$

$$\mathrm{tr}\big(\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\Sigma}\big) = \frac{(T-1)^2}{(T-N-1)(T-N-4)}\sum_{i=1}^{N}\left(f_i(\delta)^2 + \frac{f_i(\delta)\sum_{j=1}^{N}\boldsymbol{f}_j(\delta)}{T-N-2}\right). \tag{4.80}$$

Plugging (4.79)–(4.80) into (4.76), and then using (4.75) (4.76), the EU delivered by $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ becomes

$$\begin{aligned}\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = {} & \frac{k_1(T-1)}{\gamma(T-N-2)}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2 \\ & - \frac{1}{2\gamma}\frac{(T-1)^2}{(T-N-1)(T-N-4)}\sum_{i=1}^{N}\left(f_i(\delta)^2 + \frac{f_i(\delta)\sum_{j=1}^{N}\boldsymbol{f}_j(\delta)}{T-N-2}\right)\left(k_2\theta_{PC_i}^2 + \frac{k_3}{T}\right),\end{aligned} \tag{4.81}$$

which, given the matrix $\boldsymbol{A}$ in (4.19), is equivalent to (4.17). This completes the proof.

Proof of Proposition 4.4

We first derive the optimal shrinkage of the sample eigenvalues, from which we can then derive the shrinkage of their inverse. That is, we start from the formulation $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}(\boldsymbol{I}_N - \boldsymbol{\Delta})\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}'$ with $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_N)$, find the optimal δ_i's, and infer the shrinkage intensities on the inverse eigenvalues using $1 - \frac{1}{1-\delta_i}$. We can start from the sample eigenvalues instead of their inverse without loss of generality because each eigenvalue has its own shrinkage intensity δ_i, so that we will obtain the same optimal eigenvalues in both cases.

Now, the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}(\boldsymbol{I}_N - \boldsymbol{\Delta})\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}'$ is

$$N \times \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^2)\big] + \text{tr}(\boldsymbol{\Sigma}^2) - 2\text{tr}\big(\mathbb{E}[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}]\boldsymbol{\Sigma}\big). \tag{4.82}$$

Under the assumptions of the proposition, we have $\mathbb{E}[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}] = \boldsymbol{\Sigma}_{\boldsymbol{D}}$ and, from Ollila and Raninen (2019, Lemma 2),

$$\mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^2)\big] = \left(\frac{1}{T-1} + \frac{\kappa}{T}\right)\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}})^2 + \left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}}^2). \tag{4.83}$$

Letting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)'$, we have

$$\text{tr}\big(\mathbb{E}[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}]\boldsymbol{\Sigma}\big) = (\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}^2, \tag{4.84}$$

$$\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}})^2 = (\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}), \tag{4.85}$$

$$\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}}^2) = (\mathbf{1}_N - \boldsymbol{\delta})'\text{diag}(\boldsymbol{\lambda}^2)(\mathbf{1}_N - \boldsymbol{\delta}). \tag{4.86}$$

where $\boldsymbol{\Lambda}^2 = \text{diag}(\boldsymbol{\lambda}^2)$. Therefore, $\text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$ in (4.82) becomes

$$\begin{aligned}
&N \times \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) \\
&= \left(\frac{1}{T-1} + \frac{\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}) + \left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}) + \mathbf{1}_N'\boldsymbol{\lambda}^2 - 2(\mathbf{1}_N - \boldsymbol{\delta} \\
&= \left(\frac{1}{T-1} + \frac{\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}) + \left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}) + \boldsymbol{\delta}'\boldsymbol{\Lambda}^2\boldsymbol{\delta}.
\end{aligned} \tag{4.87}$$

The problem we want to solve is then to find the vector $\boldsymbol{\delta}$ that minimizes $\text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$ subject to the constraint that $\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}})$, which is equivalent to $\boldsymbol{\delta}'\boldsymbol{\lambda} = 0$. That is,

$$\begin{aligned}
\boldsymbol{\delta}^\star = \underset{\boldsymbol{\delta}\in\mathbb{R}^N}{\text{argmin}} &\left(\frac{1}{T-1} + \frac{\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}) + \left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}) + \boldsymbol{\delta}'\boldsymbol{\Lambda}^2\boldsymbol{\delta} \\
&\text{subject to } \boldsymbol{\delta}'\boldsymbol{\lambda} = 0.
\end{aligned} \tag{4.88}$$

We can show that the solution to Problem (4.88) is

$$\boldsymbol{\delta}^\star = \frac{\frac{1}{T-1} + \frac{2\kappa}{T}}{1 + \frac{1}{T-1} + \frac{2\kappa}{T}}\left(\mathbf{1}_N - \bar{\lambda}\boldsymbol{\lambda}^{-1}\right) \quad \Leftrightarrow \quad \delta_i^\star = \frac{\frac{1}{T-1} + \frac{2\kappa}{T}}{1 + \frac{1}{T-1} + \frac{2\kappa}{T}}(1 - \bar{\lambda}/\lambda_i). \tag{4.89}$$

Finally, the shrinkage intensities on the inverse sample eigenvalues are

$$\delta^{\star}_{mse,i} = 1 - \frac{1}{1-\delta^{\star}_i} = \frac{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\bar{\lambda} - \lambda_i)}{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)\bar{\lambda} + \lambda_i}. \tag{4.90}$$

It is clear that $\delta^{\star}_{mse,i} < 1$, where the inequality is strict because $\lambda_i > 0$ by assumption that $\boldsymbol{\Sigma}$ is positive definite. It is also easy to show that $\delta^{\star}_{mse,i} \to 1$ as $\lambda_i \to 0$, $\delta^{\star}_{mse,i} > 0$ if $\lambda_i < \bar{\lambda}$, and $\delta^{\star}_{mse,i} \leq 0$ if $\lambda_i \geq \bar{\lambda}$. This completes the proof.

Proof of Proposition 4.5

Part 1. Under the assumptions of the proposition, the EU of the shrinkage mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ in (4.26), with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (4.21), is obtained in the same way as for the case of linear shrinkage in Equation (4.17).

Part 2. Then, given the expression of $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (4.26), the optimal vector of shrinkage intensities $\boldsymbol{\delta}$ is

$$\boldsymbol{\delta}^{\star}_{eu} = \underset{\boldsymbol{\delta}\in\mathbb{R}^N}{\operatorname{argmax}}\, \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \mathbf{1}_N - \frac{T-N-4}{T-1} k_1 \boldsymbol{A}^{-1}\boldsymbol{\theta}^2_{PC}. \tag{4.91}$$

To derive a closed-form expression for $\boldsymbol{A}^{-1}$, we use the following Lemma.

