SHRINK WITH PURPOSE: OPTIMAL COVARIANCE MATRIX ESTIMATION FOR PORTFOLIO SELECTION

Nathan Lassance, Rodolphe Vanderveken, Frédéric Vrins

LIDAM Discussion Paper LFIN
2025 / 02

UCLouvain

LFIN | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications

Shrink with Purpose: Optimal Covariance Matrix Estimation for Portfolio Selection*

Nathan Lassance[1] Rodolphe Vanderveken[1] Frédéric Vrins[1,2]

July 11, 2025

Abstract

We introduce analytical linear and nonlinear shrinkage estimators of the sample covariance matrix that are optimal for mean-variance portfolio choice. Unlike the classical estimators 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 in mean returns. Our estimators shrink the sample eigenvalues more intensively than conventional methods, and they especially diminish the contribution of principal components with small squared Sharpe ratios. By jointly estimating the covariance matrix and the optimal portfolio in one step, our method delivers significant empirical performance gains relative to the usual two-step shrinkage approach. Our portfolios also help reduce turnover and outperform recent regularized mean-variance portfolio strategies.

Keywords: estimation risk, linear shrinkage, mean-variance portfolio, nonlinear shrinkage, out-of-sample utility, parameter uncertainty.

JEL Classification: G11.

*(1) LIDAM/LFIN, UCLouvain, Belgium, (2) Decision science department, HEC Montréal, Canada. Email addresses: nathan.lassance@uclouvain.be, rodolphe.vanderveken@uclouvain.be, frederic.vrins@uclouvain.be. The authors thank Arnaud Germain, Marco Saerens, Majeed Simaan, Guofu Zhou, and participants at the 16th Actuarial and Financial Mathematics Conference, 17th German Probability and Statistics Days, 41st International Conference of the French Finance Association, 14th International Conference of the Financial Engineering and Banking Society, and 7th QFFE Quantitative Finance and Financial Econometrics Conference for their comments. This work was supported by the Fonds de la Recherche Scientifique (FNRS) under Grant Number J.0135.25 and T.0221.22. The authors are grateful to Nathan Sanglier for excellent research assistance.

1 Introduction

The central issue that renders mean-variance portfolio theory challenging in practice is parameter uncertainty, i.e., the mean and covariance matrix of future asset returns are unknown to the investor, which induces a typically substantial gap between in-sample and out-of-sample performance (Kan and Zhou, 2007; DeMiguel et al., 2009; Barroso and Saxena, 2021; Kan et al., 2024). Among various methods for alleviating this issue, the use of shrinkage estimators has become widespread, particularly for the covariance matrix 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 where the sample covariance matrix is shrunk toward a structured target estimator (Ledoit and Wolf, 2003a,b, 2004; Chen et al., 2010; Bodnar et al., 2014), e.g., the scaled identity matrix in Ledoit and Wolf (2004), a second era of nonlinear shrinkage followed where each eigenvalue of the sample covariance matrix is shrunk with its own intensity (Ledoit and Wolf, 2012, 2015, 2017, 2020, 2022b; Hediger et al., 2023).

The common theme among these linear and nonlinear shrinkage estimators is that they minimize a statistical loss function relative to the true covariance matrix, typically measured with the mean squared error (MSE), i.e., the Frobenius norm. That is, these methods provide generic estimators that can then be used for various applications; see Ledoit and Wolf (2022a) for a review of applications across different fields. However, for a mean-variance investor, what matters is not whether the shrinkage covariance matrix estimator is close to the true one. What truly matters is how well the estimator will perform out of sample in a portfolio selection exercise. Therefore, in this paper, we develop shrinkage covariance matrix estimators under a mean-variance portfolio loss function. Changing the loss function drastically modifies the optimal shrinkage of the covariance matrix and substantially improves the out-of-sample performance of mean-variance portfolios.

The choice of a statistical loss function versus a decision-specific loss function is of wide applicability. Elmachtoub and Grigas (2022) call for switching from the standard "predict-then-optimize" paradigm to "predict-and-optimize" in solving decision problems, i.e., designing estimators under loss functions related to the application of interest.[1] We answer their call and introduce analyt-

[1] 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.

ical finite-sample linear and nonlinear shrinkage estimators of the sample covariance matrix that optimize the mean-variance expected out-of-sample utility (EU).[2] Since Kan and Zhou (2007), the EU has been used extensively for portfolio choice under parameter uncertainty; see Lassance et al. (2024a) for a review. This is fundamentally different from minimizing the MSE, in particular because we shrink the covariance matrix to alleviate the impact of estimation errors on the mean-variance portfolio, which stem from both the mean and the covariance matrix.[3] Our analytical finite-sample approach builds on the multivariate elliptical distribution for the returns[4] and contrasts with recent data-driven high-dimensional approaches to mean-variance portfolio selection like MAXSER (Ao et al., 2019), the robust SDF of Kozak et al. (2020), and the universal portfolio shrinkage approximator (UPSA) by Kelly et al. (2023). Empirically, our mean-variance portfolios using EU-optimal shrinkage covariance matrices compare positively to these benchmarks, among others.

Other papers have developed covariance matrix estimators using portfolio loss functions. Bodnar et al. (2024) linearly shrink the sample covariance matrix to the identity, and Kazak et al. (2025) propose a parametric model for the covariance matrix, both so that the global-minimum-variance (GMV) portfolio has minimum out-of-sample variance. Unlike these papers, we focus on mean-variance portfolios. Ledoit and Wolf (2012) introduce nonlinear shrinkage estimators of the covariance matrix under the MSE, which Ledoit and Wolf (2017) show are also optimal under mean-variance portfolio loss functions. However, this equivalence assumes that mean returns are known. In contrast, we introduce shrinkage estimators that account for estimation errors stemming from both the sample covariance matrix and the sample mean. Finally, Kelly et al. (2023) introduce UPSA, which shrinks the sample covariance matrix nonlinearly via leave-one-out cross-validation to maximize the mean-variance portfolio's out-of-sample utility. A distinguishing feature of our estimators is that they are available in closed form and do not require cross-validation.

We start our theoretical analysis by linear shrinkage of the sample covariance matrix toward the scaled identity matrix, as in Ledoit and Wolf (2004), which amounts to shrinking the sample

[2]Some papers also shrink the inverse covariance matrix (Frahm and Memmel, 2010; Kourtis et al., 2012; Bodnar et al., 2016; Shi et al., 2020; Mattera, 2025). We shrink the covariance matrix directly, which is more standard. Under nonlinear shrinkage, there is no difference in our method between shrinking the covariance matrix or its inverse.

[3]If the covariance matrix were known, the MSE loss function would set a shrinkage intensity of zero. In contrast, as we show in Section C of the supplementary material, it is still optimal to shrink the covariance matrix intensively to maximize the EU in that case, because it helps reduce the impact of estimation errors in sample mean returns.

[4]We accommodate the multivariate elliptical distribution and not just the normal distribution by leveraging results in Kan and Lassance (2025), which extend Kan and Zhou (2007) to the multivariate elliptical distribution.

eigenvalues toward their mean. Whereas Ledoit and Wolf (2004) consider high-dimensional asymptotics 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 or to deliver the best portfolio performance out of sample.

We then turn to nonlinear shrinkage of sample eigenvalues introduced by Ledoit and Wolf (2012), which we formulate so that each inverse eigenvalue is shrunk toward zero with its own intensity. We derive in closed form the finite-sample MSE-optimal and EU-optimal vectors of shrinkage intensities, $\boldsymbol{\delta}^\star_{mse}$ and $\boldsymbol{\delta}^\star_{eu}$, unlike the usual high-dimensional asymptotic approach. We show that $\boldsymbol{\delta}^\star_{mse}$ shrinks the contribution of lower-variance PCs more intensively, whereas $\boldsymbol{\delta}^\star_{eu}$ does so for PCs with smaller squared Sharpe ratios. Like in linear shrinkage, the shrinkage intensities $\boldsymbol{\delta}^\star_{eu}$ are substantially larger than $\boldsymbol{\delta}^\star_{mse}$. In particular, the EU shrinks all inverse eigenvalues closer to zero, whereas the MSE increases small inverse eigenvalues. Although the EU substantially shrinks the contribution of PCs with small Sharpe ratios, their contribution does not vanish, unlike common PC-sparse mean-variance portfolios (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 demonstrate that the oracle EU-optimal shrinkage intensities deliver an EU that is always positive. We use simulations to corroborate this result under bona fide shrinkage intensities and to show that this EU is substantially larger than that under the MSE-optimal intensities.

Finally, we evaluate the empirical out-of-sample performance of our proposed mean-variance portfolios using bona fide MSE-optimal and EU-optimal shrinkage intensities. We compare them to numerous benchmarks including mean-variance and GMV portfolios under MSE-optimal, PC-sparse, and cross-validated shrinkage covariance matrices, equally weighted (EW) portfolios, and

[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 $\delta^\star_{mse} = 0.05$.

four mean-variance portfolios based on regularization, PCA, high-dimensional asymptotics, and cross-validation, namely, MAXSER (Ao et al., 2019), the robust SDF of Kozak et al. (2020), 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 out-of-sample utility 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. 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 a mean-variance portfolio with an annualized net utility 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]

There are three notable empirical feats achieved by our mean-variance portfolios. First, they have a turnover substantially lower than that of the mean-variance benchmark portfolios, due to the intensive shrinkage of small eigenvalues. Second, they deliver a consistently strong performance. In particular, our EU-optimal linear shrinkage estimator delivers a positive net utility in all considered configurations except one, and our nonlinear estimator does so in all configurations. The median gain in net utility obtained by using these two estimators is positive relative to all 19 benchmarks. Third, whereas using a robust estimator of the mean vector is systematically favorable under conventional estimators of the covariance matrix, it underperforms the sample mean estimator when using our EU-optimal covariance matrices. Therefore, using the sample mean estimator is no longer a concern when its estimation risk is incorporated into the design of the shrinkage covariance matrix.

The rest of the paper is structured as follows. Section 2 introduces the problem and notation. Sections 3 and 4 derive the MSE-optimal and EU-optimal linear and nonlinear shrinkage covariance matrix estimators. Section 5 introduces bona fide estimators of the shrinkage intensities and contains simulation results. Section 6 presents our empirical analysis. Section 7 concludes. The supplementary material contains other theoretical and empirical tests, and the proofs of all results.

[6] In total, we consider 21 portfolio strategies in our main results; see Table 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. The median EU gain across the 24 configurations is positive relative to all four benchmarks.

2 Presentation of the problem

In this section, we introduce some notation and the necessary background surrounding our problem.

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 $\boldsymbol{\Sigma}$. We have the eigendecomposition $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, where $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N]$ is the matrix of eigenvectors and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$, $\lambda_1 \geq \cdots \geq \lambda_N > 0$ is the matrix of eigenvalues sorted in descending order. Given a risk-aversion coefficient $\gamma \in (0, \infty)$, the mean-variance utility of a portfolio $\boldsymbol{w} \in \mathbb{R}^N$ is

$$U(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}. \tag{1}$$

The corresponding optimal mean-variance portfolio is

$$\boldsymbol{w}^\star = \operatorname*{argmax}_{\boldsymbol{w}\in\mathbb{R}^N} U(\boldsymbol{w}) = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{2}$$

which delivers the maximum utility given by

$$U(\boldsymbol{w}^\star) = \frac{\theta^2}{2\gamma} \quad \text{with} \quad \theta^2 = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{3}$$

The traditional plug-in estimator of $\boldsymbol{w}^\star$ is

$$\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}, \tag{4}$$

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,

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \quad \text{and} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})', \tag{5}$$

where $\hat{\boldsymbol{\Sigma}}^{-1}$ exists almost surely provided $T \geq N+1$. The plug-in portfolio $\hat{\boldsymbol{w}}^\star$ performs notoriously 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 $\hat{\boldsymbol{w}}^\star$, i.e., $\mathbb{E}[U(\hat{\boldsymbol{w}}^\star)]$, 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. For that purpose, we use the stochastic representation in El Karoui (2010, 2013), i.e.,

$$\boldsymbol{r}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau) \quad \Leftrightarrow \quad \boldsymbol{r}_t \overset{d}{=} \boldsymbol{\mu} + (\tau\boldsymbol{\Sigma})^{1/2}\boldsymbol{z}, \tag{6}$$

where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}_N, \boldsymbol{I}_N)$, $\boldsymbol{0}_N$ is an N-dimensional vector of zeros, $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, 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 1 *Assume the i.i.d. sample of returns $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ is drawn from the multivariate elliptical distribution in (6), i.e., $\boldsymbol{r}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau)$. Moreover, let $T \geq N + 5$, $\boldsymbol{M} = \boldsymbol{I}_T - \boldsymbol{1}_T\boldsymbol{1}_T'/T$, $\boldsymbol{Z}$ be a $T \times N$ matrix of independent standard normal variables, $\boldsymbol{T}$ be a diagonal matrix of T independent copies of $\tau^{1/2}$, and $\boldsymbol{T} \perp \boldsymbol{Z}$. Then, the EU delivered by $\hat{\boldsymbol{w}}^\star$ in (4) is*

$$\mathbb{E}[U(\hat{\boldsymbol{w}}^\star)] = \frac{T-1}{2\gamma(T-N-2)}\left[2k_1\theta^2 - \frac{(T-1)(T-2)}{(T-N-1)(T-N-4)}\left(k_2\theta^2 + k_3\frac{N}{T}\right)\right], \tag{7}$$

where k_1, k_2, and k_3 are functions of only N, T, and the distribution of τ:

$$k_1 = \frac{T-N-2}{N}\mathbb{E}\big[\mathrm{tr}\big((\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z})^{-1}\big)\big], \tag{8}$$

$$k_2 = \frac{(T-N-1)(T-N-2)(T-N-4)}{N(T-2)}\mathbb{E}\big[\mathrm{tr}\big((\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z})^{-2}\big)\big], \tag{9}$$

$$k_3 = \frac{(T-N-1)(T-N-2)(T-N-4)}{NT(T-2)}\mathbb{E}\big[\boldsymbol{1}_T'\boldsymbol{T}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{M}\boldsymbol{T}\boldsymbol{Z})^{-2}\boldsymbol{Z}'\boldsymbol{T}\boldsymbol{1}_T\big], \tag{10}$$

provided the expectations exist, and $k_2 \geq k_1$. Finally, k_1, k_2, and k_3 reach their lower bound of one when the sample of returns is drawn from a multivariate normal distribution.

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 $N/T \to c \in (0, 1)$, we have that $k_1 = k_3$, $k_2 \geq k_1^2 \geq 1$, and both k_1 and k_2 increase with c.

2.2 Conventional MSE-optimal 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[8]

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}', \tag{11}$$

where $\boldsymbol{D}$ is a diagonal matrix. In the linear shrinkage approach of Ledoit and Wolf (2004), the matrix $\boldsymbol{D}$ is parametrized as $\boldsymbol{D} = (1-\delta)\hat{\boldsymbol{\Lambda}} + \delta\xi\boldsymbol{I}_N$, where ξ and δ are parameters to calibrate. The oracle δ is the shrinkage intensity, and the oracle $\xi = \mu_\lambda = \frac{1}{N}\text{tr}(\boldsymbol{\Lambda})$ is the shrinkage target. 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. The parameters 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 $\boldsymbol{\Sigma}$:

$$\text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \mathbb{E}\Big[||\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma}||_F^2\Big] = \frac{1}{N}\mathbb{E}\Big[\text{tr}\big((\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma})^2\big)\Big]. \tag{12}$$

Remark 1 Ledoit and Wolf (2004, 2012, 2015, 2017, 2020) and Bodnar et al. (2014, 2016) do not need to evaluate the expectation in (12) because they minimize the nonrandom high-dimensional asymptotic limit of the Frobenius norm as $N \to \infty$, $T \to \infty$, and $N/T \to c < \infty$, $c \neq 1$. 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 find that their solution for the optimal matrix $\boldsymbol{D}$ is equivalent to that minimizing the asymptotic Frobenius norm.