Lemma 4.1. *Let $\boldsymbol{b} = (b_1, \ldots, b_N)'$ be an $N \times 1$ vector and c a constant. Define the $N \times N$ matrix $\boldsymbol{B}$ as*

$$\boldsymbol{B}_{ij} = \begin{cases} b_i & \text{if } i = j, \\ cb_i & \text{if } i \neq j. \end{cases} \tag{4.92}$$

Then, the inverse of $\boldsymbol{B}$ exists provided $c \notin \{-\frac{1}{N-1}, 1\}$ and $b_i \neq 0$ for all i, and is given by

$$(\boldsymbol{B}^{-1})_{ij} = \begin{cases} \frac{1+(N-2)c}{(1-c)(1+(N-1)c)} b_i^{-1} & \text{if } i = j, \\ -\frac{c}{(1-c)(1+(N-1)c)} b_j^{-1} & \text{if } i \neq j. \end{cases} \tag{4.93}$$

Proof of Lemma 4.1. *We can rewrite $\boldsymbol{B}$ as*

$$\boldsymbol{B} = (1-c)\boldsymbol{D}_b + c\boldsymbol{b}\mathbf{1}'_N, \tag{4.94}$$

where $\boldsymbol{D}_b = diag(b)$. Using the Sherman–Morrison formula, we have

$$\boldsymbol{B}^{-1} = \frac{1}{1-c}\left(\boldsymbol{D}_b^{-1} - \frac{c\boldsymbol{D}_b^{-1}\boldsymbol{b}\mathbf{1}'_N\boldsymbol{D}_b^{-1}}{1-c+c\mathbf{1}'_N\boldsymbol{D}_b^{-1}\boldsymbol{b}}\right), \tag{4.95}$$

where $\mathbf{1}'_N\boldsymbol{D}_b^{-1}\boldsymbol{b} = N$ and $(\boldsymbol{D}_b^{-1}\boldsymbol{b}\mathbf{1}'_N\boldsymbol{D}_b^{-1})_{ij} = b_j^{-1}$. Therefore, the entries of $\boldsymbol{B}^{-1}$ are

$$\boldsymbol{B}^{-1}_{ii} = \frac{b_i^{-1}}{1-c}\left(1 - \frac{c}{1+(N-1)c}\right) = \frac{1+(N-2)c}{(1-c)(1+(N-1)c)} b_i^{-1}, \tag{4.96}$$

$$\boldsymbol{B}^{-1}_{ij} = -\frac{c}{(1-c)(1+(N-1)c} b_j^{-1}, \tag{4.97}$$

which corresponds to (4.93).

Using Lemma 4.1, the inverse of $\boldsymbol{A}$ in (4.19) is

$$\boldsymbol{A}_{ij}^{-1} = \begin{cases} \frac{(T-3)(T-N-1)}{(T-2)(T-N-2)}(k_2\theta_{PC_i}^2 + k_3/T)^{-1} & \text{if } i = j, \\ -\frac{(T-N-1)}{(T-2)(T-N-2)}(k_2\theta_{PC_j}^2 + k_3/T)^{-1} & \text{if } i \neq j. \end{cases} \tag{4.98}$$

Plugging (4.98) into (4.91) yields the desired result in (4.28), where $R_i \in [0,1]$ in (4.25) because $\theta_{PC_i}^2 \geq$, $k_2 \geq k_1$, and $k_1, k_2, k_3 > 0$. We have $\delta_{eu,i}^\star \geq 0$ for all i, the lower being achieved as $T \to \infty$ because, then, $R_i \to 1$, $\frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)} \to 1$, and $\frac{N}{T-2}\bar{R} \to 0$. Concerning an upper bound on $\delta_{eu,i}^\star$, note that

$$\delta_{eu,i}^\star = 1 + \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} \sum_{j \neq i}^{N} R_j - \frac{(T-3)(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} R_i. \tag{4.99}$$

Therefore, an upper bound on $\delta_{eu,i}^\star$ is

$$\delta_{eu,i}^\star \leq 1 + \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} \sum_{j \neq i}^{N} R_j \leq 1 + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} \bar{R}, \tag{4.100}$$

which is achieved when $R_i = 0$.

Part 3. Plugging $\boldsymbol{\delta}_{eu}^\star$ in (4.91) into (4.26) gives

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1^2(T-N-4)}{2\gamma(T-N-2)} (\boldsymbol{\theta}_{PC}^2)' \boldsymbol{A}^{-1} \boldsymbol{\theta}_{PC}^2. \tag{4.101}$$

Given the expression for $\boldsymbol{A}^{-1}$ in (4.98), we have

$$\begin{aligned} (\boldsymbol{\theta}_{PC}^2)' \boldsymbol{A}^{-1} \boldsymbol{\theta}_{PC}^2 &= \frac{T-N-1}{k_1(T-N-2)} \sum_{i=1}^{N} \left(R_i - \frac{N}{T-2}\bar{R} \right) \theta_{PC_i}^2 \\ &= \frac{T-1}{k_1(T-N-4)} \sum_{i=1}^{N} (1 - \delta_{eu,i}^\star) \theta_{PC_i}^2, \end{aligned} \tag{4.102}$$

and plugging this into (4.101) yields (4.30). Clearly, the EU obtained with $\boldsymbol{\delta}_{eu}^\star$ is larger than that obtained when constraining $\boldsymbol{\delta} = \delta \mathbf{1}_N$ and finding the optimal δ. To find the EU in the latter case, note that plugging $\boldsymbol{\delta} = \delta \mathbf{1}_N$ into (4.26) yields

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-1)(1-\delta)\theta^2}{\gamma(T-N-2)} - \frac{(T-1)^2(1-\delta)^2 \mathbf{1}_N' \boldsymbol{A} \mathbf{1}}{2\gamma(T-N-2)(T-N-4)}, \tag{4.103}$$

where $\theta^2 = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ is the maximum squared Sharpe ratio and $\mathbf{1}_N' \boldsymbol{A} \mathbf{1} = \frac{T-2}{T-N-1}(k_2\theta^2 + k_3 N/T)$. Therefore, the optimal $\delta_{eu,cst}^\star$ maximizing (4.103) is given by (4.31). Plugging $\delta_{eu,cst}^\star$ into (4.103) then yields

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{(T-N-1)(T-N-4)}{2\gamma(T-2)(T-N-2)} \left(\frac{k_1^2\theta^4}{k_2\theta^2 + k_3 N/T} \right) \geq 0, \tag{4.104}$$

which is equivalent to the lower bound in (4.30) because $\theta^2 = \sum_{i=1}^{N} \theta_{PC_i}^2$.