2.3 Proposed EU-optimal approach to shrinking the covariance matrix

The common theme of the shrinkage methods in Section 2.2 is that they fit the true covariance matrix $\boldsymbol{\Sigma}$. However, for an investor, the aim is rather to obtain an estimator 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 so as to maximize the EU of the mean-variance portfolio, i.e.,[9]

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \mathbb{E}[U(\hat{\boldsymbol{w}}_{\boldsymbol{D}})] = \mathbb{E}\Big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D}}\Big] \quad \text{with} \quad \hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}. \tag{13}$$

[8] Kelly et al. (2023) refer to the class of estimators (11) as spectral shrinkage estimators.

[9] One may be concerned about using the sample mean $\hat{\boldsymbol{\mu}}$, even if we shrink the sample covariance matrix. However, in Section F of the supplementary material, we show empirically that once we shrink the covariance matrix with our EU-optimal estimators, which account for estimation errors in $\hat{\boldsymbol{\mu}}$, using the sample mean outperforms the robust Bayes-Stein mean estimator of Jorion (1986). We also find empirically that using our EU-optimal covariance matrices and the sample mean outperforms plugging existing robust estimators of both the mean and covariance matrix.

There are four main differences between the proposed EU-optimal shrinkage and the conventional MSE-optimal shrinkage. First, we calibrate the shrinkage intensities using a portfolio objective, the EU, rather than a statistical loss function. Second, we account for the impact that estimation errors stemming from the sample mean vector $\hat{\boldsymbol{\mu}}$ have on portfolio performance. Third, we account for more estimation errors stemming from the sample eigenvalues. Specifically, in linear shrinkage, we account for estimation errors in μ_λ 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 a finite-sample setting in which N and T are fixed rather than a high-dimensional asymptotic setting.

Remark 2 To evaluate the required expectations in a finite-sample setting with N and T fixed, most of the literature on shrinkage mean-variance portfolio estimators employs the multivariate normal distributional assumption (Kan and Zhou, 2007; Tu and Zhou, 2011; Kan et al., 2021; Lassance et al., 2024a). We consider a more general setting in this paper by assuming that the sample of returns is drawn from a multivariate elliptical distribution; see Equation (6). To achieve this, we leverage some of the results in Kan and Lassance (2025).

3 Linear shrinkage

In this section, we consider the linear shrinkage of sample eigenvalues as pioneered by Ledoit and Wolf (2004). In that case, the shrinkage covariance matrix takes the form

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}' \quad \text{with} \quad \boldsymbol{D} = (1-\delta)\hat{\boldsymbol{\Lambda}} + \delta\hat{\mu}_\lambda \boldsymbol{I}_N,\ \hat{\mu}_\lambda = \frac{1}{N}\text{tr}(\hat{\boldsymbol{\Lambda}}). \tag{14}$$

Note that Ledoit and Wolf (2004) derive the MSE-optimal shrinkage intensity δ for the oracle μ_λ, because $\hat{\mu}_\lambda$ is a consistent estimator. However, given that we work in a finite-sample setting, we account for estimation errors in μ_λ and shrink the sample eigenvalues towards their mean $\hat{\mu}_\lambda$ in (14).

3.1 MSE-optimal linear shrinkage

In the next proposition, we derive the shrinkage intensity δ minimizing the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (14).

Proposition 2 *Assume the i.i.d. sample of returns* $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ *is drawn from the multivariate elliptical distribution, i.e.,* $\boldsymbol{r}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau)$. *Let* $3\kappa = \mathbb{V}\text{ar}[\tau^2]$ *be the excess kurtosis of any asset*

return, which we assume exists. Then, the shrinkage intensity $\delta \in \mathbb{R}$ *minimizing the MSE in* (12) *of the linear shrinkage estimator* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *in* (14) *is*

$$\delta^{\star}_{mse} = \frac{\frac{(N-2)s+N+2}{T-1} + \frac{\kappa[2(N-1)s+N+2]}{T}}{Ns + \frac{(N-2)s+N+2}{T-1} + \frac{\kappa[2(N-1)s+N+2]}{T}} \in [0,1], \tag{15}$$

where the lower bound of zero is achieved only as $T \to \infty$ *and*

$$s = \frac{N\frac{\mathrm{tr}(\boldsymbol{\Lambda}^2)}{\mathrm{tr}(\boldsymbol{\Lambda})^2} - 1}{N-1} \in [0,1) \tag{16}$$

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 2 shows that $\delta^{\star}_{mse}$ does not depend on $\boldsymbol{\mu}$, does depend on $\boldsymbol{\Sigma}$ but only via its eigenvalues through the sphericity parameter s, and depends on the elliptical fat tails only via the kurtosis parameter κ. Moreover, $\delta^{\star}_{mse} > 0$ as long as the sample size T is finite.

Remark 3 Proposition 2 extends results in Ollila and Raninen (2019), who derive the MSE-optimal δ under the multivariate elliptical distribution but considering the oracle μ_λ instead of $\hat{\mu}_\lambda$ in (14). It also extends results in Chen et al. (2010), who also derive the MSE-optimal δ considering $\hat{\mu}_\lambda$ in (14) but for $\kappa = 0$, i.e., under the multivariate normal distribution.

3.2 EU-optimal linear shrinkage

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 (14). As discussed in Remark 5 below, we derive the EU under an approximation concerning the distribution of the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (14).

Proposition 3 *Assume that* $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$*, where* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *is given by* (14)*, are obtained from an i.i.d. sample* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ *drawn from the multivariate elliptical distribution, i.e.,* $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$ *with* $\boldsymbol{\Sigma}_{\boldsymbol{D}} = \boldsymbol{V}\big[(1-\delta)\boldsymbol{\Lambda} + \delta\mu_\lambda \boldsymbol{I}_N\big]\boldsymbol{V}'$ *the population counterpart of* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$*. Then, provided* $T \geq N+5$*, the EU in* (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*

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-1)\boldsymbol{f}(\delta)'\boldsymbol{\theta}^2_{PC}}{\gamma(T-N-2)} - \frac{(T-1)^2\boldsymbol{f}(\delta)'\boldsymbol{A}\boldsymbol{f}(\delta)}{2\gamma(T-N-2)(T-N-4)}, \tag{17}$$

where $\boldsymbol{\theta}^2_{PC}$ *is the vector of squared Sharpe ratios of principal components (PCs),* $(\boldsymbol{\theta}^2_{PC})_i = \theta^2_{PC_i} = (\boldsymbol{v}_i'\boldsymbol{\mu})^2/\lambda_i$*, the entries of the vector* $\boldsymbol{f}(\delta)$ *are* $f_i(\delta) = \lambda_i((1-\delta)\lambda_i + \delta\mu_\lambda)^{-1}$*, the entries of the*

matrix $\boldsymbol{A}$ *are* $A_{ij} = (k_2\theta^2_{PC_i} + k_3/T)(\mathbb{1}_{i=j} + (T-N-1)^{-1}\mathbb{1}_{i\neq j})$*, and the parameters* (k_1, k_2, k_3) *are given by* (8)–(10)*, which we assume exist.*

We then define the optimal shrinkage intensity $\delta^\star_{eu}$ as

$$\delta^\star_{eu} = \underset{\delta\in\mathbb{R}}{\operatorname{argmax}}\, \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) \quad \text{with} \quad \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = (17). \tag{18}$$

There is no closed-form expression for $\delta^\star_{eu}$, but it can easily be found numerically.[10]

Unlike the MSE-optimal shrinkage intensity, $\delta^\star_{mse}$ in (15), which does not depend on $\boldsymbol{\mu}$ and only depends on $\boldsymbol{\Sigma}$ via the sphericity parameter s, the EU-optimal $\delta^\star_{eu}$ in (18) depends on both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ via the squared Sharpe ratios of PCs, $\theta^2_{PC_i}$, as well as the eigenvalues λ_i. In particular, the shrinkage intensity δ determines the shrinkage factor $f_i(\delta)$, which down-weights $\theta^2_{PC_i}$ if λ_i is below the average μ_λ and over-weights the remaining ones. Also, $\delta^\star_{eu}$ depends on the elliptical distribution via (k_1, k_2, k_3), versus only the kurtosis parameter κ for $\delta^\star_{mse}$.

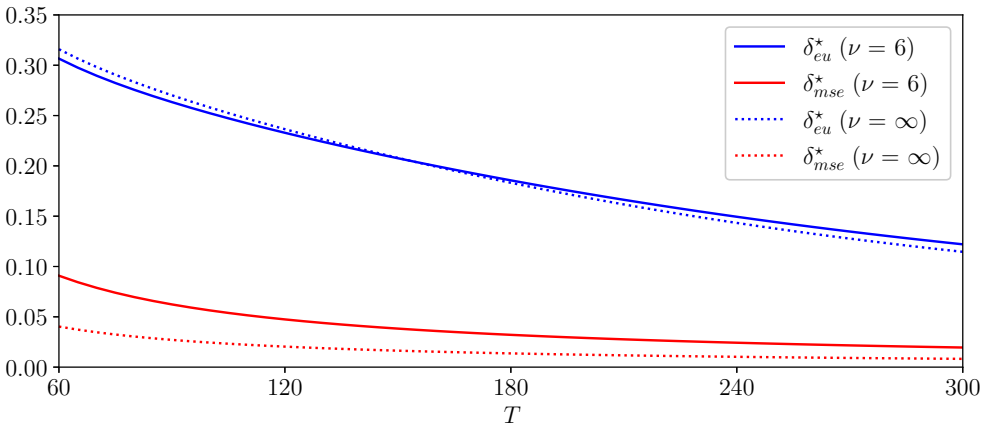
We expect the EU-optimal shrinkage intensity, $\delta^\star_{eu}$, to be larger than the MSE-optimal one, $\delta^\star_{mse}$. This is because $\delta^\star_{eu}$ accounts for the substantial impact of estimation errors stemming from the sample mean $\hat{\boldsymbol{\mu}}$ on the EU. Testament to this, we show in Section C of the supplementary material that, even when $\boldsymbol{\Sigma}$ is known, it is optimal under the EU to shrink it intensively to alleviate the impact of estimation errors in $\hat{\boldsymbol{\mu}}$, whereas the MSE would set an intensity $\delta = 0$ in that case.

Remark 4 It seems difficult to show that some non-zero shrinkage is always optimal, i.e., $\delta^\star_{eu} \neq 0$. However, in Section C.1 of the supplementary material, we consider a simpler setting where $\boldsymbol{\Sigma}$ is known and estimation errors only stem from $\hat{\boldsymbol{\mu}}$. In that case, we prove that $\frac{\partial \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})}{\partial \delta}\big|_{\delta=0} > 0$ provided T is finite and the λ_i's are not all equal, implying $\delta^\star_{eu} \neq 0$.

Remark 5 By assuming a similar distribution for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = (1-\delta)\hat{\boldsymbol{\Sigma}} + \delta\hat{\mu}_\lambda \boldsymbol{I}_N$ to that of $\hat{\boldsymbol{\Sigma}}$ in Proposition 3, we overstate sampling variability because $\hat{\mu}_\lambda \boldsymbol{I}_N$ has a lower variance than $\hat{\boldsymbol{\Sigma}}$. Specifically, we prove in Section A of the supplementary material that $\mathrm{tr}(\mathbb{V}[\mathrm{vec}(\hat{\boldsymbol{\Sigma}})]) > \mathrm{tr}(\mathbb{V}[\mathrm{vec}(\hat{\mu}_\lambda \boldsymbol{I}_N)])$. Therefore, our approach to determining $\delta^\star_{eu}$ is conservative, which is expected to compensate for the estimation errors stemming from the bona fide version of $\delta^\star_{eu}$. In Section B of the supplementary material, we show via simulations that $\delta^\star_{eu}$ is larger than, but close to, the exact optimal δ.

[10] $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ is proportional to $1/\gamma$, and thus, $\delta^\star_{eu}$ does not depend on γ. This is also true for the nonlinear shrinkage intensities we consider in Section 4.2.

Figure 1: Comparison of MSE and EU-optimal linear shrinkage intensities



Notes. This figure depicts the linear shrinkage intensities $\delta^\star_{mse}$ and $\delta^\star_{eu}$ in (15) and (18), which minimize the MSE of the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (14) and maximize the EU in (17) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, respectively. 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 $\gamma = 3$.

3.3 Comparison of MSE and EU-optimal linear shrinkage intensities

We now compare $\delta^\star_{mse}$ and $\delta^\star_{eu}$ in (15) and (18), respectively. $\delta^\star_{mse}$ 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 spanning July 1926 to July 2024. We consider a multivariate t-distribution with degrees of freedom $\nu \in \{6, \infty\}$, which gives $\kappa = 2/(\nu - 4)$. If $\nu = \infty$, then $k_1 = k_2 = k_3 = 1$ and $\kappa = 0$. Figure 1 depicts the intensities $\delta^\star_{eu}$ and $\delta^\star_{mse}$ as a function of the sample size $T \in [60, 300]$. We set a risk-aversion coefficient $\gamma = 3$ without loss of generality.

The main result in Figure 1 is that $\delta^\star_{eu}$ is substantially larger than $\delta^\star_{mse}$ because it accounts for the impact of estimation errors in both the sample mean and covariance matrix. For example, when $T = 120$ and $\nu = 6$, $\delta^\star_{eu} = 0.24$ is almost five times larger than $\delta^\star_{mse} = 0.05$. In Section 5.2, we use simulations to evaluate the EU delivered by bona fide estimators of $\delta^\star_{eu}$ and $\delta^\star_{mse}$.

4 Nonlinear shrinkage

In this section, we consider the nonlinear shrinkage of sample eigenvalues as developed by Ledoit and Wolf (2012, 2015, 2017, 2020, 2022b). 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. For ease of interpretation of our results in this section, we formulate our nonlinear shrinkage covariance matrix estimator so that we shrink the inverse sample eigenvalues toward zero:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}' \quad \text{with} \quad \boldsymbol{D} = (\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\hat{\boldsymbol{\Lambda}}, \tag{19}$$

where $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta}) = \text{diag}(\delta_1, \ldots, \delta_N)$ is the diagonal matrix of shrinkage intensities. Equation (19) implies that δ_i is the shrinkage intensity of the inverse sample eigenvalue $1/\hat{\lambda}_i$ since $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1} = \hat{\boldsymbol{V}}(\boldsymbol{I}_N - \boldsymbol{\Delta})\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{V}}'$. In particular, if δ_i is close to one, $1/\hat{\lambda}_i$ is heavily shrunk toward zero.

Remark 6 The standard nonlinear shrinkage approach, as in Ledoit and Wolf (2012, 2015, 2017, 2020, 2022b), sets $\boldsymbol{D}$ to be a constant diagonal matrix to calibrate. In this setting, the matrix $\boldsymbol{D} = \text{diag}(\hat{\boldsymbol{v}}_1'\boldsymbol{\Sigma}\hat{\boldsymbol{v}}_1, \ldots, \hat{\boldsymbol{v}}_N'\boldsymbol{\Sigma}\hat{\boldsymbol{v}}_N)$ minimizes the Frobenius norm $||\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} - \boldsymbol{\Sigma}||_F$, 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 (19), we account for estimation errors in sample eigenvalues for determining their optimal shrinkage.

4.1 MSE-optimal nonlinear shrinkage

Finding the optimal vector of shrinkage intensities $\boldsymbol{\delta}$ minimizing the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (19) in a finite-sample setting is difficult because it involves the first two moments of the sample eigenvalues $\hat{\lambda}_i$. In the next proposition, we derive the MSE-optimal $\boldsymbol{\delta}$ assuming, as in Proposition 3, that $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ is obtained from an i.i.d. multivariate elliptical sample with population covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$. As discussed in Remark 7 below, we follow Ledoit and Wolf (2017, Proposition 4.1) and constrain the shrinkage intensities such 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 in (14).

Proposition 4 *Assume that the nonlinear shrinkage estimator $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (19) is obtained from an i.i.d. sample $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ drawn from the multivariate elliptical distribution, i.e., $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$ with $\boldsymbol{\Sigma}_{\boldsymbol{D}} = \boldsymbol{V}(\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}\boldsymbol{V}'$ the population counterpart of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$. Then, the i-th entry of the vector of shrinkage intensities $\boldsymbol{\delta}^{\star}_{mse} \in \mathbb{R}^N$ minimizing the MSE in (12) of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ subject to the constraint $\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}})$ is*

$$\delta^{\star}_{mse,i} = \frac{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\mu_\lambda - \lambda_i)}{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)\mu_\lambda + \lambda_i} < 1. \tag{20}$$

It satisfies $\delta^{\star}_{mse,i} \to 1$ as $\lambda_i \to 0$, $\delta^{\star}_{mse,i} \geq 0$ if $\lambda_i \leq \mu_\lambda$, and $\delta^{\star}_{mse,i} \leq 0$ otherwise.

To the best of our knowledge, Proposition 4 is the first non-asymptotic (approximate) MSE-optimal solution to nonlinear shrinkage. Hediger et al. (2023) also consider nonlinear shrinkage in finite samples for elliptical distributions but in a different manner, i.e., they use nonlinear shrinkage to correct Tyler (1987)'s robust estimator.

Proposition 4 shows that the inverses of below-average sample eigenvalues are shrunk toward zero (i.e., $\delta_{mse,i} \geq 0$ if $\lambda_i \leq \mu_\lambda$), and the remaining inverse eigenvalues are enlarged (i.e., $\delta_{mse,i} \leq 0$ if $\lambda_i \geq \mu_\lambda$). This is consistent with linear shrinkage, except that each eigenvalue is assigned a different shrinkage intensity. Moreover, $\delta^\star_{mse,i} \to 1$ as $\lambda_i \to 0$, i.e., there is little contribution from low-variance PCs. Finally, all shrunk sample eigenvalues remain strictly positive because $\delta^\star_{mse,i} < 1$ for all i.

Remark 7 We can derive the unconstrained MSE-optimal nonlinear shrinkage intensities, i.e., without constraining the trace of $\hat{\boldsymbol{\Sigma}}$ to be preserved. We achieve this in Section D of the supplementary material, and show that the constrained and unconstrained intensities agree as $\lambda_i \to 0$ and ∞. However, we also show that the unconstrained solution is not well-behaved, as the optimal δ_i diverges to $\pm\infty$ as λ_i approaches a specific value below μ_λ. Moreover, we show empirically that, due to this divergence, the unconstrained MSE-optimal shrinkage estimator delivers a poor out-of-sample performance. Therefore, by imposing the trace constraint, we render the MSE nonlinear shrinkage estimator well-behaved and a tougher benchmark for our EU-optimal estimator.

4.2 EU-optimal nonlinear shrinkage

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 (19), the optimal vector of shrinkage intensities $\boldsymbol{\delta}$, and the EU it delivers. We use the conservative approximation discussed in Remark 5. We do not impose a constraint that the trace of the sample covariance matrix is preserved, unlike Proposition 4, because the unconstrained EU-optimal solution is well-behaved. However, such a constraint can easily be enforced, and we show in Section D of the supplementary material that EU-optimal nonlinear shrinkage delivers a strong portfolio performance empirically both with and without the constraint.

Proposition 5 *Assume that* $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$*, where* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *is given by* (19)*, are obtained from an i.i.d. sample* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ *drawn from the multivariate elliptical distribution, i.e.,* $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$ *with*

$\boldsymbol{\Sigma_D} = \boldsymbol{V}(\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}\boldsymbol{V}'$ *the population counterpart of* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$. *Define*

$$R_i = \frac{k_1\theta_{PC_i}^2}{k_2\theta_{PC_i}^2 + k_3/T} \in [0,1] \tag{21}$$

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* (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*

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-1)(\mathbf{1}_N - \boldsymbol{\delta})'}{\gamma(T-N-2)}\boldsymbol{\theta}_{PC}^2 - \frac{(T-1)^2(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{A}(\mathbf{1}_N - \boldsymbol{\delta})}{2\gamma(T-N-2)(T-N-4)}, \tag{22}$$

where $\mathbf{1}_N$ *is an* N*-dimensional vector of ones, the vector* $\boldsymbol{\theta}_{PC}^2$ *and the matrix* $\boldsymbol{A}$ *are defined in Proposition 3, and the parameters* (k_1, k_2, k_3) *are given by* (8)–(10), *which we assume exist.*

2. *The* i*-th entry of the vector of shrinkage intensities* $\boldsymbol{\delta}_{eu}^{\star} \in \mathbb{R}^N$ *maximizing* $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *in* (22) *is*

$$\delta_{eu,i}^{\star} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i - \frac{N}{T-2}\bar{R}\right). \tag{23}$$

It satisfies

$$\delta_{eu,i}^{\star} \in \left[0\ , 1 + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\bar{R}\right], \tag{24}$$

where the lower bound is achieved as $T \to \infty$ *and the upper bound when* $\theta_{PC_i}^2 = 0$.

3. *Define* $\boldsymbol{\Delta}_{eu}^{\star} = \mathrm{diag}(\boldsymbol{\delta}_{eu}^{\star})$ *and* $\boldsymbol{D}_{eu}^{\star} = (\boldsymbol{I}_N - \boldsymbol{\Delta}_{eu}^{\star})^{-1}\hat{\boldsymbol{\Lambda}}$. *The EU in* (22) *delivered by* $\boldsymbol{\delta}_{eu}^{\star}$ *is*

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}_{eu}^{\star}}) = \frac{k_1(T-1)\sum_{i=1}^{N}\left[(1-\delta_{eu,i}^{\star})\theta_{PC_i}^2\right]}{2\gamma(T-N-2)} \geq \frac{k_1(T-1)\sum_{i=1}^{N}\left[(1-\delta_{eu,cst}^{\star})\theta_{PC_i}^2\right]}{2\gamma(T-N-2)} \geq 0. \tag{25}$$

The first lower bound is obtained with $\delta_i = \delta_{eu,cst}^{\star}$ *for all* i, *where* $\delta_{eu,cst}^{\star}$ *maximizes* $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *in* (22) *subject to the constraint* $\boldsymbol{\delta} = \delta\mathbf{1}_N$:

$$\delta_{eu,cst}^{\star} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)}\left(\frac{k_1\theta^2}{k_2\theta^2 + k_3 N/T}\right) \in [0,1]. \tag{26}$$

4. *Assume* $k_1/k_2 > \frac{N}{T-2}\bar{R}$. *Then,* $\delta_{eu,i}^{\star} < 1$ *if* $\theta_{PC_i}^2 > \bar{\theta}_{PC}^2$, *and* $\delta_{eu,i}^{\star} \geq 1$ *otherwise, where*

$$\bar{\theta}_{PC}^2 = \frac{k_3\frac{N}{T}\bar{R}}{k_1(T-2) - k_2 N\bar{R}}. \tag{27}$$

Proposition 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, $\theta_{PC_i}^2$, and they also depend on

the elliptical distribution via (k_1, k_2, k_3). Unlike in linear shrinkage, the optimal vector of nonlinear shrinkage intensities $\boldsymbol{\delta}^\star_{eu}$ can be expressed in closed form via (23).

As in linear shrinkage, we typically expect the EU-optimal shrinkage intensities, $\boldsymbol{\delta}^\star_{eu}$, to be larger than the MSE-optimal ones, $\boldsymbol{\delta}^\star_{mse}$, because they account for estimation errors in $\hat{\boldsymbol{\mu}}$. Indeed, in Section C of the supplementary material, we show that $\delta^\star_{mse,i} = 0$ but $\delta^\star_{eu,i} = (1/T)/(\theta^2_{PC_i} + 1/T)$ when $\boldsymbol{\Sigma}$ is known. This is also suggested from their respective bounds: $\delta^\star_{mse,i}$ is always below one and negative for some i, whereas $\delta^\star_{eu,i}$ is always above zero and also above one for some i.

Part 3 of Proposition 5 shows that $\boldsymbol{\delta}^\star_{eu}$ always delivers a positive EU theoretically, which holds remarkably well in the simulations of Section 5.2 and the empirical analysis of Section 6. Such a result does not hold with $\delta^\star_{eu}$ under linear shrinkage, as the simulations of Section 5.2 clearly confirm. The reason for this is not because we use N different shrinkage intensities instead of one. Indeed, Part 3 of Proposition 5 shows that even if we constrain all shrinkage intensities to be equal, which amounts to shrinking linearly toward a zero matrix, the optimal $\delta^\star_{eu,cst}$ in (26) delivers a positive EU. Rather, linear shrinkage can deliver a negative EU because the Ledoit and Wolf (2004) target covariance matrix $\hat{\mu}_\lambda \boldsymbol{I}_N$ increases the first few inverse sample eigenvalues and decreases the other ones, whereas Proposition 5 shows that it is optimal to decrease all of them since $\delta^\star_{eu,i} \geq 0$.

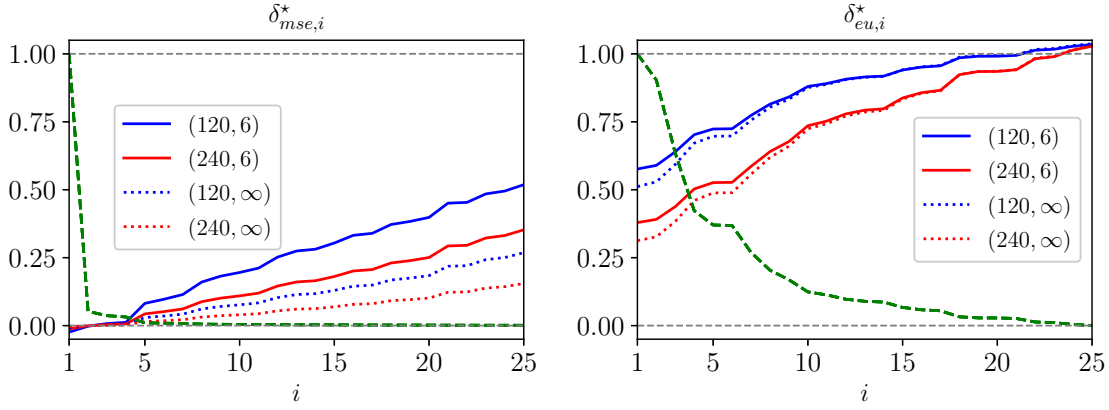
Finally, another difference between the MSE and EU-optimal nonlinear shrinkage intensities resides in the bounds they satisfy. We have $\delta^\star_{eu,i} \geq 0$ whereas $\delta^\star_{mse,i}$ can turn negative, which means that under the EU it is always optimal to decrease the inverse sample eigenvalues. It also holds that $\delta^\star_{mse,i} < 1$ whereas $\delta^\star_{eu,i}$ can surpass one for low-Sharpe-ratio PCs, i.e., $\theta^2_{PC_i} < \bar{\theta}^2_{PC}$ in (27). In that case, the shrunk eigenvalues can turn negative because $\hat{\lambda}_i/(1 - \delta^\star_{eu,i}) < 0$ if $\delta^\star_{eu,i} > 1$.

Remark 8 One may be reluctant to using negative shrunk eigenvalues, even if it is EU-optimal. Therefore, in Section E of the supplementary material, we study two ways of keeping all shrinkage intensities below one. First, we consider the shrinkage intensities $\boldsymbol{\delta}$ that maximize $\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (22) under the constraints $\delta_i \leq 1 \ \forall i$. We find that this solution is typically close to the unconstrained one in (23) because even the $\delta^\star_{eu,i}$ above one are typically close to one. Second, observing that

$$\sum_{i=1}^{N} \delta^\star_{eu,i} = N - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)} \sum_{i=1}^{N} R_i, \tag{28}$$

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 $\sum_{i=1}^{N} \delta_i = \sum_{i=1}^{N} \delta^\star_{eu,i}$.

Figure 2: Comparison of MSE and EU-optimal nonlinear shrinkage intensities



Notes. This figure depicts the nonlinear shrinkage intensities $\delta^\star_{mse,i}$ (left panel) and $\delta^\star_{eu,i}$ (right panel) in Equations (20) and (23) for $i = 1, \ldots, N = 25$, which minimize the MSE of the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (19) and maximize the EU in (22) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, 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 left panel and $\theta^2_{PC_i}/\theta^2_{PC_1}$ in the right panel. The horizontal gray dashed lines depict a shrinkage intensity δ_i equal to zero and one. The asset returns follow a multivariate t-distribution with degrees of freedom $\nu = \{6, \infty\}$, and each line in both panels represents a different choice of (T, ν) with $T = \{120, 240\}$. 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. We set a risk-aversion coefficient $\gamma = 3$.

We find that this solution is a dampened version of $\boldsymbol{\delta}^\star_{eu}$, 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 $\boldsymbol{\delta}^\star_{eu}$. Thus, our method can also output a nonlinear shrinkage covariance matrix with non-negative eigenvalues without sacrificing performance.

4.3 Comparison of MSE and EU-optimal nonlinear shrinkage intensities

We now compare the oracle nonlinear shrinkage intensities $\boldsymbol{\delta}^\star_{mse}$ and $\boldsymbol{\delta}^\star_{eu}$ given by (20) and (23), respectively. $\boldsymbol{\delta}^\star_{mse}$ depends on T, $\boldsymbol{\Lambda}$, and κ, whereas $\boldsymbol{\delta}^\star_{eu}$ depends on N, T, $\boldsymbol{\theta}^2_{PC}$, and (k_1, k_2, k_3). We calibrate these parameters as in Figure 1. The two panels of Figure 2 depict $\delta^\star_{mse,i}$ and $\delta^\star_{eu,i}$ for $i = 1, \ldots, 25$ using $T = 120$ and 240. The $\delta^\star_{mse,i}$ are in increasing order because the eigenvalues λ_i are in decreasing order. For visibility, we order the $\theta^2_{PC_i}$ in decreasing order, which is not the case in general, so that the $\delta^\star_{eu,i}$ are in increasing order like the $\delta^\star_{mse,i}$.

We make two main observations from Figure 2. First, both panels are in line with Propositions 4 and 5. Under the MSE, lower-variance PCs are shrunk more intensively, and $\delta^\star_{mse,i} < 0$ for the very first few eigenvalues that satisfy $\lambda_i > \mu_\lambda$. Under the EU, lower-Sharpe-ratio PCs are shrunk more

intensively, and $\delta^\star_{eu,i} > 1$ for the PCs that satisfy $\theta^2_{PC,i} < \bar{\theta}^2_{PC}$. Second, both panels show that $\boldsymbol{\delta}^\star_{eu}$ shrinks all inverse eigenvalues much more intensively toward zero than $\boldsymbol{\delta}^\star_{mse}$, which is because the EU accounts for estimation errors in both the sample mean and covariance matrix. For example, when $T = 120$ and $\nu = 6$, $\delta^\star_{mse,i} < 0.52$ whereas $\delta^\star_{eu,i} > 0.59$ for all i. In Section 5.2, we use simulations to evaluate the EU delivered by bona fide estimators of $\boldsymbol{\delta}^\star_{eu}$ and $\boldsymbol{\delta}^\star_{mse}$.

5 Bona fide shrinkage intensities and simulations

In Section 5.1, we introduce bona fide estimators of the shrinkage intensities. In Section 5.2, we compare the EU obtained under the bona fide versus oracle, and EU versus MSE-optimal, intensities.

5.1 Bona fide shrinkage intensities

We now estimate the MSE-optimal shrinkage intensities, $\delta^\star_{mse}$ in (15) and $\boldsymbol{\delta}^\star_{mse}$ in (20) for linear and nonlinear shrinkage, respectively, and the EU-optimal shrinkage intensities, $\delta^\star_{eu}$ in (18) and $\boldsymbol{\delta}^\star_{eu}$ in (23), respectively, which are oracles because they depend on the true $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and distribution of τ.