Part 4. From (4.28), the condition $\delta_{eu,i}^{\star} < 1$ is equivalent to

$$\begin{aligned} \delta_{eu,i}^{\star} < 1 \;&\Leftrightarrow\; \frac{k_1 \theta_{PC_i}^2}{k_2 \theta_{PC_i}^2 + k_3/T} > \frac{N}{T-2}\bar{R} \\ &\Leftrightarrow\; \theta_{PC_i}^2 \left(k_1 - \frac{k_2 N}{T-2}\bar{R} \right) > \frac{k_3 \frac{N}{T}}{T-2}\bar{R}. \end{aligned} \tag{4.105}$$

Assuming $k_1/k_2 > \frac{N}{T-2}\bar{R}$, we have $k_1 - \frac{k_2 N}{T-2}\bar{R} > 0$, and thus the condition (4.105) simplifies to $\theta_{PC_i}^2 > \bar{\theta}_{PC}^2$ in (4.32). Otherwise, $\delta_{eu,i}^{\star} \geq 1$. This completes the proof.

Proof of Proposition 4.6

The EU of the shrinkage mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, with $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (4.51), is

$$\begin{aligned} \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) &= \frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\mu} - \frac{1}{2\gamma}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\mu} + \frac{1}{T}\mathrm{tr}\big(\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\big) \right) \\ &= \frac{1}{\gamma}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2 - \frac{1}{2\gamma}\sum_{i=1}^{N} f_i(\delta)^2 \left(\theta_{PC_i}^2 + \frac{1}{T} \right), \end{aligned} \tag{4.106}$$

where $\boldsymbol{f}(\delta)$ is defined in (4.18), which corresponds to (4.52). This completes the proof.

Proof of Proposition 4.7

Part 1. The EU of the shrinkage mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ in (4.55), with $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (4.54), is obtained in the same way as for the case of linear shrinkage in Equation (4.52).

Part 2. Given the expression for $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (4.55), it is easy to show that the optimal δ_i's are given by (4.56), and $\delta_{eu,i}^{\star} \in [0,1]$ because $\theta_{PC_i}^2 \geq 0$. To show that $\delta_{eu,i}^{\star}$ in (4.56) is lower than its counterpart under the sample covariance matrix in (4.28), we must show that

$$1 - \frac{\theta_{PC_i}^2}{\theta_{PC_i}^2 + 1/T} \leq 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i - \frac{N}{T-2}\bar{R} \right), \tag{4.107}$$

which holds because $\frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)} \leq 1$ and $\frac{\theta_{PC_i}^2}{\theta_{PC_i}^2+1/T} \geq R_i = \frac{k_1\theta_{PC_i}^2}{k_2\theta_{PC_i}^2+k_3/T}$ given that $k_1, k_2, k_3 \geq 1$ and $k_2 \geq k_1$.

Part 3. Plugging $\delta_{eu,i}^{\star}$ in (4.56) into (4.55), we directly obtain that the EU delivered by $\delta_{eu,i}^{\star}$ is given by (4.57). This completes the proof.

Proof of Proposition 4.8

Part 1. The difference between $\tilde{\delta}^{\star}_{eu,i}$ in (4.62) and $\delta^{\star}_{eu,i}$ in (4.28) is

$$\begin{aligned}
&\tilde{\delta}^{\star}_{eu,i} - \delta^{\star}_{eu,i} \\
&= 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)} R_i - \left(1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)} \left(R_i - \frac{N}{T-2}\bar{R}\right)\right) \\
&= \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}(R_i - \bar{R}),
\end{aligned} \tag{4.108}$$

which corresponds to (4.63). Therefore, it is clear that $\tilde{\delta}^{\star}_{eu,i} \geq \delta^{\star}_{eu,i}$ if $R_i \geq \bar{R}$, and $\tilde{\delta}^{\star}_{eu,i} < \delta^{\star}_{eu,i}$ otherwise.

Part 2. Plugging $\delta_i = \tilde{\delta}^{\star}_{eu,i}$ into (4.26), the EU delivered by the $\tilde{\delta}^{\star}_{eu,i}$'s is

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-N-1)(T-N-4)}{\gamma(T-2)(T-N-2)} \sum_{i=1}^{N} R_i \theta^2_{PC_i} - \frac{(T-N-1)^2(T-N-4)}{2\gamma(T-2)^2(T-N-2)} \boldsymbol{R}' \boldsymbol{A} \boldsymbol{R}, \tag{4.109}$$

where $\boldsymbol{R} = (R_1, \ldots, R_N)'$. Given the matrix $\boldsymbol{A}$ in (4.19), we have

$$\boldsymbol{R}' \boldsymbol{A} \boldsymbol{R} = \frac{k_1(T-2)}{T-N-1} \left(\left(1 - \frac{N}{T-2}\right) \sum_{i=1}^{N} R_i \theta^2_{PC_i} + \frac{N}{T-2} \bar{R} \theta^2 \right), \tag{4.110}$$

and thus, $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (4.109) simplifies to

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-N-1)(T-N-4)(T+N-2)}{2\gamma(T-2)^2(T-N-2)} \sum_{i=1}^{N} R_i \left(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2}\right), \tag{4.111}$$

which corresponds to (4.64). To show that this EU is positive, we need to show that $\sum_{i=1}^{N} R_i \big(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2}\big) \geq 0$. This inequality holds because, from (4.30), we have that

$$\sum_{i=1}^{N} (1 - \delta^{\star}_{eu,i}) \theta^2_{PC_i} \geq 0 \;\Leftrightarrow\; \sum_{i=1}^{N} R_i \left(\theta^2_{PC_i} - \frac{\theta^2}{T-2}\right) \geq 0 \;\Leftrightarrow\; \sum_{i=1}^{N} R_i \left(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2}\right) \geq 0. \tag{4.112}$$

This completes the proof.