Starting with $\delta^\star_{mse}$ and $\boldsymbol{\delta}^\star_{mse}$, 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.,

$$\hat{\kappa} = \frac{1}{3N}\sum_{i=1}^{N} \hat{K}_i, \quad \hat{K}_i = \frac{(T-1)[(T+1)\hat{k}_i + 6]}{(T-2)(T-3)}, \tag{29}$$

where $\hat{k}_i = m_4(\boldsymbol{r}_i)/m_2(\boldsymbol{r}_i)^2 - 3$ and $m_k(\boldsymbol{r}_i) = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i)^k$. We estimate the sphericity parameter s using the asymptotically unbiased estimator in Ollila and Raninen (2019, Equation (27)), i.e.,

$$\hat{s} = \frac{1}{N-1}\left[\frac{(\hat{\kappa}+T)(T-1)^2}{(T-2)(3\hat{\kappa}(T-1)+T(T+1))}\left(\frac{N\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^2)}{\mathrm{tr}(\hat{\boldsymbol{\Lambda}})^2} - \frac{N(\hat{\kappa}+\frac{T}{T-1})}{\hat{\kappa}+T}\right) - 1\right], \tag{30}$$

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.,

$$\hat{\lambda}_{i,mse} = (1 - \hat{\delta}^\star_{mse})\hat{\lambda}_i + \hat{\delta}^\star_{mse}\hat{\mu}_\lambda, \tag{31}$$

where $\hat{\delta}^\star_{mse}$ is the bona fide estimator of the oracle $\delta^\star_{mse}$ obtained from $\hat{\kappa}$ and $\hat{s}$ above.

Turning to $\delta^\star_{eu}$ and $\boldsymbol{\delta}^\star_{eu}$, we need to estimate, in addition to the eigenvalues λ_i, the squared Sharpe

ratios of PCs $\theta^2_{PC_i}$ and the fat-tail parameters (k_1, k_2, k_3). We estimate the $\theta^2_{PC_i}$ in a structured way to alleviate the sensitivity to estimation errors in $\hat{\boldsymbol{\mu}}$.[11] Specifically, following Kozak et al. (2020) who argue that low-variance PCs with high Sharpe ratios are unlikely to persist out of sample,[12] we model $\theta^2_{PC_i}$ as decreasing with i. Specifically, we model the $\theta^2_{PC_i}$ with a geometric sequence, 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 (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

$$\theta^2_{PC_1} \frac{1 - g^N}{1 - g} = \theta^2. \tag{32}$$

This means that, under this model, estimating all $\theta^2_{PC_i}$ 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.,

$$\hat{\theta}^2_a = \frac{(T - N - 2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{N/2}(1 + \hat{\theta}^2)^{-(T-2)/2}}{T \times B_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2, (T - N)/2)}, \tag{33}$$

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

$$\hat{\theta}^2_{a,PC_1} = \frac{(T - 3)\hat{\theta}^2_{PC_1} - 1}{T} + \frac{2(\hat{\theta}^2_{PC_1})^{1/2}(1 + \hat{\theta}^2_{PC_1})^{-(T-2)/2}}{T \times B_{\hat{\theta}^2_{PC_1}/(1+\hat{\theta}^2_{PC_1})}(1/2, (T - 1)/2)}, \tag{34}$$

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 simulations to evaluate the expectations in (8)–(10), by estimating 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 $N/T \to c \in (0, 1)$, we have the asymptotic limit $(k_1, k_2, k_3) \to (\eta, \varphi, \eta)$, where $\eta \geq 1$ is the unique positive solution to

$$\mathbb{E}\big[(1 - c + c\eta\tau)^{-1}\big] = 1, \tag{35}$$

where the random variable τ determines the elliptical distribution in (6), and

$$\varphi = (1 - c)\left(\eta^{-2} - \mathbb{E}\left[\frac{c\tau^2}{(1 - c + c\eta\tau)^2}\right]\right)^{-1} \geq \eta^2. \tag{36}$$

[11] In unreported results, we find that using the plug-in estimators of the $\theta^2_{PC_i}$ instead of the structured approach we propose results in a small performance loss empirically, which does not affect the overall conclusions.

[12] Kozak et al. (2020, p.277) write: "[...] Sharpe ratios of low-eigenvalue PCs should be smaller than those of the high-eigenvalue PCs that are the major source of risk premia."

We estimate η and φ via the sample estimator of the distribution of τ in El Karoui (2010, 2013):

$$\hat{\tau}_t = \frac{(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})'(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}})}{\frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}})'(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}})}, \quad t = 1, \ldots, T, \tag{37}$$

which is consistent as $N \to \infty$. The sample estimate of $\hat{\eta}$ is the unique positive solution to

$$\frac{1}{T}\sum_{t=1}^{T}\left(1 - \frac{N}{T} + \frac{N}{T}\hat{\eta}\hat{\tau}_t\right)^{-1} = 1, \tag{38}$$

and the sample estimate of φ is

$$\hat{\varphi} = \left(1 - \frac{N}{T}\right)\left(\hat{\eta}^{-2} - \frac{1}{T}\sum_{t=1}^{T}\frac{\frac{N}{T}\hat{\tau}_t^2}{\left(1 - \frac{N}{T} + \frac{N}{T}\hat{\eta}\hat{\tau}_t\right)^2}\right)^{-1}. \tag{39}$$

5.2 Comparison of EU via simulations

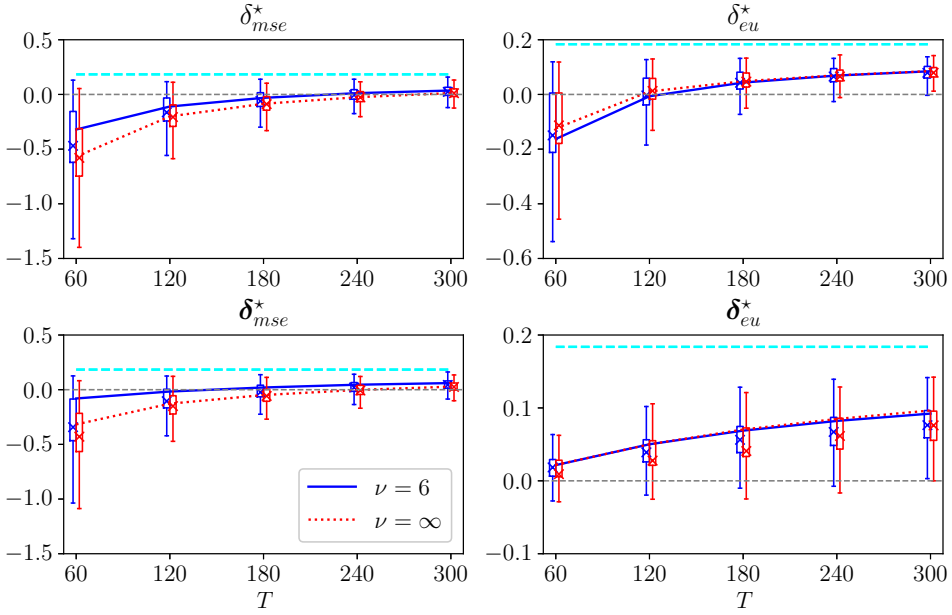
In this section, we use simulations to evaluate the EU under the bona fide and oracle linear and nonlinear shrinkage intensities. We calibrate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024, as previously. Given $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we simulate $M = 10{,}000$ times a sample of size $T \in \{60, 120, 180, 240, 300\}$ of multivariate t-distributed returns, with $\nu \in \{6, \infty\}$ degrees of freedom, for the N assets, i.e., $(\boldsymbol{r}_{1,m}, \ldots, \boldsymbol{r}_{T,m})$ with $m = 1, \ldots, M$. In each simulation m, we compute the sample mean $\hat{\boldsymbol{\mu}}_m$, the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_m$, and the bona fide estimators of the EU and MSE-optimal linear and nonlinear shrinkage intensities introduced in Section 5.1. Given that we know the population parameters, we can also compute the oracle shrinkage intensities, which do not vary across simulations. Then, given a shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D},\boldsymbol{m}}$ in simulation m, we construct the shrinkage mean-variance portfolio as

$$\hat{\boldsymbol{w}}_{\boldsymbol{D},m} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D},\boldsymbol{m}}^{-1}\hat{\boldsymbol{\mu}}_m, \tag{40}$$

where we set $\gamma = 3$ without loss of generality. We consider eight different types of shrinkage mean-variance portfolios, corresponding to linear or nonlinear shrinkage, MSE or EU-optimal shrinkage, and bona fide or oracle intensities. Finally, for a given shrinkage method, we obtain the out-of-sample utility in simulation m as $\hat{\boldsymbol{w}}'_{\boldsymbol{D},m}\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}'_{\boldsymbol{D},m}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D},m}$, and the EU by averaging them:

$$\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\boldsymbol{w}}'_{\boldsymbol{D},m}\boldsymbol{\mu} - \frac{\gamma}{2}\hat{\boldsymbol{w}}'_{\boldsymbol{D},m}\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D},m}\right). \tag{41}$$

Figure 3: EU delivered by the bona fide versus oracle linear and nonlinear shrinkage intensities



Notes. This figure depicts the annualized EU of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ is the shrinkage covariance matrix obtained under the MSE-optimal linear shrinkage intensity $\delta^\star_{mse}$ (top-left panel), EU-optimal linear shrinkage intensity $\delta^\star_{eu}$ (top-right panel), MSE-optimal nonlinear shrinkage intensities $\boldsymbol{\delta}^\star_{mse}$ (bottom-left panel), and EU-optimal nonlinear shrinkage intensities $\boldsymbol{\delta}^\star_{eu}$ (bottom-right panel). We depict the EU as a function of the sample size $T \in \{60, 120, 180, 240, 300\}$. We obtain the EU by simulating asset returns from a multivariate t-distribution with degrees of freedom $\nu \in \{6, \infty\}$ whose mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are calibrated to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024; see Section 5.2 for details on the simulation. In each panel, we depict, for each choice of ν and T, 1) the annualized EU without parameter uncertainty, i.e., $12U(\boldsymbol{w}^\star)$ in (3) (horizontal dashed cyan line), 2) the annualized EU under the oracle shrinkage intensities (solid blue and dotted red lines), and 3) boxplots of out-of-sample utilities under the bona fide shrinkage intensities as well as the EU with an "$\times$" marker in each boxplot. We set a risk-aversion coefficient $\gamma = 3$.

In Figure 3, we report the simulation results in four panels, corresponding to MSE-optimal and EU-optimal linear and nonlinear shrinkage. In each panel, we depict, for $\nu \in \{6, \infty\}$ and $T \in \{60, 120, 180, 240, 300\}$, 1) the annualized EU without parameter uncertainty, i.e., $12U(\boldsymbol{w}^\star)$ in (3), 2) the annualized EU under the oracle shrinkage intensities, and 3) the annualized EU under the bona fide shrinkage intensities as well as boxplots of out-of-sample utilities around this EU.

We make a number of observations from Figure 3. First, the MSE-optimal shrinkage intensities $\delta^\star_{mse}$ and $\boldsymbol{\delta}^\star_{mse}$ need a rather large sample size to deliver a positive EU, even when using the oracle intensities. Second, EU-optimal shrinkage delivers a substantially larger EU than MSE-optimal shrinkage. For example, when $T = 120$, the annualized EU gain delivered by the bona fide EU-optimal versus MSE-optimal intensities is 0.16 and 0.22 in linear shrinkage, and 0.15 and 0.18 in nonlinear shrinkage, for $\nu = 6$ and ∞, respectively. Third, $\delta^\star_{eu}$ delivers a positive EU as of $T = 120$ under both the oracle and bona fide intensities, and $\boldsymbol{\delta}^\star_{eu}$ delivers a positive EU for all T considered

even under the bona fide intensities This last result is consistent with the theoretically positive EU delivered by EU-optimal nonlinear shrinkage; see part 3 of Proposition 5. Fourth, the boxplots of out-of-sample utilities under EU-optimal shrinkage are less wide than under MSE-optimal shrinkage. This is because we shrink the sample eigenvalues more intensively under the EU than the MSE, and thus, the EU-optimal shrinkage mean-variance portfolio is less exposed to estimation errors. Finally, the EU loss incurred by relying on bona fide instead of oracle shrinkage intensities appears limited, particularly as T increases, supporting the estimation methodology proposed in Section 5.1.[13]

Overall, these results highlight the relevance for a mean-variance investor of calibrating the shrinkage covariance matrix to a portfolio-motivated objective function, which is well corroborated by the extensive empirical analysis contained in Section 6.

6 Empirical analysis

We now evaluate the empirical portfolio performance obtained with our shrinkage covariance matrix estimators. In Section 6.1, we present the six datasets we use. In Section 6.2, we detail the shrinkage covariance matrix estimators we consider as well as the benchmark portfolio strategies. In Section 6.3, we explain the out-of-sample methodology. In Section 6.4, we discuss the results.

6.1 Datasets

Table 1 lists the six datasets of monthly excess returns we consider. The first three 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) and obtained from jkpfactors.com. The last two datasets are composed of the long and short legs obtained from the eight low-turnover characteristics (16NMV) and the 23 characteristics (46NMV) in Novy-Marx and Velikov (2015), available on Robert Novy-Marx's website.

[13] In the right panels of Figure 3, it can happen that the bona fide shrinkage intensities deliver a larger EU than the oracle ones. This can be explained to the use of a finite number of simulations and because the oracle intensities in Propositions 3 and 5 do not maximize the exact EU but an approximation.

Table 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	16NMV	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	46NMV	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 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 jkpfactors.com. The last two datasets are collected from Robert Novy-Marx's website.

6.2 Portfolio strategies

We consider 21 strategies in total, listed in Table 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 the sample mean $\hat{\boldsymbol{\mu}}$ and rotation-equivariant shrinkage covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}'$ for different choices of $\boldsymbol{D}$. In Section F of the supplementary material, we also implement these first nine mean-variance portfolios using the Jorion (1986) Bayes-Stein mean estimator. First, the EU-optimal linear shrinkage estimator in Proposition 3 (EUL). Second, the EU-optimal nonlinear shrinkage estimator in Proposition 5 (EUNL). Third, the MSE-optimal linear shrinkage estimator in Proposition 2 (MSEL). Fourth, the MSE-optimal nonlinear shrinkage estimator in Proposition 4 (MSENL). In Section 5.1, we introduce bona fide shrinkage intensities for these first four estimators.[14] Fifth, the sample covariance matrix in (5) (SAMPLE). Sixth, the asymptotic MSE-optimal linear shrinkage estimator of Ledoit and Wolf (2004) (LW2004). Seventh, the asymptotic MSE-optimal nonlinear shrinkage estimator of Ledoit and Wolf (2020) (LW2020).

The eighth covariance matrix estimator is the MSE-optimal linear shrinkage estimator of Bodnar et al. (2014) (BGP). Their shrinkage estimator considers a general target matrix $\boldsymbol{\Sigma}_0$, which when choosing $\boldsymbol{\Sigma}_0 = \hat{\mu}_\lambda \boldsymbol{I}_N$ and applying their quations (4.3)–(4.4) gives

$$\hat{\boldsymbol{\Sigma}}_{BGP} = (1 - \hat{\delta}_{BGP})\hat{\boldsymbol{\Sigma}} + \hat{\delta}_{BGP}\hat{\mu}_\lambda \boldsymbol{I}_N \quad \text{with} \quad \hat{\delta}_{BGP} = \frac{N/T}{(N-1)\hat{s}_p}, \tag{42}$$

[14] In Section G 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 distributional assumption. 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.

Table 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 (13) with linear shrinkage estimator maximizing the EU (Proposition 3)
2	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with nonlinear shrinkage estimator maximizing the EU (Proposition 5)
3	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with linear shrinkage estimator minimizing the MSE (Proposition 2)
4	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with nonlinear shrinkage estimator minimizing the MSE (Proposition 4)
5	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with sample estimator in (5)
6	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with linear shrinkage estimator of Ledoit and Wolf (2004)
7	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with nonlinear shrinkage estimator of Ledoit and Wolf (2020)
8	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with linear shrinkage estimator of Bodnar et al. (2014)
9	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) with PCA estimator of Chen and Yuan (2016)
10	GMV	SAMPLE	GMV portfolio with sample estimator in (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 (44) with sample estimator in (5)
14	GMVRF	LW2004	GMVRF portfolio in (44) with linear shrinkage estimator of Ledoit and Wolf (2004)
15	GMVRF	SSHH	GMVRF portfolio in (44) 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 (45)
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 6. The first two portfolios, EUL and EUNL, are the strategies introduced in this paper. The first nine portfolios are of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13) using different rotation-equivariant estimators of the covariance matrix of the form $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (11) for different choices of the diagonal matrix $\boldsymbol{D}$. More details about the 21 portfolio strategies are available in Section 6.2.

where $\hat{s}_p = \big(\frac{N\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^2)}{\mathrm{tr}(\hat{\boldsymbol{\Lambda}})^2} - 1\big)/(N-1)$ is the sample plug-in estimator of the sphericity parameter s.

The ninth shrinkage covariance matrix estimator uses PCA to keep 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 eigenvalues sorted in decreasing order. Then, we construct the PCA mean-variance portfolio as $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{V}}_{\hat{K}}\hat{\boldsymbol{\Lambda}}_{\hat{K}}^{-1}\hat{\boldsymbol{V}}_{\hat{K}}'\hat{\boldsymbol{\mu}}$, where the estimator $\hat{K}$ is that in Bai and Ng (2002), i.e.,

$$\hat{K} = \underset{K\in[1,N-1]}{\mathrm{argmin}} \ \ln\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}^2\right) + K\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \tag{43}$$

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 2 are the global-minimum-variance (GMV)

[15]The PCA estimator amounts to taking $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1} = \hat{\boldsymbol{V}}\boldsymbol{D}^{-1}\hat{\boldsymbol{V}}'$ with $\boldsymbol{D}^{-1} = \mathrm{diag}\big(1/\hat{\lambda}_1, \ldots, 1/\hat{\lambda}_{\hat{K}}, \boldsymbol{0}_{N-\hat{K}}\big)$.

portfolio, $\hat{\boldsymbol{w}}_g = (\mathbf{1}_N'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N)^{-1}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N$, where $\hat{\boldsymbol{\Sigma}}$ is either the sample estimator, the LW2004 estimator, or the cross-validated maximum-likelihood estimator of Shi et al. (2025) (SSHH). For the latter, we use the diagonal target matrix $\hat{\mu}_\lambda \boldsymbol{I}_N$ and 10 folds.[16] The following three portfolio strategies consist of the EU-optimal combination of the GMV portfolio with the risk-free asset (GMVRF) in Kan and Lassance (2025, Section VI.C.2.), i.e.,

$$\hat{\boldsymbol{w}}_{gmvrf} = \frac{\hat{\eta}(T-N-1)(T-N-4)}{\gamma\hat{\varphi}T(T-2)}\left(\frac{\mathbf{1}_N'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}_N'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}\right)\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N, \tag{44}$$

where we use the same three estimators $\hat{\boldsymbol{\Sigma}}$ as for GMV, and $\hat{\eta}$ and $\hat{\varphi}$ are defined in (38)–(39).[17]

Two other portfolio strategies we consider as benchmarks are the equally weighted portfolio (EW), $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$, and the EU-optimal combination of the EW portfolio with the risk-free asset (EWRF) considered in, e.g., Lassance et al. (2024b), i.e.,

$$\hat{\boldsymbol{w}}_{ewrf} = \frac{T-3}{\gamma T}\left(\frac{\boldsymbol{w}_{ew}'\hat{\boldsymbol{\mu}}}{\boldsymbol{w}_{ew}'\hat{\boldsymbol{\Sigma}}\boldsymbol{w}_{ew}}\right)\boldsymbol{w}_{ew}. \tag{45}$$

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). Second, the universal portfolio shrinkage estimator (UPSA) of Kelly et al. (2023), which shrinks the sample eigenvalues nonlinearly to maximize the leave-one-out EU.[18] 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. (2025, Section F), which optimally combines the mean-variance portfolio of the first K PCs with its orthogonal arbitrage portfolio estimated via the empirical Bayes method of Da et al. (2024), where the number of PCs K is estimated with the method of Bai and Ng (2002) in (43).

[16] R codes implementing the SSHH covariance matrix are available at 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.

[18] Python codes implementing the UPSA portfolio strategy are available at github.com/pourmohammadimohammad/Universal_Portfolio_Shrinkage/tree/main. We consider the version of UPSA that allows an investment in the risk-free asset, for consistency with our portfolio strategies.

6.3 Out-of-sample methodology

We measure portfolio performance with 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 a sample size $T = 120$ 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 observations. We then compute the net out-of-sample returns, $r_{net,k,t} = r_{gross,k,t}$ if $t = T+1$ and

$$r_{net,k,t} = (1 + r_{gross,k,t})(1 - c \times \text{turnover}_{k,t-1}) - 1 \quad \text{if} \quad t = T+2, \ldots, T_{tot}, \tag{46}$$

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

$$\text{turnover}_{k,t} = \sum_{i=1}^{N} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, T_{tot}, \tag{47}$$

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 $c = 10$ basis points in line with recent values of average bid-ask spreads reported by Engle et al. (2012) and Frazzini et al. (2018).[19] Finally, we compare the portfolio strategies in terms of annualized out-of-sample utility net of transaction costs,

$$U_k = 12 \times \Big(\hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2\Big), \tag{48}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the sample mean and variance of $r_{net,k,t}$. In Sections H and I of the supplementary material, we also report the average portfolio turnover and the net out-of-sample Sharpe ratio of the different portfolio strategies, respectively.

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.

[19] Transaction costs are larger in the older part of the sample, but our objective is to adjust returns to transaction costs of a magnitude that investors encounter nowadays.

6.4 Results

In Table 3, we report the annualized net out-of-sample utility, in percentage points, of the 21 portfolio strategies listed in Table 2 across the six datasets listed in Table 1, following the methodology detailed in Section 6.3. We consider $\gamma \in \{3, 10\}$ and $T = \{120, 240\}$. Figure 4 depicts the median gain in annualized net out-of-sample utility of EUL and EUNL versus all other benchmarks, where the median is taken over 12 observations (six datasets with two sample sizes), for $\gamma = 3$. These medians are all positive, indicating a consistent performance gain delivered by our estimators.

Our first objective is to assess the empirical performance gain of our proposed shrinkage portfolios, EUL and EUNL, relative to their MSE-optimal counterparts, MSEL and MSENL. 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 utility 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 MSENL. 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 5, EUNL systematically delivers a positive net utility, whereas MSENL delivers a negative net utility in 50% of the cases. Among the 21 competitors, only EUNL and GMVRF achieve this feat.

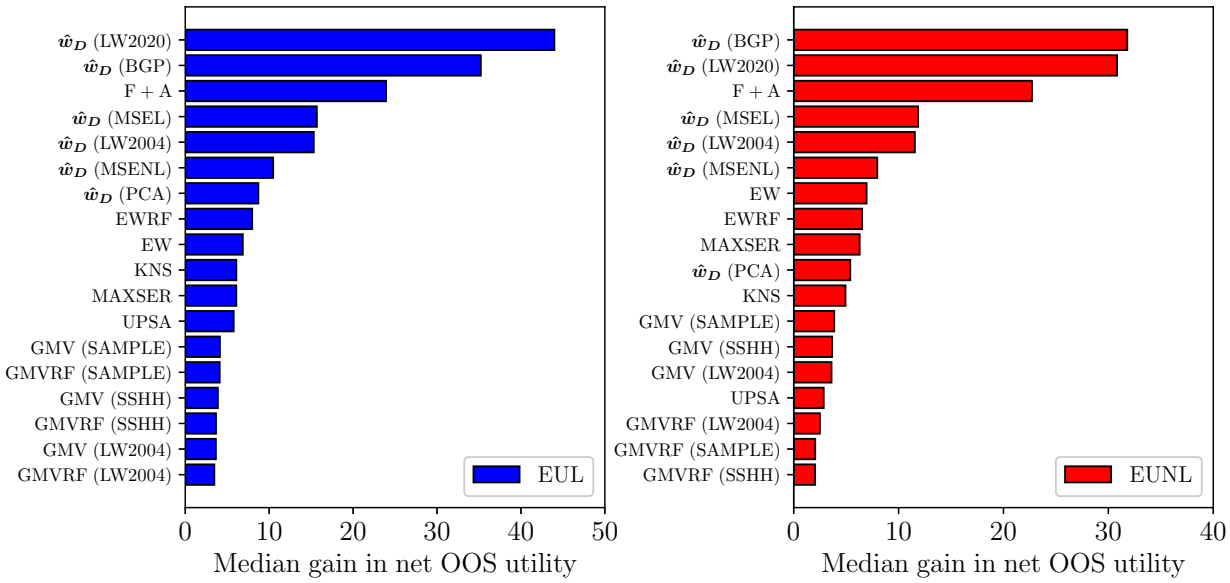
These results support the theoretical evidence in Sections 3 and 4, as well as the simulation evidence in Section 5.2, and indicate the performance benefits of calibrating shrinkage covariance matrices with a portfolio loss function instead of the MSE. In addition, because EUL and EUNL shrink the sample eigenvalues more intensively than MSEL and MSENL, they also achieve a lower portfolio turnover. Specifically, Table SM.5 of the supplementary material shows that, on average, MSEL has a portfolio turnover 198% and 188% larger than EUL, and MSENL 340% and 220% larger than EUNL, for $T = 120$ and 240, respectively.

Table 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 NMV	46 NMV	25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (13)	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●
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	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●
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	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 (13)	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●
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	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●
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	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 6.2 and listed in Table 2. The first nine of them are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ 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 EU-optimal linear and nonlinear shrinkage estimators introduced in this paper. We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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 net 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.

Figure 4: Median gain in annualized net out-of-sample utility in empirical data



Notes. This figure depicts the median difference between the annualized net out-of-sample utility, in percentage points, obtained with the proposed EUL (left panel) and EUNL (right panel) portfolios relative to all the benchmarks listed in Table 2, except the $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ SAMPLE benchmark for visibility. The medians are taken over 12 observations, corresponding to the six datasets in Table 1 times two sample sizes $T = 120$ and 240. We order the benchmarks from worst to best in terms of median utility gain. The figure is constructed following the methodology described in Section 6.3. We consider a risk-aversion coefficient $\gamma = 3$.

Comparing EUL and EUNL, there is 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 difference can be explained because nonlinear shrinkage involves N 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 SM.5 of the supplementary material shows that, on average, EUNL has a turnover 36.5% and 16.1% smaller than EUL for $T = 120$ and 240, respectively.

After comparing EUL, EUNL, MSEL, and MSENL, we now compare 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-optimal 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, which are systematically outperformed by LW2004. As a result, our proposed EU-optimal shrinkage estimators, EUL and

EUNL, almost systematically improve upon each of LW2004, LW2020, and BGP, which is consistent with the improvement relative to MSEL and MSENL discussed above.

The PCA-based estimator delivers superior and more stable results relative to the other benchmarks, with a positive net utility in most cases. This improved performance can be explained because PCA only keeps the first few eigenvalues, avoiding issues related to inverting small eigenvalues. Consequently, as documented by Chen and Yuan (2016), PCA delivers stable portfolio weights. Indeed, Table SM.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. Still, our EU-optimal estimators EUL and EUNL outperform PCA in all configurations except one, even though the utility 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."

In Section F of the supplementary material, we show that the superiority of EUL and EUNL versus MSEL, MSENL, SAMPLE, LW2004, LW2020, BGP, and PCA is robust to using the Bayes-Stein mean estimator of Jorion (1986) instead of the sample mean. Moreover, whereas the benchmarks systematically profit from using the Jorion mean estimator, it is no longer the case for EUL and EUNL, and particularly EUNL. Therefore, when the mean-variance portfolio is implemented with our EU-optimal covariance matrix estimators, using the sample mean is no longer a concern.

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, 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 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.

Finally, we compare our EUL and EUNL portfolios to the recently proposed MAXSER, UPSA, KNS, and F+A benchmarks. Overall, these benchmarks can offer substantial improvements relative to the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ implemented with the sample covariance matrix estimator and existing shrinkage estimators. However, unlike our EUL and EUNL portfolios, these benchmarks deliver out-of-sample net 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 SM.5 of the supplementary material, these benchmarks also face a substantially larger turnover than our EUL and EUNL portfolios.

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 paper, 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 typically 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 stemming from mean returns in the calibration of the shrinkage intensities. This proposal addresses the call for 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 small squared Sharpe ratios. 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-optimal shrinkage covariance matrices compare favorably out of sample to MSE-optimal covariance matrices (Ledoit and Wolf, 2004, 2020; Bodnar et al., 2014), 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 equally weighted portfolios. Moreover, mean-variance portfolios constructed with our EU-optimal 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 paper to shrink both the mean vector and the covariance matrix to maximize the EU, similar to Bodnar et al. (2024) who shrink both 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 instead of the EU. In addition, mean-variance efficient portfolio weights provide the stochastic discount factor (SDF) in asset pricing. MSE-optimal 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.

References

Ao, M., Li, Y., and Zheng, X. (2019), "Approaching mean-variance efficiency for large portfolios." *Review of Financial Studies*, 32(7):2890–2919.

Bai, J. and Ng, S. (2002), "Determining the number of factors in approximate factor models." *Econometrica*, 70(1):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):1222–1278.

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*, 132:215–228.

Bodnar, T., Gupta, A. K., and Parolya, N. (2016), "Direct shrinkage estimation of large dimensional precision matrix." *Journal of Multivariate Analysis*, 146:223–236.

Bodnar, T., Parolya, N., and Thorsen, E. (2024), "Two is better than one: Regularized shrinkage of large minimum variance portfolio." *Journal of Machine Learning Research*, 25:1–32.

Cai, Z., Cui, Z., Lassance, N., and Simaan, M. (2024), "The economic value of mean squared error: Evidence from portfolio selection." *SSRN working paper*.

Chen, J. and Yuan, M. (2016), "Efficient portfolio selection in a large market." *Journal of Financial Econometrics*, 14(3):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):5016–5029.

Chernov, M., Dahlquist, M., and Lochstoer, L. (2023), "Pricing currency risks." *Journal of Finance*, 78(2):693–730.

Da, R., Nagel, S., and Xiu, D. (2024), "The statistical limit of arbitrage." *National Bureau of Economic Research*, Technical report.

DeMiguel, V., Garlappi, L., Nogales, F., and Uppal, R. (2009), "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms." *Management Science*, 55(5):798–812.

El Karoui, N. (2010), "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation." *Annals of Statistics*, 38(6):3487–3566.

El Karoui, N. (2013), "On the realized risk of high-dimensional Markowitz portfolios." *SIAM Journal on Financial Mathematics*, 4:737–783.

Elmachtoub, A. and Grigas, P. (2022), "Smart "Predict, then Optimize"." *Management Science*, 68(1):9–26.

Engle, R., Ferstenberg, R., and Russell, J. (2012), "Measuring and modeling execution cost and risk." *Journal of Portfolio Management*, 38(2):14–28.

Frahm, G. and Memmel, C. (2010), "Dominating estimators for minimum-variance portfolios." *Journal of Econometrics*, 159(2):289–302.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2018), "Trading costs." *SSRN working paper*.

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*, 71:1657–1668.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023), "Is there a replication crisis in finance?" *Journal of Finance*, 78(5):246–2518.

Jorion, P. (1986), "Bayes-Stein estimation for portfolio analysis." *Journal of Financial and Quantitative Analysis*, 21(3):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., Wang, X., and Zheng, X. (2024), "In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models." *Journal of Financial Economics*, 155.