Proof of Proposition 4.9

Let $\boldsymbol{\kappa} = \mathbf{1}_N - \boldsymbol{\delta}$. Then, Problem (4.65) is equivalent to

$$\begin{aligned}
\boldsymbol{\kappa}^{+}_{eu} = \underset{\boldsymbol{\kappa} \in \mathbb{R}^N}{\mathrm{argmax}} \; & \frac{k_1(T-1)}{\gamma(T-N-2)} \boldsymbol{\kappa}' \boldsymbol{\theta}^2_{PC} - \frac{(T-1)^2}{2\gamma(T-N-2)(T-N-4)} \boldsymbol{\kappa}' \boldsymbol{A} \boldsymbol{\kappa} \\
& \text{subject to} \;\; \kappa_i \geq 0 \; \forall i,
\end{aligned} \tag{4.113}$$

whose solution is

$$\boldsymbol{\kappa}_{eu}^{+} = \frac{T-N-4}{T-1} k_1 \boldsymbol{A}^{-1}(\boldsymbol{\theta}_{PC}^2 + \boldsymbol{\eta}) \;\Leftrightarrow\; \boldsymbol{\delta}_{eu}^{+} = \mathbf{1}_N - \frac{T-N-4}{T-1} k_1 \boldsymbol{A}^{-1}(\boldsymbol{\theta}_{PC}^2 + \boldsymbol{\eta}), \quad (4.114)$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N) \geq \mathbf{0}_N$ is the vector of Lagrange multipliers at the optimum. Given the expression for $\boldsymbol{A}^{-1}$ in (4.98), $\boldsymbol{\delta}_{eu}^{+}$ takes a form similar to (4.28) with R_i replaced by R_i^+ in (4.66), which gives the solution in (4.67). When $\eta_i > 0$, the constraint $\delta_i \leq 1$ is binding, and thus $\delta_{eu,i}^{+} = 1$. Otherwise, if $\eta_i = 0$, the constraint is non-binding, $R_i^+ = R_i$, and $\delta_{eu,i}^{+}$ relates to $\delta_{eu,i}^{\star}$ in (4.28) as

$$\begin{aligned}
\delta_{eu,i}^{+} &= 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)} \left(R_i^+ - \frac{N}{T-2} \bar{R}^+ \right) \\
&= 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)} \left(R_i - \frac{N}{T-2} \bar{R}^+ \right) \\
&= \delta_{eu,i}^{\star} + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)} \left(\bar{R}^+ - \bar{R} \right), \qquad (4.115)
\end{aligned}$$

which corresponds to (4.68), where $\bar{R}^+ \geq \bar{R}$ because $R_j^+ \geq R_j$ for all j. This completes the proof.

Chapter 5

Conclusion

This thesis has been organized as a collection of three research papers, one per chapter. In this last chapter, we provide a general conclusion to the thesis, based on all chapters. In Section 5.1, we summarize the main results introduced in Chapters 2, 3 and 4. Then, in Section 5.2, we list several avenues for future research.

5.1 Summary of main results

This thesis investigates how investors can perform *optimal portfolio selection under parameter uncertainty*. Indeed, while the foundational mean-variance framework remains central in portfolio theory, its practical implementation is severely challenged by the need to estimate key parameters such as the expected returns and the covariance matrix of asset returns. This estimation challenge, known as *parameter uncertainty*, is of particular importance as ignoring it often leads to poor out-of-sample performance.

The objective of this thesis has been to design practical, theoretically grounded and empirically robust portfolio strategies which account for parameter uncertainty. Using expected out-of-sample utility (EU) as a unifying criterion, the thesis presents three complementary chapters, each offering a different approach to addressing parameter uncertainty. In all three chapters, an extensive empirical analysis is carried out in order to ensure the practical relevance of the methodologies introduced for mean-variance investors.

In Chapter 2, we investigate the portfolio combination problem and provide several insights. First, which portfolios to combine? We show that combining the SMV (optimal without estimation risk) with the EW portfolio does deliver the best performance because it is not exposed to estimation risk, displays low turnover and offers large diversification gains. We also show that combining more than three funds delivers performance losses because of the estimation risk in combination coefficients. Second, how to estimate those combination coefficients? We show that constraints that are suboptimal in theory actually help in practice by alleviating estimation errors and are thus relevant to the investor. We propose a shrinkage estimation strategy to take advantage of these constraints. Third, what

is the impact of estimation risk? When N/T increases, combining the SMV with another robust portfolio becomes necessary to obtain a satisfying out-of-sample performance. In these settings, the EW portfolio is shown to perform particularly well because of its non-exposure to estimation risk. Fourth, when can we expect combinations to outperform the combined portfolios individually? The SMV portfolio is largely outperformed by all considered combinations. The risk-free asset is also consistently outperformed by the three-fund combination, but only when it is implemented without constraints on combination coefficients. Compared to the combination of the EW portfolio and the risk-free asset, the portfolio combination delivers large gains in most settings, but can underperform when there is little cross-sectional variation in mean returns and when estimation risk is large. Fifth, does the distributional assumption matter? Not so much according to our evidence, as in a rather general elliptical model for asset returns, the out-of-sample utility loss incurred by an investor who wrongly assumes that the data is i.i.d. normal is very small. Overall, we provide theoretical and practical insights into the portfolio combination problem, helping to inform the construction of more reliable portfolios in the presence of parameter uncertainty. We show that the shrinkage unconstrained combination of the SMV and EW portfolios offers the best all-around out-of-sample performance among the main portfolio combinations in the literature.

Chapter 3 shifts the focus from how to combine portfolios to a more structural question: How many assets should be included in a portfolio in the first place? In this chapter, we offer a novel perspective on the optimal *portfolio size* and challenge the conventional wisdom that investors should invest in as many assets as possible to diversify away idiosyncratic risk. Contrary to this conventional wisdom that broader diversification is always beneficial, we show that estimation risk introduces a tradeoff: Adding more assets does provide better diversification, which is beneficial to portfolio performance, but can also worsen out-of-sample performance due to an increasing exposure to estimation risk. Specifically, because of *parameter uncertainty*, we show that it is optimal to invest in a limited number N of assets. We thus offer an additional rational motivation for the concentrated portfolios typically observed in practice among individual investors, on top of already well-known explanations such as behavioral biases (prospect theory, overconfidence, geographical closeness and familiarity, private incentives, or preferences for higher moments) or other financial arguments (cost of information, transaction costs, short-sale constraints). We derive the optimal portfolio size that maximizes the EU, and propose a flexible implementation method to answer the natural follow-up question: How to select the N assets? We consider a range of different selection rules which are both simple to implement and sensible in the sense that they all have economic intuition. Empirically, we find that our optimally size-restricted portfolios consistently outperform their unrestricted counterparts, across a variety of selection rules and datasets, and even after accounting for transaction costs. For some asset selection rules, our optimally restricted portfolios

also systematically outperform the robust EW and minimum-variance portfolios, which are hard to beat in high dimension. Overall, our methodology renders portfolio theory valuable in the challenging but practically relevant case in which the sample size is close to the size of the investment universe.