Kan, R., Wang, X., and Zhou, G. (2021), "Optimal portfolio choice with estimation risk: No risk-free asset case." *Management Science*, 68(3):2047–2068.

Kan, R. and Zhou, G. (2007), "Optimal portfolio choice with parameter uncertainty." *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Kazak, E., Li, Y., Nolte, I., and Nolte, S. (2025), "Realized weight regression for dynamic realized global minimum variance portfolio allocation." *SSRN working paper*.

Kelly, B., Malamud, S., Pourmohammadi, M., and Trojani, F. (2023), "Universal portfolio shrinkage." *SSRN working paper*.

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):2522–2531.

Kozak, S., Nagel, S., and Santosh, S. (2020), "Shrinking the cross-section." *Journal of Financial Economics*, 135(2):271–292.

Lassance, N., Martín-Utrera, A., and Simaan, M. (2024a), "The risk of expected utility under parameter uncertainty." *Management Science*, 70(11):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):875–889.

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

Ledoit, O. and Wolf, M. (2003a), "Honey, I shrunk the sample covariance matrix." *Journal of Portfolio Management*, 30(4):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):603–621.

Ledoit, O. and Wolf, M. (2004), "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2008), "Robust performance hypothesis testing with the Sharpe ratio." *Journal of Empirical Finance*, 15(5):850–859.

Ledoit, O. and Wolf, M. (2012), "Nonlinear shrinkage estimation of large-dimensional covariance matrices." *Annals of Statistics*, 40(2):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*, 139:360–384.

Ledoit, O. and Wolf, M. (2017), "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks." *Review of Financial Studies*, 30(12):4349–4388.

Ledoit, O. and Wolf, M. (2020), "Analytical nonlinear shrinkage of large-dimensional covariance matrices." *Annals of Statistics*, 48(5):3043–3065.

Ledoit, O. and Wolf, M. (2022a), "The power of (non-)linear shrinking: A review and guide to covariance matrix estimation." *Journal of Financial Econometrics*, 20(1):187–218.

Ledoit, O. and Wolf, M. (2022b), "Quadratic shrinkage for large covariance matrices." *Bernoulli*, 28(3):1519–1547.

Mattera, R. (2025), "Improved precision matrix estimation for mean-variance portfolio selection." *Statistics*, 59(30):661–681.

Novy-Marx, R. and Velikov, M. (2015), "A taxonomy of anomalies and their trading costs." *Review of Financial Studies*, 29(1):104–147.

Ollila, E. and Raninen, E. (2019), "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions." *IEEE Transactions on Signal Processing*, 67(10):2707–2719.

Politis, D. and Romano, J. (1994), "The stationary bootstrap." *Journal of the American Statistical Association*, 89(428):1303–1313.

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):2700–2731.

Tu, J. and Zhou, G. (2011), "Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies." *Journal of Financial Economics*, 99(1):204–215.

Tyler, D. E. (1987), "A distribution-free M-estimator of multivariate scatter." *Annals of Statistics*, 15(1):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*, 594:400–415.

Supplementary Material to

Shrink with Purpose: Optimal Covariance Matrix Estimation for Portfolio Selection

Nathan Lassance Rodolphe Vanderveken Frédéric Vrins

This supplementary material to the paper contains 10 sections. In Section A, we compare the sampling variability of the sample and target covariance matrices. In Section B, we test the accuracy of the proposed approximation to the optimal linear shrinkage intensity. In Section C, we study the EU-optimal linear and nonlinear shrinkage intensities when the covariance matrix is known. In Section D, we compare nonlinear shrinkage intensities with and without a trace-preservation constraint. In Section E, we study two ways of constraining the EU-optimal nonlinearly shrunk eigenvalues to be positive. In Section G, we compare the empirical performance under the multivariate normal versus elliptical assumption. In Section F, we analyze the empirical performance of shrinkage mean-variance portfolios implemented with a robust mean estimator. In Section H, we investigate the empirical stability of our shrinkage intensities and its impact on portfolio turnover. In Section I, we report the empirical out-of-sample Sharpe ratio of the considered portfolio strategies. Finally, in Section J, we provide the proofs of all theoretical results.

A Sampling variability of sample and target estimators

To compare the sampling variability of two matrix estimators, $\hat{\boldsymbol{M}}_1$ and $\hat{\boldsymbol{M}}_2$, we can compare $\operatorname{tr}(\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{M}}_1)])$ and $\operatorname{tr}(\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{M}}_2)])$, i.e., the sum of the variances of each entry in $\hat{\boldsymbol{M}}_1$ and $\hat{\boldsymbol{M}}_2$. In the next proposition, we compare $\hat{\boldsymbol{\Sigma}}$ and $\hat{\mu}_\lambda \boldsymbol{I}_N$ in this manner.

Proposition SM.1 *Assume the i.i.d. sample of returns $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$ is drawn from the multivariate elliptical distribution, i.e., $\boldsymbol{r}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau)$. Let $3\kappa = \mathbb{V}\mathrm{ar}[\tau^2]$ be the excess kurtosis of any asset return, which we assume exists. Then,*

$$\begin{aligned}
&\operatorname{tr}(\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\Sigma}})]) - \operatorname{tr}(\mathbb{V}[\operatorname{vec}(\hat{\mu}_\lambda \boldsymbol{I}_N)]) \\
&= \frac{1}{N}\left[\left(\frac{N}{T-1} + \frac{(N-1)\kappa}{T}\right)\operatorname{tr}(\boldsymbol{\Sigma})^2 + \left(\frac{N-2}{T-1} + \frac{2(N-1)\kappa}{T}\right)\operatorname{tr}(\boldsymbol{\Sigma}^2)\right] \geq 0, \qquad \text{(SM1)}
\end{aligned}$$

where the lower bound of zero is achieved if and only if $N = 1$ or $T \to \infty$.

Therefore, because we assume $N \geq 2$ in this paper, $\hat{\boldsymbol{\Sigma}}$ has a strictly larger sampling variability than $\hat{\mu}_\lambda \boldsymbol{I}_N$ as long as the sample size T is finite.

B Accuracy of approximation to linear shrinkage intensity

In this section, we evaluate the accuracy of the approximation used in Proposition 3 to derive the EU-optimal linear shrinkage intensity, $\delta_{eu}^{\star}$ in (18). Proposition 5 uses a similar approximation to derive the EU-optimal nonlinear shrinkage intensities.

We proceed as follows. As in Section 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. 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 (18). We can also find the EU-optimal shrinkage intensity without approximation, denoted $\delta_{eu}^{\star\star}$, via Monte Carlo simulations. Specifically, 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 (5). Then, given a shrinkage intensity δ, we construct the shrinkage mean-variance portfolio in each simulation m as

$$\hat{\boldsymbol{w}}_{\boldsymbol{D},m} = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D},\boldsymbol{m}}^{-1} \hat{\boldsymbol{\mu}}_m \quad \text{where} \quad \hat{\boldsymbol{\Sigma}}_{\boldsymbol{D},\boldsymbol{m}} = (1-\delta)\hat{\boldsymbol{\Sigma}}_m + \delta \frac{\operatorname{tr}(\hat{\boldsymbol{\Sigma}}_m)}{N} \boldsymbol{I}_N. \tag{SM2}$$

Averaging across all M simulations, the simulated EU for a given δ is

$$\overline{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{M} \sum_{m=1}^{M} \Big(\hat{\boldsymbol{w}}_{\boldsymbol{D},m}' \boldsymbol{\mu} - \frac{\gamma}{2} \hat{\boldsymbol{w}}_{\boldsymbol{D},m}' \boldsymbol{\Sigma} \hat{\boldsymbol{w}}_{\boldsymbol{D},m} \Big). \tag{SM3}$$

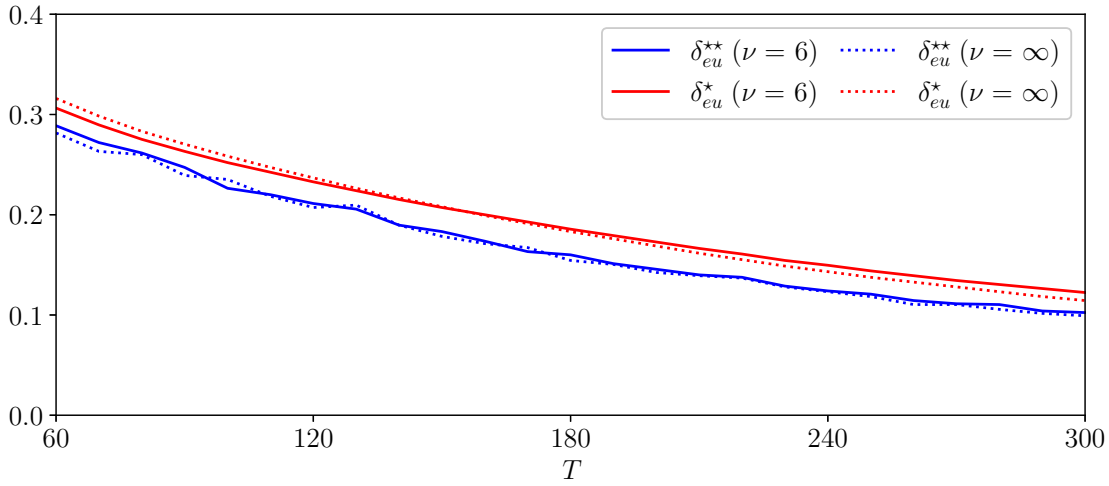
Finally, we find the EU-optimal shrinkage intensity $\delta_{eu}^{\star\star}$ maximizing (SM3),

$$\delta_{eu}^{\star\star} = \operatorname*{argmax}_{\delta \in \mathbb{R}} \overline{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}), \tag{SM4}$$

which we obtain by considering a grid of 1,000 values of δ.

In Figure SM.1, we compare $\delta_{eu}^{\star}$ with $\delta_{eu}^{\star\star}$ for $\nu \in \{6, \infty\}$ and $T \in [60, 300]$. We make two main observations. First, $\delta_{eu}^{\star}$ and $\delta_{eu}^{\star\star}$ are consistently close to one another, indicating that the approximation used in Proposition 3 is accurate. Second, $\delta_{eu}^{\star}$ is consistently larger than $\delta_{eu}^{\star\star}$, indicating that the approximation yields conservative shrinkage intensities. This result is expected because the approximation overstates sampling variability in the shrinkage covariance matrix, as discussed in Remark 5. These results are robust to considering our other empirical datasets.

Figure SM.1: 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 $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (14). The intensity $\delta_{eu}^{\star}$ in (18) (red) is based on the approximation in Proposition 3, whereas the intensity $\delta_{eu}^{\star\star}$ in (SM4) (blue) is without approximation and obtained via Monte Carlo simulations as explained in Section 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 $\nu \in \{6, \infty\}$.

C Shrinkage with a 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 objective function.

C.1 Linear shrinkage with a known covariance matrix

The linear shrinkage covariance matrix when $\boldsymbol{\Sigma}$ is known is

$$\boldsymbol{\Sigma}_{\boldsymbol{D}} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}' \quad \text{with} \quad \boldsymbol{D} = (1-\delta)\boldsymbol{\Lambda} + \delta\mu_\lambda \boldsymbol{I}_N, \ \mu_\lambda = \frac{1}{N}\text{tr}(\boldsymbol{\Lambda}). \tag{SM5}$$

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 (SM5). Unlike the case where we shrink the sample covariance matrix in Proposition 3, we don't need any distributional assumption for the returns and we can have $N \geq T$. To distinguish with the case where 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 SM.2 *The EU in* (13) *of the shrinkage mean-variance portfolio* $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$*, where*

$\boldsymbol{\Sigma_D}$ *is given by* (SM5)*, is*

$$\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{\gamma}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2 - \frac{1}{2\gamma}\sum_{i=1}^{N} f_i(\delta)^2\left(\theta_{PC_i}^2 + \frac{1}{T}\right), \tag{SM6}$$

where $f_i(\delta)$ *is defined in Proposition 3. Moreover, some non-zero shrinkage is always optimal, in the sense that*

$$\left.\frac{\partial\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})}{\partial\delta}\right|_{\delta=0} > 0, \tag{SM7}$$

if T *is finite and the* λ_i*'s are not all equal.*

Like when we shrink the sample covariance matrix, there is no closed-form expression for the EU-optimal shrinkage intensity,

$$\tilde{\delta}_{eu}^{\star} = \underset{\delta\in\mathbb{R}}{\text{argmax}}\,\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) \quad \text{with} \quad \widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = (\text{SM6}), \tag{SM8}$$

but it can easily be found numerically. In Figure SM.2, we depict $\tilde{\delta}_{eu}^{\star}$ using a setting similar to that in Figure 1. For comparison, we also depict the EU-optimal shrinkage intensity under the sample covariance matrix, $\delta_{eu}^{\star}$ in (18). We depict the shrinkage intensities as a function of the sample size $T \in [60, 300]$. We make two main observations. First, $\tilde{\delta}_{eu}^{\star}$ is much larger than zero, which contrasts with $\delta_{mse}^{\star} = 0$ under a known covariance matrix. Second, $\tilde{\delta}_{eu}^{\star}$ is lower than $\delta_{eu}^{\star}$, i.e., we shrink less intensively when the covariance matrix is known, as expected.

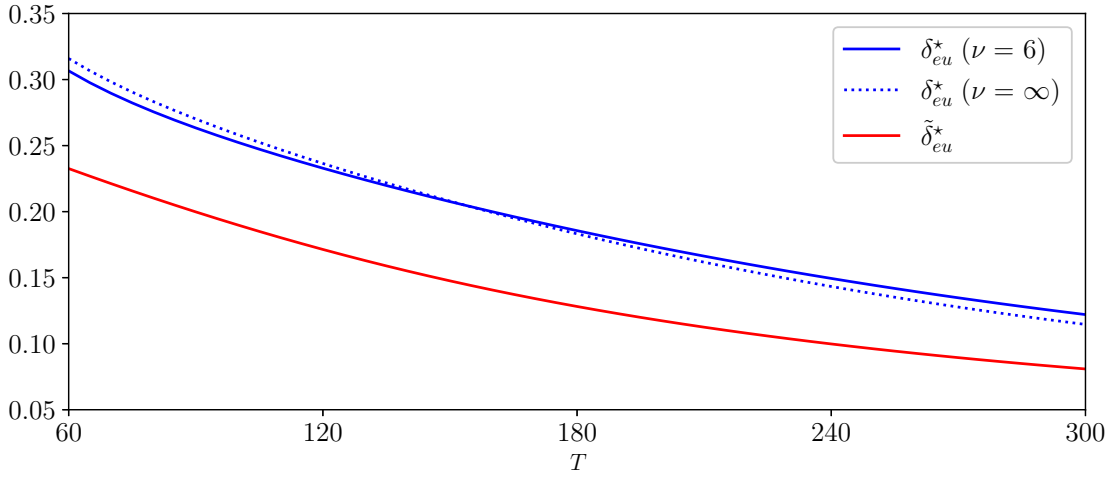
C.2 Nonlinear shrinkage with a known covariance matrix

The nonlinear shrinkage covariance matrix when $\boldsymbol{\Sigma}$ is known is

$$\boldsymbol{\Sigma_D} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}' \quad \text{with} \quad \boldsymbol{D} = (\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}, \tag{SM9}$$

where $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta}) = \text{diag}(\delta_1, \ldots, \delta_N)$ is a diagonal matrix containing the N shrinkage intensities 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 $\boldsymbol{\Sigma_D}$ in (SM9). Unlike the case where we shrink the sample covariance matrix in Proposition 5, we don't need any distributional assumption for the returns and we can have $N \geq T$. As in Section C.1, we denote the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ by $\widetilde{\text{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$.

Figure SM.2: EU-optimal linear shrinkage intensity with a 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 $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where the shrinkage covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (SM5) 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 (18), for degrees of freedom $\nu \in \{6, \infty\}$ of 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.

Proposition SM.3 *The following holds concerning the shrinkage mean-variance portfolio* $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$*, where* $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ *is given by* (SM9)*:*

1. *The EU in* (13) *of* $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ *is*

$$\widetilde{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{1}{\gamma}\sum_{i=1}^{N}(1-\delta_i)\theta^2_{PC_i} - \frac{1}{2\gamma}\sum_{i=1}^{N}(1-\delta_i)^2\left(\theta^2_{PC_i} + \frac{1}{T}\right). \tag{SM10}$$

2. *The* i*-th entry of the vector of shrinkage intensities* $\tilde{\boldsymbol{\delta}}^\star_{eu} \in \mathbb{R}^N$ *maximizing* $\widetilde{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ *in* (SM10) *is*

$$\tilde{\delta}^\star_{eu,i} = \frac{1/T}{\theta^2_{PC_i} + 1/T} \in [0,1], \tag{SM11}$$

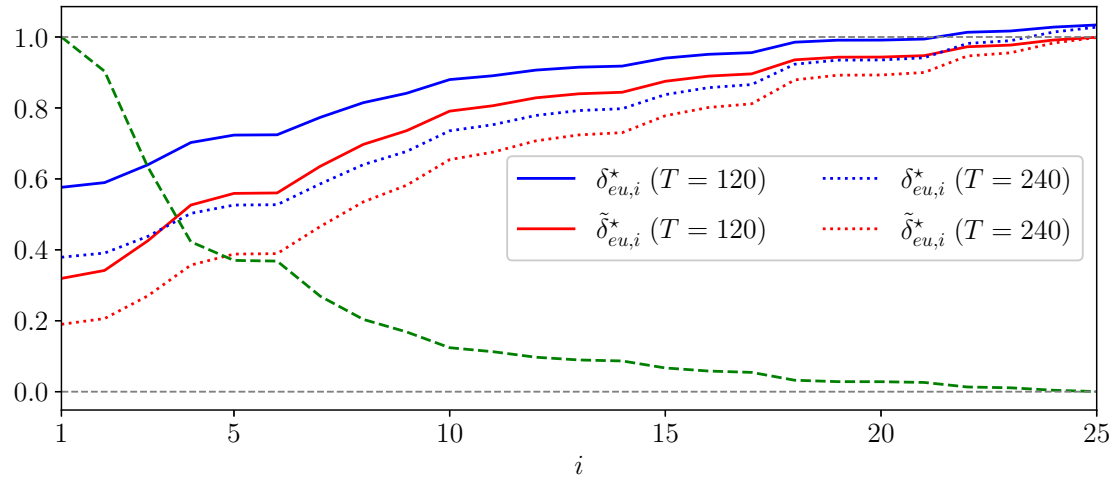
which is lower than the one under the sample covariance matrix, $\delta^\star_{eu,i}$ *in* (23)*.*

3. *Define* $\widetilde{\boldsymbol{\Delta}}^\star_{eu} = \mathrm{diag}(\tilde{\boldsymbol{\delta}}^\star_{eu})$ *and* $\widetilde{\boldsymbol{D}}^\star_{eu} = (\boldsymbol{I}_N - \widetilde{\boldsymbol{\Delta}}^\star_{eu})^{-1}\boldsymbol{\Lambda}$*. The EU in* (SM10) *delivered by* $\tilde{\boldsymbol{\delta}}^\star_{eu}$ *is*

$$\widetilde{\mathrm{EU}}(\hat{\boldsymbol{w}}_{\widetilde{\boldsymbol{D}}^\star_{eu}}) = \frac{1}{2\gamma}\sum_{i=1}^{N}\left(1-\tilde{\delta}^\star_{eu,i}\right)\theta^2_{PC_i} \geq 0. \tag{SM12}$$

Proposition SM.3 shows that even when the covariance matrix $\boldsymbol{\Sigma}$ is known, the eigenvalues with a small corresponding squared Sharpe ratio $\theta^2_{PC_i}$ are shrunk intensively. In particular $\tilde{\delta}^\star_{eu,i} = 1$

Figure SM.3: EU-optimal nonlinear shrinkage intensities with a 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 $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where the shrinkage covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (SM9) 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 (23), for degrees of freedom $\nu \in \{6, \infty\}$ of the multivariate t-distribution. We depict the shrinkage intensities for the eigenvalues λ_i sorted in descending order of their corresponding squared Sharpe ratios $\theta^2_{PC_i}$. We consider a sample size $T \in \{120, 240\}$. 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 of zero and one.

when $\theta^2_{PC_i} = 0$. We also have $\tilde{\delta}^\star_{eu,i} \in [0, 1]$ for all i.

Figure SM.3 depicts the shrinkage intensities $\tilde{\delta}^\star_{eu,i}$ in (SM11) using a setting similar to that in Figure 2. For comparison, we also depict the optimal shrinkage intensities under the sample covariance matrix, $\delta^\star_{eu,i}$ in (23). We depict the shrinkage intensities for a sample size $T \in \{120, 240\}$. We make three main observations. First, the $\tilde{\delta}^\star_{eu,i}$ are much larger than zero, which contrasts with $\delta^\star_{mse,i} = 0$ for all i under a known covariance matrix. Second, in line with Proposition SM.3, the shrinkage intensities $\tilde{\delta}^\star_{eu,i}$ are lower than $\delta^\star_{eu,i}$ for all i, i.e., we shrink less intensively when the covariance matrix is known. Third, $\tilde{\delta}^\star_{eu,i}$ approaches one when $\theta^2_{PC_i}$ approaches zero.

D Unconstrained MSE-optimal nonlinear shrinkage

In Proposition 4, we derive the MSE-optimal nonlinear shrinkage intensities $\boldsymbol{\delta}^\star_{mse}$ under the constraint that the trace of the sample covariance matrix is preserved, i.e., $\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}})$. In this section, we derive the unconstrained MSE-optimal nonlinear shrinkage intensities and show that they are not well-behaved and underperform the constrained shrinkage intensities empirically. Thus,

the trace constraint renders the MSE-optimal nonlinear shrinkage estimator a tougher benchmark.

In the next proposition, we start by deriving the unconstrained counterpart of the MSE-optimal nonlinear shrinkage intensities in Proposition 4 of the paper.

Proposition SM.4 *Assume that the nonlinear shrinkage estimator* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *in* (19) *is obtained from an i.i.d. sample* $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ *drawn from the multivariate elliptical distribution, i.e.,* $\boldsymbol{x}_t \sim \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{D}}, \tau)$ *with* $\boldsymbol{\Sigma}_{\boldsymbol{D}} = \boldsymbol{V}(\boldsymbol{I}_N - \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}\boldsymbol{V}'$ *the population counterpart of* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$*. Then, the* i*-the entry of the vector of unconstrained shrinkage intensities* $\boldsymbol{\delta}^{\star}_{mse,unc} \in \mathbb{R}^N$ *minimizing the MSE in* (12) *of* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ *is*

$$\delta^{\star}_{mse,unc,i} = \frac{\left(\frac{N}{T-1} + \frac{\kappa N}{T}\right)\mu_\lambda + \left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)\left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)\left(1 + \frac{N}{T-1} + \frac{\kappa N}{T}\right)\lambda_i}{\left(\frac{N}{T-1} + \frac{\kappa N}{T}\right)\mu_\lambda - \left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)\left(1 + \frac{N}{T-1} + \frac{\kappa N}{T}\right)\lambda_i}. \tag{SM13}$$

The constrained intensities $\delta^{\star}_{mse,i}$ *in* (20) *and the unconstrained intensities* $\delta^{\star}_{mse,unc,i}$ *in* (SM13) *coincide as* $\lambda_i \to 0$ *or* ∞*. Specifically, we have*

$$\lim_{\lambda_i \to 0} \delta^{\star}_{mse,i} = \lim_{\lambda_i \to 0} \delta^{\star}_{mse,unc,i} = 1, \quad \textit{and} \tag{SM14}$$

$$\lim_{\lambda_i \to \infty} \delta^{\star}_{mse,i} = \lim_{\lambda_i \to \infty} \delta^{\star}_{mse,unc,i} = -\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right). \tag{SM15}$$

However, $\delta^{\star}_{mse,unc,i} \to \pm\infty$ *as* $\lambda_i/\mu_\lambda \to \zeta$ *from below (above), where*

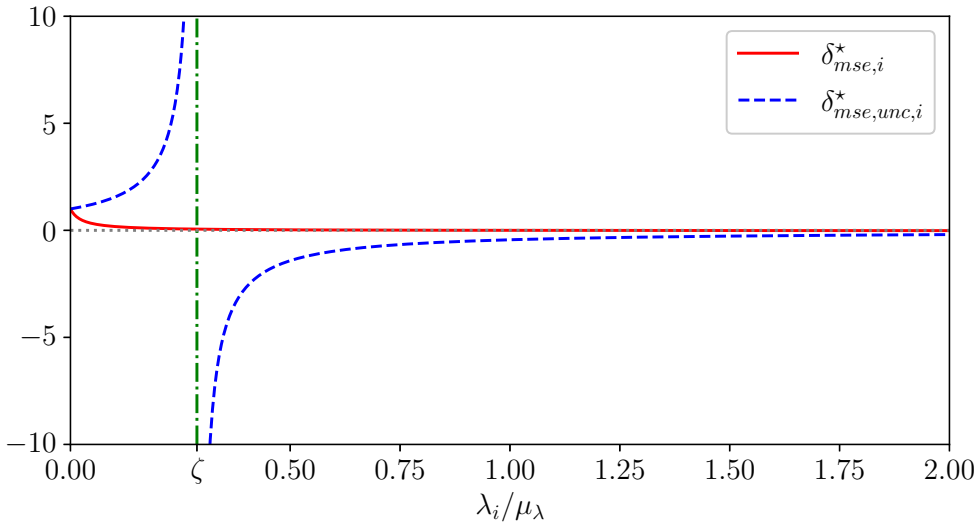
$$\zeta = \frac{\frac{N}{T-1} + \frac{\kappa N}{T}}{\left(1 + \frac{N}{T-1} + \frac{\kappa N}{T}\right)\left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)} < 1 \tag{SM16}$$

is the value of λ_i/μ_λ *for which the denominator of* (SM13) *is equal to zero.*

Proposition SM.4 shows that the unconstrained MSE-optimal nonlinear shrinkage intensities $\delta^{\star}_{mse,unc,i}$ are not well-behaved, as they diverge to $\pm\infty$ as λ_i/μ_λ tends to ζ in (SM16). We illustrate this result in Figure SM.4, in which we depict the constrained and unconstrained intensities, i.e., $\delta^{\star}_{mse,i}$ in (20) and $\delta^{\star}_{mse,unc,i}$ in (SM13), as a function of λ_i/μ_λ for $N = 25$, $T = 120$, and $\kappa = 1$.

Finally, in Table SM.1, we report the annualized empirical net out-of-sample utility delivered by mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ is constructed using four types of nonlinear shrinkage intensities. First, the EU-optimal unconstrained intensities in (23). Second, the constrained EU-optimal intensities, obtained by maximizing (22) under the constraint that $\sum_{i=1}^{N}(1 - (1 - \delta_i)^{-1})\lambda_i = 0$. Third, the MSE-optimal unconstrained intensities in (SM13). Fourth, the MSE-optimal constrained intensities in (20). We introduce bona

Figure SM.4: Constrained versus unconstrained MSE-optimal nonlinear shrinkage intensities



Notes. This figure depicts the constrained nonlinear shrinkage intensities $\delta^\star_{mse,i}$ in (20) (solid red) and the unconstrained nonlinear shrinkage intensities $\delta^\star_{mse,unc,i}$ in (SM13) (dashed blue), as a function of the ratio $\lambda_i/\mu_\lambda \in [0, 2]$. Both intensities minimize the MSE of the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (19), but $\delta^\star_{mse,i}$ is constrained to preserve the trace of the sample covariance matrix, i.e., $\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \text{tr}(\hat{\boldsymbol{\Sigma}})$. The vertical dash-dotted green line depicts ζ, the value of λ_i/μ_λ for which the denominator of $\delta^\star_{mse,unc,i}$ in (SM13) is equal to zero. We calibrate μ_λ to a dataset of $N = 25$ portfolios sorted on size and book-to-market spanning July 1926 to July 2024, and set $T = 120$ and $\kappa = 1$. The horizontal gray dashed line depicts a shrinkage intensity of zero.

fide estimators of these shrinkage intensities in Section 5.1. We report net utilities across our six datasets, a sample size $T = 120$ and 240, and a risk-aversion coefficient $\gamma = 3$ and 10. We observe that for MSE-optimal shrinkage, the unconstrained intensities deliver extreme negative net utilities, due to the divergence of the intensities identified in Proposition SM.4 and Figure SM.4. The MSE-optimal constrained intensities solve that issue and restore more satisfying net utilities. For EU-optimal shrinkage, there is no clear winner between the constrained and unconstrained intensities, as they each outperform in half of the datasets.

E Positive nonlinearly shrunk eigenvalues

We demonstrate in Proposition 5, and illustrate in Figure 2, that the EU-optimal nonlinear shrinkage intensities $\delta^\star_{eu,i}$ in (23) are larger than one provided that $\theta^2_{PC_i} < \bar{\theta}^2_{PC}$ in (27). 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 keeping all nonlinear shrinkage intensities below one, without sacrificing the empirical performance.

Table SM.1: Annualized net out-of-sample utility in empirical data (in percentage points) delivered by constrained versus unconstrained nonlinear shrinkage intensities

Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	$T = 120$						$T = 240$					
		Dataset						Dataset					
		25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV	25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL-U	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-C	6.52	8.45	−15.2	70.2	4.40	58.3	10.9	8.99	−8.83	66.2	6.59	20.3
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL-U	−2E+10	−6E+09	−7E+11	−1E+09	−3327	−6E+07	−2E+07	−2E+09	−1E+14	−5E+13	−1E+10	−2E+10
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL-C	−1.93	6.12	−37.9	24.3	−5.30	−13.7	4.38	7.93	−15.5	53.3	4.73	−159
Panel (b): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL-U	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-C	1.28	2.44	−5.78	20.9	1.12	16.6	3.11	2.64	−2.88	20.1	1.92	6.16
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL-U	−4E+08	−1E+08	−2E+10	−3E+07	−1236	−1E+06	−3E+05	−5E+07	−4E+12	−1E+12	−3E+08	−5E+08
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL-C	−0.64	1.82	−11.5	6.86	−1.60	−4.98	1.28	2.37	−4.70	15.7	1.41	−48.9

Notes. This table reports the annualized net out-of-sample utility, in percentage points, delivered by mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$, where the shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ is constructed using four types of nonlinear shrinkage intensities. First, the EU-optimal unconstrained intensities in (23) (EUNL-U). Second, the constrained EU-optimal intensities, obtained by maximizing (22) under the constraint that $\sum_{i=1}^{N}(1-(1-\delta_i)^{-1})\lambda_i = 0$ (EUNL-C). Third, the MSE-optimal unconstrained intensities in (SM13) (MSENL-U). Fourth, the MSE-optimal constrained intensities in (20) (MSENL-C). The constraint we impose is that the trace of the sample covariance matrix is preserved by the shrinkage estimator. We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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).

The first method we propose starts from the observation that the sum of the $\delta^{\star}_{eu,i}$ in (23) is equal to

$$\sum_{i=1}^{N}\delta^{\star}_{eu,i} = N - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)}\sum_{i=1}^{N}R_i, \tag{SM17}$$

where $R_i = k_1\theta^2_{PC_i}/(k_2\theta^2_{PC_i}+k_3/T)$. Therefore, we propose to use the vector of shrinkage intensities $\boldsymbol{\delta}^{\circ}_{eu} \in [0,1]^N$ whose i-th entry is given by

$$\delta^{\circ}_{eu,i} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)}R_i, \tag{SM18}$$

which satisfies $\delta^{\circ}_{eu,i} \in [0,1]$, 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 the intensities $\delta^{\circ}_{eu,i}$.