In Chapter 4, we address the construction of the covariance matrix itself—a central input to any mean-variance portfolio. This problem is of high importance, highlighted by the warning of Ledoit and Wolf (2003a), who state that "nobody should be using the sample covariance matrix for the purpose of portfolio optimization" and instead, one should use a shrinkage estimator. Similarly, in this chapter, we recommend caution in using off-the-shelf shrinkage estimators. This is because existing linear and nonlinear shrinkage estimators aim to minimize the mean squared error (MSE) relative to the true covariance matrix, which is not the investor's objective. Instead, we design linear and nonlinear shrinkage estimators of the covariance matrix that optimize the mean-variance investor's expected out-of-sample utility (EU). These EU-based estimators shrink the eigenvalues of the sample covariance matrix more intensively than under the MSE, particularly so for those with low corresponding Sharpe ratios, and effectively reduce portfolio turnover. Our empirical analysis shows that these estimators outperform on average a wide array of leading alternatives, including MSE and PCA based estimators as well as naive strategies and recent proposals in the literature for estimating the mean-variance portfolio. This chapter not only advances the theory of shrinkage estimation for portfolio selection but also contributes to a broader shift in decision-making frameworks: from "predict-then-optimize" to "predict-and-optimize".

Collectively, the contributions of this thesis form a coherent set of approaches for dealing with parameter uncertainty in portfolio selection in the mean-variance framework, contributing with novel theoretical insights, practical methodologies and empirical evidence. They illustrate that properly accounting for parameter uncertainty can render portfolio theory viable even when it is most affected by estimation errors, i.e., in high dimension, and yield substantial out-of-sample performance gains relative to naive approaches.

5.2 Future research

The literature on portfolio selection is extensive, and this work opens several promising directions for future research. Natural and direct extensions of this work include, but are not limited to:

- Extending the framework to optimize the expected out-of-sample Sharpe ratio (ESR), which—unlike the expected out-of-sample utility—does not penalize leverage and is therefore a less conservative measure of performance. This is because the Sharpe ratio of a portfolio $\boldsymbol{w}$ is invariant to a scaling of the weights. The out-of-sample Sharpe ratio of an estimated portfolio $\hat{\boldsymbol{w}}$ given by $\hat{\boldsymbol{w}}'\boldsymbol{\mu}/\sqrt{\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}}$ is a random variable, similarly to

the out-of-sample utility. We could thus consider the expected out-of-sample Sharpe ratio as a performance measure, given by

$$\mathrm{ESR}(\hat{\boldsymbol{w}}) = \mathbb{E}\left[\frac{\hat{\boldsymbol{w}}'\boldsymbol{\mu}}{\sqrt{\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}}}\right]. \tag{5.1}$$

The ESR of any portfolio given by (5.1) is difficult to evaluate analytically. Therefore, we could follow DeMiguel et al. (2013) and use the following approximation,

$$\mathrm{ESR}(\hat{\boldsymbol{w}}) = \frac{\mathbb{E}[\hat{\boldsymbol{w}}'\boldsymbol{\mu}]}{\sqrt{\mathbb{E}[\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}]}}, \tag{5.2}$$

which is much easier to evaluate in practice. Generally speaking, strategies that significantly outperform in terms of mean-variance out-of-sample utility also tend to do so in terms of the Sharpe ratio. In fact, we show in Chapters 2 and 3 that even though the strategies are calibrated on the EU, they tend to deliver a satisfying out-of-sample Sharpe ratio as well. However, this is not systematic, as in Chapter 4 we show that the EU-based shrinkage methods do not outperform benchmarks as significantly when considering the out-of-sample Sharpe ratio relative to the utility. Therefore, an interesting idea would be to evaluate whether shrinkage intensities calibrated on the ESR would yield different results. For instance, in the linear shrinkage setting that we consider in Section 4.3, we could use the following shrinkage intensity,

$$\delta^{\star}_{esr} = \underset{\delta\in\mathbb{R}}{\operatorname{argmax}}\, \mathrm{ESR}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}). \tag{5.3}$$

This approach would yield several questions. Is $\delta^{\star}_{esr}$ larger or lower than $\delta^{\star}_{eu}$? Do they depend on the same underlying parameters, e.g. the squared Sharpe ratios of the principal components, and in a similar fashion? Can an analytical expression be derived for the expression in (5.3)? Given the conservative nature of the EU criterion, $\delta^{\star}_{eu}$ benefits from stabilizing portfolio weights and reducing turnover. Would that also hold for $\delta^{\star}_{esr}$?

Taking a step back, considering the Sharpe ratio as a performance measure is also valuable for practical reasons. Indeed, the mean-variance utility depends on the risk-aversion coefficient γ of the investor, which is known to be difficult to estimate in practice. While we show that our methods accommodate investors with any γ, it still could be that an investor under- or overestimates his or her own risk-aversion, and that may lead to disappointing out-of-sample performance. Because of this, practitioners often tend to avoid using utility-based methods and rather rely on methods based on indicators which are better understood by themselves and their customers, the Sharpe ratio being one good example.

- Designing new shrinkage estimators for both the mean vector and the covariance matrix simultaneously, with the same objective of maximizing the EU. In Chapter 4,

we design shrinkage estimators for the covariance matrix only, but still consider the sample estimator of the mean returns vector. This estimator is known to be particularly noisy, and we account for these estimation errors by maximizing the EU. However, it would still be interesting to consider a simultaneous shrinkage of both $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$. Following Jorion (1986), the shrinkage estimator of $\boldsymbol{\mu}$ could for instance be of the following form in a linear shrinkage setting,

$$\hat{\boldsymbol{\mu}}_{\delta_{\boldsymbol{\mu}}} = (1 - \delta_{\boldsymbol{\mu}})\hat{\boldsymbol{\mu}} + \delta_{\boldsymbol{\mu}}\boldsymbol{\theta}, \tag{5.4}$$

where $\boldsymbol{\theta}$ denotes the $N \times 1$ dimensional shrinkage target, and $\delta_{\boldsymbol{\mu}}$ is the shrinkage intensity towards that target. An example of a specification for $\boldsymbol{\theta}$ could for instance be $\hat{\bar{\mu}}\mathbf{1}_N$, where $\hat{\bar{\mu}}$ denotes the average of the sample average returns. In this specification, we shrink all estimates of the expected returns toward their average, in a similar fashion to what we do for the sample eigenvalues of the covariance matrix. Other specifications can be considered, for instance Jorion (1986) consider a shrinkage target set to the return of the minimum variance portfolio. Following the framework introduced in Chapter 4, we could then also shrink the covariance matrix,

$$\hat{\boldsymbol{\Sigma}}_{\delta_{\boldsymbol{\Sigma}}} = (1 - \delta_{\boldsymbol{\Sigma}})\hat{\boldsymbol{\Sigma}} + \delta_{\boldsymbol{\Sigma}}\hat{\bar{\lambda}}\boldsymbol{I}_N, \tag{5.5}$$

where $\delta_{\boldsymbol{\Sigma}}$ would be the shrinkage intensity, and then rely on a simultaneous shrinkage portfolio defined as

$$\hat{\boldsymbol{w}}(\delta_{\boldsymbol{\mu}}, \delta_{\boldsymbol{\Sigma}}) = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\delta_{\boldsymbol{\Sigma}}}^{-1}\hat{\boldsymbol{\mu}}_{\delta_{\boldsymbol{\mu}}}. \tag{5.6}$$

The objective would then to find the optimal pair of shrinkage intensities $(\delta_{\boldsymbol{\mu}}^\star, \delta_{\boldsymbol{\Sigma}}^\star)$ that jointly maximizes the EU of the $\hat{\boldsymbol{w}}(\delta_{\boldsymbol{\mu}}, \delta_{\boldsymbol{\Sigma}})$ portfolio,

$$(\delta_{\boldsymbol{\mu}}^\star, \delta_{\boldsymbol{\Sigma}}^\star) = \underset{(\delta_{\boldsymbol{\mu}}, \delta_{\boldsymbol{\Sigma}}) \in \mathbb{R}^2}{\operatorname{argmax}} \operatorname{EU}(\hat{\boldsymbol{w}}(\delta_{\boldsymbol{\mu}}, \delta_{\boldsymbol{\Sigma}})). \tag{5.7}$$

This formulation yields many questions. How would $\delta_{\boldsymbol{\mu}}^\star$ and $\delta_{\boldsymbol{\Sigma}}^\star$ compare to one another, as well as to $\delta_{eu}^\star$? The $(\delta_{\boldsymbol{\mu}}^\star, \delta_{\boldsymbol{\Sigma}}^\star)$ pair can only deliver a higher expected utility than $\delta_{eu}^\star$ in (4.20) alone since it is always possible to set $\delta_{\boldsymbol{\mu}} = 0$. Does this theoretical insight hold in an empirical and/or simulations setting where both shrinkage intensities must be estimated? This may not necessarily be the case, as we show in Chapter 2 that in the context of portfolio combinations, having to estimate two instead of one combination coefficients can have a significant impact on portfolio performance. Also, given the form of $\hat{\boldsymbol{w}}(\delta_{\boldsymbol{\mu}}, \delta_{\boldsymbol{\Sigma}})$, isn't there a risk of entanglement of $\delta_{\boldsymbol{\mu}}$ and $\delta_{\boldsymbol{\Sigma}}$? How can we ensure consistency between the shrinkage targets for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$? How does the performance of this strategy compare to using two robust estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ already available in the literature? How would this approach translate in the nonlinear setting?

In the same vein, another direct extension of our work would be to evaluate the EU-based shrinkage intensities for other portfolio strategies, such as the two- and three-fund combinations considered in Chapters 2 and 3. The portfolio combination coefficients and the shrinkage intensity would then have to be evaluated simultaneously, which raises similar questions as the ones mentioned above.

- Applying our shrinkage estimators in asset pricing, particularly to construct stochastic discount factors (SDFs) with smaller out-of-sample pricing errors. Indeed, the problems of finding the SDF in asset pricing and computing portfolio weights are very close to one another, and it is therefore possible that using our methods may improve the explanatory power of asset pricing models.

Addressing the limitations of our work is also a promising avenue for future research. The two main limitations of our approach lie in the assumptions it relies on. First, it requires a distributional assumption for future asset returns if one wishes to evaluate analytically the EU of any sample portfolio. Closed-form expressions are valuable as they are computationally efficient and provide insights into the problem at hand. Throughout this work, we postulate that asset returns are drawn from a multivariate elliptical distribution, which is a class of distribution allowing for fat tails. For instance, the multivariate student and Laplace distributions belong to this class. Although less restrictive than the traditional multivariate normal, this class of distribution still assumes symmetry, i.e., zero skewness. Relaxing this assumption may be of interest given that financial returns' distribution typically exhibit skewness, see e.g. Conrad et al. (2013). The second limitation we use is the stationary setting, where the distribution and parameters of asset returns are constant over time. While the elliptical distribution is relatively flexible by allowing for fat tails, and these assumptions do not prevent strong empirical performance, relaxing them could lead to further out-of-sample performance gains.

Finally, a promising direction for future research lies in applying our methodology beyond portfolio selection. In particular, the forecast combination problem is very similar to the portfolio selection problem (Roccazzella et al., 2022; Wang et al., 2023), suggesting that our framework could be adapted to enhance forecasting accuracy and reduce the variance of out-of-sample forecast errors. For example, it could build upon and extend the contributions of Blanc and Setzer (2020).

References

Adams, Z., Füss, R. and Glück, T. (2017). Are correlations constant? Empirical and theoretical results on popular correlation models in finance. *Journal of Banking and Finance*, *84*, pp. 9–24.

Ang, A., Hodrick, R., Xing, Y. and Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, *61*(1), pp. 259–299.

Ao, M., Li, Y. and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies*, *32*(7), pp. 2890–2919.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), pp. 191–221.

Barroso, P. and Saxena, K. (2021). Lest we forget: Using out-of-sample forecast errors in portfolio optimization. *Review of Financial Studies*, *35*(3), pp. 1222–1278.

Behr, P., Guettler, A. and Miebs, F. (2013). On portfolio optimization: Imposing the right constraints. *Journal of Banking and Finance*, *37*(4), pp. 1232–1242.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*(5), pp. 28–43.

Blanc, S. and Setzer, T. (2020). Bias–variance trade-off and shrinkage of weights in forecast combination. *Management Science*, *66*(12), pp. 5720–5737.

Bodnar, T., Dette, H., Parolya, N. and Thorstén, E. (2022). Sampling distributions of optimal portfolio weights and characteristics in low and large dimensions. *Random Matrices: Theory and Applications*, *11*(1).

Bodnar, T., Gupta, A. K. and Parolya, N. (2014). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis*, *132*, pp. 215–228.

Bodnar, T., Gupta, A. K. and Parolya, N. (2016). Direct shrinkage estimation of large dimensional precision matrix. *Journal of Multivariate Analysis*, *146*, pp. 223–236.

Bodnar, T., Okhrin, Y. and Parolya, N. (2023). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, *41*(1), pp. 140–156.

Bodnar, T., Parolya, N. and Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research*, *266*(1), pp. 371–390.

Bodnar, T., Parolya, N. and Thorsén, E. (2024). Two is better than one: regularized shrinkage of large minimum variance portfolios. *Journal of Machine Learning Research*, *25*(1).

Boyle, P., Garlappi, L., Uppal, R. and Wang, T. (2012). Keynes meets Markowitz: The trade-off between familiarity and diversification. *Management Science*, *58*(2), pp. 253–272.

Branger, N., Lučivjanská, K. and Weissensteiner, A. (2019). Optimal granularity for portfolio choice. *Journal of Empirical Finance*, *50*, pp. 125–146.

Cai, Z., Cui, Z., Lassance, N. and Simaan, M. (2024). The economic value of mean squared error: Evidence from portfolio selection. *SSRN working paper*.

Callot, L., Caner, M., Önder, Ö. and Ulaşan, E. (2021). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*, *39*(2), pp. 520–531.

Calvet, L., Campbell, J. and Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, *115*(5), pp. 707–747.

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

Campbell, J. (2006). Household finance. *Journal of Finance*, *61*(4), pp. 1553–1604.

Chamberlain, G. (1983). A characterization of the distributions that imply mean-variance utility functions. *Journal of Economic Theory*, *29*(1), pp. 185–201.

Chen, J. and Yuan, M. (2016). Efficient portfolio selection in a large market. *Journal of Financial Econometrics*, *14*(3), pp. 496–524.

Chen, Y., Wiesel, A., Eldar, Y. C. and Hero, A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, *58*(10), pp. 5016–5029.

Chernov, M., Dahlquist, M. and Lochstoer, L. (2023). Pricing currency risks. *Journal of Finance*, *78*(2), pp. 693–730.

Chung, M., Lee, Y., Kim, J. H., Kim, W. C. and Fabozzi, F. (2022). The effects of errors in means, variances, and correlations on the mean-variance framework. *Quantitative Finance*, *22*(10), pp. 1893–1903.

Clements, A., Scott, A. and Silvennoinen, A. (2015). On the benefits of equicorrelation for portfolio allocation. *Journal of Forecasting*, *34*(6), pp. 507–522.

Conrad, J., Dittmar, R. F. and Ghysels, E. (2013). Ex ante skewness and expected stock returns. *Journal of Finance*, *68*(1), pp. 85–124.

Da, R., Nagel, S. and Xiu, D. (2024). The statistical limit of arbitrage. *NBER working paper*.

DeMiguel, V., Garlappi, L., Nogales, F. and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*(5), pp. 798–812.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009b). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, *22*(5), pp. 1915–1953.

DeMiguel, V., Martín-Utrera, A. and Nogales, F. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking and Finance*, *37*(8), pp. 3018–3034.

DeMiguel, V., Martín-Utrera, A. and Nogales, F. (2015). Parameter uncertainty in multiperiod portfolio optimization with transaction costs. *Journal of Financial and Quantitative Analysis*, *50*(6), pp. 1443–1471.

Du, J.-H., Guo, Y. and Wang, X. (2023). High-dimensional portfolio selection with cardinality constraints. *Journal of the American Statistical Association*, *118*(542), pp. 779–791.

El Karoui, N. (2010). High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *The Annals of Statistics*, *38*(6), pp. 3487–3566.

El Karoui, N. (2013). On the realized risk of high-dimensional Markowitz portfolios. *SIAM Journal on Financial Mathematics*, *4*, pp. 737–783.

Elmachtoub, A. and Grigas, P. (2022). Smart "Predict, then Optimize". *Management Science*, *68*(1), pp. 9–26.

Elton, E. and Gruber, M. (1973). Estimating the dependence structure of share prices–Implications for portfolio selection. *Journal of Finance*, *28*(5), pp. 1203–1232.

Engle, R., Ferstenberg, R. and Russell, J. (2012). Measuring and modeling execution cost and risk. *Journal of Portfolio Management*, *38*(2), pp. 14–28.

Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, *30*(2), pp. 212–228.

Fan, J., Zhang, J. and Yu, K. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, *107*(498), pp. 592–606.

Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, *159*(2), pp. 289–302.

Frazzini, A., Israel, R. and Moskowitz, T. J. (2018). Trading costs. *SSRN working paper*.

Frost, P. and Savarino, J. (1988). For better performance: Constrain portfolio weights. *Journal of Portfolio Management*, *15*(1), pp. 29–34.

Gandy, A. and Veraart, L. (2013). The effect of estimation in high-dimensional portfolios. *Mathematical Finance*, *23*(3), pp. 531–559.

Gao, J. and Li, D. (2013). Optimal cardinality constrained portfolio selection. *Operations Research*, *61*(3), pp. 745–761.

Goetzmann, W. and Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, *12*(3), pp. 433–446.

Haff, L. R. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, *9*(4), pp. 531–544.

Hansen, P., Lunde, H. and Nason, J. (2011). The model confidence set. *Econometrica*, *79*(2), pp. 453–497.

Hediger, S., Naf, J. and Wolf, M. (2023). R-NL: Covariance matrix estimation for elliptical distributions based on nonlinear shrinkage. *IEEE Transactions on Signal Processing*, *71*, pp. 1657–1668.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, *58*(4), pp. 1651–1684.

James, W. and Stein, C. (1961). 'Estimation with quadratic loss'. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 1961. University of California Press, pp. 361–379.

Javed, F., Mazur, S. and Ngailo, E. (2021). Higher order moments of the estimated tangency portfolio weights. *Journal of Applied Statistics*, *48*(3), pp. 517–535.

Jensen, T. I., Kelly, B. and Pedersen, L. H. (2023). Is there a replication crisis in finance? *Journal of Finance*, *78*(5), pp. 246–2518.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, *21*(3), pp. 279–292.

Kan, R. and Lassance, N. (2025). Optimal portfolio choice with fat tails and parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *forthcoming*.

Kan, R., Lassance, N. and Wang, X. (2024a). The distribution of sample mean-variance portfolio weights. *Random Matrices: Theory and Applications*, *13*(01), p. 2450002.

Kan, R. and Smith, D. (2008). The distribution of the sample minimum-variance frontier. *Management Science*, *54*(7), pp. 1364–1380.

Kan, R. and Wang, X. (2023). Optimal portfolio choice with unknown benchmark efficiency. *Management Science*, *70*(9), pp. 6117–6138.

Kan, R., Wang, X. and Zheng, X. (2024b). In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models. *Journal of Financial Economics*, *155*, p. 103837.

Kan, R., Wang, X. and Zhou, G. (2021). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science*, *68*(3), pp. 2047–2068.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *42*(3), pp. 621–656.

Kazak, E. and Pohlmeier, W. (2022). Bagged pretested portfolio selection. *Journal of Business & Economic Statistics*, *41*(4), pp. 1116–1131.

Kelly, B., Malamud, S., Pourmohammadi, M. and Trojani, F. (2023). Universal portfolio shrinkage. *NBER working paper*.

Kirby, C. and Ostdiek, B. (2012). It's all in the timing: Simple active portfolio strategies that outperform naïve diversification. *Journal of Financial and Quantitative Analysis*, *47*(2), pp. 437–467.

Koijen, R. and Yogo, M. (2019). A demand system approach to asset pricing. *Journal of Political Economy*, *127*(4), pp. 1475–1515.

Kourtis, A., Dotsis, G. and Markellos, R. N. (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking & Finance*, *36*(9), pp. 2522–2531.

Kozak, S., Nagel, S. and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, *135*(2), pp. 271–292.

Kremer, P. J., Lee, S., Bogdan, M. and Paterlini, S. (2020). Sparse portfolio selection via the sorted l1-Norm. *Journal of Banking & Finance*, *110*, p. 105687.

Lassance, N. and Martín-Utrera, A. (2024). Does the factor zoo pay off? A portfolio view on mispricing and the limited gains from new anomalies. *SSRN working paper*.

Lassance, N., Martín-Utrera, A. and Simaan, M. (2024a). The risk of expected utility under parameter uncertainty. *Management Science*, *70*(11), pp. 7644–7663.

Lassance, N., Vanderveken, R. and Vrins, F. (2024b). On the combination of naive and mean-variance portfolio strategies. *Journal of Business & Economic Statistics*, *42*(3), pp. 875–889.

Lassance, N., Vanderveken, R. and Vrins, F. (2025a). Optimal portfolio size under parameter uncertainty. *SSRN working paper*.

Lassance, N., Vanderveken, R. and Vrins, F. (2025b). Optimal shrinkage of the covariance matrix for portfolio selection. *Work in progress*.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), pp. 850–859.

Ledoit, O. and Wolf, M. (2003a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, *30*(4), pp. 110–119.

Ledoit, O. and Wolf, M. (2003b). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), pp. 603–621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), pp. 365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, *40*(2), pp. 1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, *139*, pp. 360–384.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*, *30*(12), pp. 4349–4388.

Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, *48*(5), pp. 3043–3065.

Ledoit, O. and Wolf, M. (2022a). Quadratic shrinkage for large covariance matrices. *Bernoulli*, *28*(3), pp. 1519–1547.

Ledoit, O. and Wolf, M. (2022b). The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, *20*(1), pp. 187–218.

Levy, H. and Levy, M. (2014). The benefits of differential variance-based constraints in portfolio optimization. *European Journal of Operational Research*, *234*(2), pp. 372–381.

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business & Economic Statistics*, *33*(3), pp. 381–392.

MacKinlay, A. and Pástor, L. (2000). Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial Studies*, *13*(4), pp. 883–916.

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

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.

Martellini, L. and Ziemann, V. (2010). Improved estimates of higher-order comoments and implications for portfolio selection. *Review of Financial Studies*, *23*(4), pp. 1467–1502.

Mattera, R. (2025). Improved precision matrix estimation for mean-variance portfolio selection. *Statistics*, *forthcoming*.

Merton, R. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), pp. 323–361.

Michaud, R. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(1), pp. 31–42.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *Journal of Finance*, *72*(4), pp. 1611–1643.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. New Jersey: John Wiley & Sons.

Novy-Marx, R. and Velikov, M. (Nov. 2015). A taxonomy of anomalies and their trading costs. *Review of Financial Studies*, *29*(1), pp. 104–147.

Okhrin, Y. and Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics*, *134*(1), pp. 235–256.

Ollila, E. and Raninen, E. (2019). Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions. *IEEE Transactions on Signal Processing*, *67*(10), pp. 2707–2719.

Pflug, G., Pichler, A. and Wozabal, D. (2012). The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking and Finance*, *36*(2), pp. 410–417.

Politis, D. and Romano, J. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), pp. 1303–1313.

Rencher, A. and Schaalje, G. B. (2008). *Linear Models in Statistics (2nd ed.)* New Jersey: John Wiley & Sons.

Roccazzella, F., Gambetti, P. and Vrins, F. (2022). Optimal and robust combination of forecasts via constrained optimization and shrinkage. *International Journal of Forecasting*, *38*(1), pp. 97–116.

Schuhmacher, F., Kohrs, H. and Auer, B. (2021). Justifying mean-variance portfolio selection when asset returns are skewed. *Management Science*, *67*(12), pp. 7812–7824.

Shi, F. S., Shu, L., He, F. and Huang, W. (2025). Improving minimum-variance portfolio through shrinkage of large covariance matrices. *Economic Modelling*, *144*.

Shi, F. S., Shu, L., Yang, A. and He, F. (2020). Improving minimum-variance portfolios by alleviating overdispersion of eigenvalues. *Journal of Financial and Quantitative Analysis*, *55*(8), pp. 2700–2731.

Simaan, M. and Simaan, Y. (2019). Rational explanation for rule-of-thumb practices in asset allocation. *Quantitative Finance*, *19*(12), pp. 2095–2109.

Stein, C. (1956). 'Inadmissibility of the usual estimator for the mean of a multivariate normal distribution'. *Proceedings of the third Berkeley symposium on mathematical statistics and probability*. Vol. 1, pp. 197–206.

Tu, J. and Zhou, G. (2004). Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics*, *72*(2), pp. 385–421.

Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, *99*(1), pp. 204–215.

Tyler, D. E. (1987). A distribution-free M-Estimator of multivariate scatter. *Annals of Statistics*, *15*(1), pp. 234–251.

Vanderschueren, T., Verdonck, T., Baesens, B. and Verbeke, W. (2022). Predict-then-optimize or predict-and-optimize? An empirical evaluation of cost-sensitive learning strategies. *Information Sciences*, *594*, pp. 400–415.

Wang, X., Hyndman, R. J., Li, F. and Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, *39*(4), pp. 1518–1547.

Yan, C. and Zhang, H. (2017). Mean-variance versus naïve diversification: The role of mispricing. *Journal of International Financial Markets, Institutions and Money*, *48*, pp. 61–81.

Yen, Y.-M. (2016). Sparse weighted-norm minimum variance portfolios. *Review of Finance*, *20*(3), pp. 1259–1287.

Yuan, M. and Zhou, G. (2024). Why naïve 1/N diversification is not so naïve, and how to beat it? *Journal of Financial and Quantitative Analysis*, *59*(8), pp. 3601–3632.

Zhao, Z., Ledoit, O. and Jiang, H. (2021). Risk reduction and efficiency increase in large portfolios: Gross-exposure constraints and shrinkage of the covariance matrix. *Journal of Financial Econometrics*, *21*(1), pp. 1–33.