Proposition SM.5 *The following holds concerning the vector of nonlinear shrinkage intensities* $\boldsymbol{\delta}^{\circ}_{eu}$ *in* (SM18)*:*

1. *The difference between* $\delta^{\circ}_{eu,i}$ *and the* i*-th optimal nonlinear shrinkage intensity* $\delta^{\star}_{eu,i}$ *in* (23) *is*

$$\delta^{\circ}_{eu,i} - \delta^{\star}_{eu,i} = \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\big(R_i - \bar{R}\big). \tag{SM19}$$

We have $\delta^{\circ}_{eu,i} \geq \delta^{\star}_{eu,i}$ *if* $R_i \geq \bar{R}$*, and* $\delta^{\circ}_{eu,i} \leq \delta^{\star}_{eu,i}$ *otherwise.*

2. *Define* $\boldsymbol{\Delta}^{\circ}_{eu} = \text{diag}(\boldsymbol{\delta}^{\circ}_{eu})$ *and* $\boldsymbol{D}^{\circ}_{eu} = (\boldsymbol{I}_N - \boldsymbol{\Delta}^{\circ}_{eu})^{-1}\hat{\boldsymbol{\Lambda}}$*. The EU in* (22) *delivered by* $\boldsymbol{\delta}^{\circ}_{eu}$ *is*

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}^{\circ}_{eu}}) = \frac{k_1(T-1)(T+N-2)}{2\gamma(T-2)(T-N-2)}\sum_{i=1}^{N}(1-\delta^{\circ}_{eu,i})\left(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2}\right) \geq 0. \tag{SM20}$$

Part 1 of Proposition SM.5 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 SM.5 shows that $\boldsymbol{\delta}^{\circ}_{eu}$ always delivers a positive EU even though it is not EU-optimal.

The second method we propose to keep the nonlinear shrinkage intensities below one is to solve the following constrained quadratic optimization problem:

$$\boldsymbol{\delta}^{+}_{eu} = \underset{\boldsymbol{\delta}\in\mathbb{R}^N}{\text{argmax}}\,\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) \quad \text{subject to} \quad \delta_i \leq 1\ \forall i \quad \text{with} \quad \text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = (22), \tag{SM21}$$

where the $+$ subscript refers to the shrunk eigenvalues being non-negative and $\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ is a quadratic function of $\boldsymbol{\delta}$. In the next proposition, we derive an expression for $\boldsymbol{\delta}^{+}_{eu}$.

Proposition SM.6 *Let* $\eta_i \geq 0$ *be the Lagrange multiplier corresponding to the* i*-th inequality constraint in* (SM21) *at the optimum. Define*

$$R^{+}_i = \frac{k_1(\theta^2_{PC_i} + \eta_i)}{k_2\theta^2_{PC_i} + \frac{k_3}{T}} \geq R_i, \tag{SM22}$$

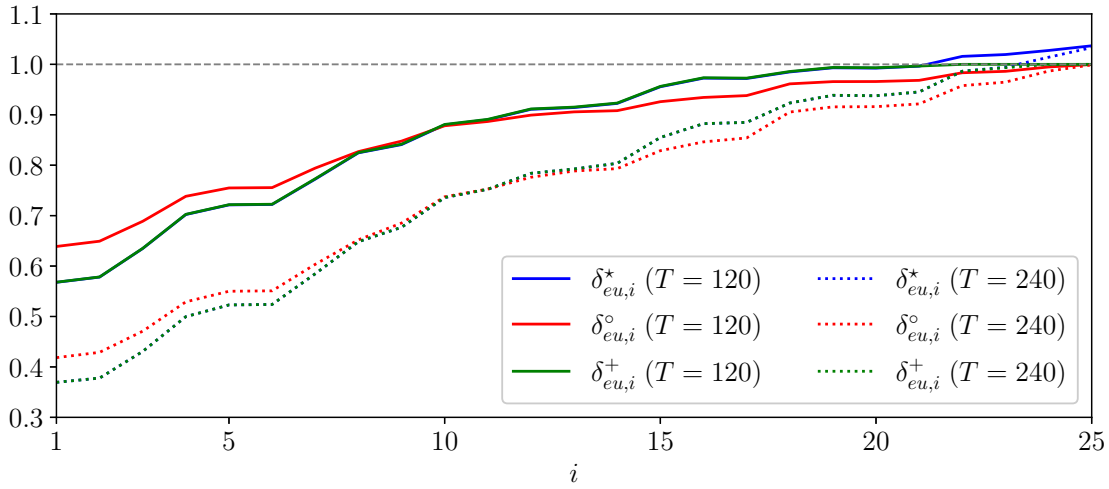
and $\bar{R}^{+} = \frac{1}{N}\sum_{i=1}^{N} R^{+}_i$*. Then, the* i*-th entry of the vector of constrained nonlinear shrinkage intensities* $\boldsymbol{\delta}^{+}_{eu}$ *in* (SM21) *is*

$$\delta^{+}_{eu,i} = 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R^{+}_i - \frac{N}{T-2}\bar{R}^{+}\right), \tag{SM23}$$

and it satisfies

$$\delta^{+}_{eu,i} = \begin{cases} 1 & \text{if} \quad \eta_i > 0, \\ \delta^{\star}_{eu,i} + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\big(\bar{R}^{+} - \bar{R}\big) \geq \delta^{\star}_{eu,i} & \text{if} \quad \eta_i = 0, \end{cases} \tag{SM24}$$

Figure SM.5: 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 $\delta^{+}_{eu,i}$ in Propositions 5, SM.5, and SM.6, respectively. The $\delta^{\star}_{eu,i}$ maximize the EU in (22) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (19). The $\delta^{\circ}_{eu,i}$ are a dampened version of the $\delta^{\star}_{eu,i}$ such that $\delta^{\circ}_{eu,i} \in [0,1]$ and $\sum_{i=1}^{N}\delta^{\circ}_{eu,i} = \sum_{i=1}^{N}\delta^{\star}_{eu,i}$. The $\delta^{+}_{eu,i}$ also maximize the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$, but under the constraint that $\delta^{+}_{i} \leq 1\ \forall i$. 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 a sample size $T \in \{120, 240\}$. 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 $\nu \in \{6, \infty\}$.

where $\delta^{\star}_{eu,i}$ in (23) is the i-th unconstrained EU-optimal nonlinear shrinkage intensity.

Proposition SM.6 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.

Figure SM.5 depicts the nonlinear shrinkage intensities $\delta^{\star}_{eu,i}$, $\delta^{\circ}_{eu,i}$, and $\delta^{+}_{eu,i}$ using a setting similar to that in Figure 2, 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 (SM19), $\delta^{+}_{eu,i}$ and $\delta^{\star}_{eu,i}$ are almost equal for the indices i such that $\delta^{\star}_{eu,i} < 1$. Overall, the three nonlinear shrinkage methods are reasonably close.

Next, we compare the empirical performance delivered by $\boldsymbol{\delta}^{\star}_{eu}$, $\boldsymbol{\delta}^{\circ}_{eu}$, and $\boldsymbol{\delta}^{+}_{eu}$. Table SM.2 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 6. We observe that the three nonlinear shrinkage covariance matrix estimators deliver a similar empirical performance. If anything, the constrained intensities $\boldsymbol{\delta}^{+}_{eu}$ perform slightly better overall.

Table SM.2: Annualized net out-of-sample utility in empirical data (in percentage points) delivered by the nonlinear shrinkage intensities $\boldsymbol{\delta}^{\star}_{eu}$, $\boldsymbol{\delta}^{\circ}_{eu}$, and $\boldsymbol{\delta}^{+}_{eu}$

Portfolio	$\boldsymbol{\delta}$	$T = 120$						$T = 240$					
		Dataset						Dataset					
		25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV	25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\boldsymbol{\delta}^{\star}_{eu}$	9.36	11.0	3.42	20.5	2.82	10.5	12.2	11.0	6.88	16.6	3.24	6.28
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\boldsymbol{\delta}^{\circ}_{eu}$	9.08	10.9	3.52	20.5	2.80	10.4	12.2	10.9	6.72	16.6	3.23	6.26
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\boldsymbol{\delta}^{+}_{eu}$	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): $\gamma = 10$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\boldsymbol{\delta}^{\star}_{eu}$	2.80	3.31	1.03	6.15	0.85	3.15	3.66	3.28	2.06	4.99	0.97	1.89
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\boldsymbol{\delta}^{\circ}_{eu}$	2.72	3.27	1.06	6.16	0.84	3.13	3.65	3.27	2.02	4.99	0.97	1.88
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	$\boldsymbol{\delta}^{+}_{eu}$	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 $\boldsymbol{\delta}^{\star}_{eu}$, $\boldsymbol{\delta}^{\circ}_{eu}$ and $\boldsymbol{\delta}^{+}_{eu}$ in Propositions 5, SM.5, and SM.6, respectively. $\boldsymbol{\delta}^{\star}_{eu}$ maximizes the EU in (22) of the mean-variance portfolio $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ given by (19). $\boldsymbol{\delta}^{\circ}_{eu}$ is a dampened version of $\boldsymbol{\delta}^{\star}_{eu}$ such that $\delta^{\circ}_{eu,i} \in [0, 1]$ for all i and $\sum_{i=1}^{N} \delta^{\circ}_{eu,i} = \sum_{i=1}^{N} \delta^{\star}_{eu,i}$. $\boldsymbol{\delta}^{+}_{eu}$ also maximizes the EU of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$, but under the constraint that $\delta^{+}_{i} \leq 1$ for all i. We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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).

F Empirical performance with a robust mean estimator

In the paper, we argue that calibrating shrinkage estimators using the EU objective function is relevant because it accounts for the impact of estimation errors in both $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ on the mean-variance portfolio, whereas traditional MSE-optimal shrinkage estimators only consider errors in $\hat{\boldsymbol{\Sigma}}$. In this section, we evaluate empirically whether relying on our EU-optimal linear and nonlinear shrinkage estimators is more favorable than using a conventional MSE-optimal shrinkage estimator of $\boldsymbol{\Sigma}$ jointly with a robust estimator of $\boldsymbol{\mu}$ that alleviates estimation errors in $\hat{\boldsymbol{\mu}}$.

Table SM.3 reports the annualized net out-of-sample utility delivered by the first nine mean-variance portfolios in Table 2 that use shrinkage covariance matrices of the form $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ for different choices of the diagonal matrix $\boldsymbol{D}$. For each portfolio, we plug either the sample mean estimator or the Bayes-Stein shrinkage mean estimator of Jorion (1986). We make two main observations from Table SM.3. First, our shrinkage portfolios EUL and EUNL continue to largely outperform the benchmark mean-variance portfolios, no matter if we use the sample mean estimator or the Jorion mean estimator, thus addressing our main question in this section. For example, considering

Table SM.3: Annualized net out-of-sample utility in empirical data (in percentage points) delivered by the sample mean versus a robust mean estimator

Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	$\hat{\boldsymbol{\mu}}$	$T = 120$						$T = 240$					
			Dataset						Dataset					
			25	10	25	13	16	46	25	10	25	13	16	46
			SBM	MOM	OPIN	JKP	NMV	NMV	SBM	MOM	OPIN	JKP	NMV	NMV
Panel (a): $\gamma = 3$														
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL	SAMPLE	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	JORION	8.56	**11.5**	0.01	**48.6**	3.38	**65.6**	10.9	**11.7**	6.34	57.6	7.21	**18.4**
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	SAMPLE	**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	JORION	8.13	9.29	**3.43**	21.2	1.86	9.09	11.0	9.95	**6.89**	16.9	2.63	5.88
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	SAMPLE	−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	JORION	5.51	10.2	−14.8	23.6	1.04	7.01	8.50	10.6	−0.04	54.0	8.85	−130
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	SAMPLE	−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	JORION	7.75	10.9	−10.7	30.7	2.14	41.7	10.1	11.1	1.37	57.6	**9.44**	−87.9
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	SAMPLE	−109	−14.9	−126	−207	−48.0	−1467	−29.7	0.35	−34.4	−26.3	−11.4	−610
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	JORION	−37.2	2.52	−43.5	−163	−13.3	−968	−5.64	7.51	−5.52	−6.91	3.05	−427
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	SAMPLE	−8.68	4.33	−48.0	24.7	−8.67	−74.0	0.58	6.83	−18.9	54.0	3.29	−215
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	JORION	5.42	10.3	−14.5	31.4	1.10	6.08	8.63	10.6	0.18	**58.3**	9.13	−128
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	SAMPLE	−45.5	−7.32	−61.8	−106	−28.8	−338	−17.9	2.02	−21.0	−5.72	−6.03	−375
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	JORION	−8.22	6.12	−19.3	−78.5	−5.46	−173	0.22	8.53	−0.21	10.4	5.51	−250
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	SAMPLE	−40.6	−3.82	−84.6	−32.7	−22.1	−411	−15.9	3.11	−28.7	17.0	−4.56	−419
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	JORION	−6.99	7.16	−27.9	−18.3	−3.40	−219	1.20	8.90	−3.41	28.4	5.92	−282
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	SAMPLE	1.90	7.79	2.00	15.2	−0.88	3.22	0.24	8.31	3.11	6.27	−2.28	−8.64
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	JORION	2.24	8.49	3.12	22.3	−1.61	5.96	0.63	8.96	3.84	10.8	−1.62	−4.63
Panel (b): $\gamma = 10$														
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUL	SAMPLE	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	JORION	2.56	**3.46**	−0.01	**14.5**	1.01	**19.7**	3.26	**3.52**	1.90	**17.2**	2.16	**5.48**
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	EUNL	SAMPLE	**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	JORION	2.44	2.79	**1.03**	6.38	0.56	2.73	3.31	2.98	**2.07**	5.07	0.79	1.76
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSEL	SAMPLE	−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	JORION	1.62	3.05	−4.50	6.63	0.31	1.34	2.53	3.16	−0.03	15.9	2.65	−39.9
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	MSENL	SAMPLE	−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	JORION	2.30	3.26	−3.26	8.84	0.64	12.1	3.02	3.33	0.40	17.0	**2.83**	−27.1
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	SAMPLE	−33.5	−4.53	−38.2	−66.0	−14.6	−488	−9.10	0.09	−10.4	−8.73	−3.45	−190
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	SAMPLE	JORION	−11.5	0.74	−13.2	−52.2	−4.04	−320	−1.78	2.24	−1.68	−2.77	0.90	−133
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	SAMPLE	−2.69	1.29	−14.5	6.94	−2.62	−23.7	0.13	2.04	−5.72	15.9	0.98	−66.3
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2004	JORION	1.59	3.08	−4.38	9.05	0.32	1.05	2.57	3.18	0.04	17.2	2.74	−39.5
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	SAMPLE	−14.0	−2.24	−18.7	−34.0	−8.71	−107	−5.50	0.59	−6.35	−2.38	−1.84	−116
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	LW2020	JORION	−2.57	1.82	−5.84	−25.3	−1.66	−55.4	0.00	2.55	−0.08	2.57	1.65	−77.0
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	SAMPLE	−12.5	−1.18	−25.6	−10.8	−6.69	−130	−4.89	0.92	−8.66	4.64	−1.39	−130
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	BGP	JORION	−2.20	2.14	−8.45	−6.33	−1.04	−69.9	0.30	2.66	−1.04	8.12	1.77	−87.2
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	SAMPLE	0.57	2.34	0.60	4.47	−0.26	0.96	0.07	2.49	0.93	1.82	−0.68	−2.61
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$	PCA	JORION	0.67	2.55	0.93	6.64	−0.48	1.79	0.19	2.69	1.15	3.18	−0.48	−1.40

Notes.. This table reports the annualized net-of-cost out-of-sample utility, in percentage points, for 18 mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ using rotation-equivariant covariance matrices $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}\boldsymbol{D}\hat{\boldsymbol{V}}'$ for nine different choices of the diagonal matrix $\boldsymbol{D}$ described in Table 2, as well as two different choices of mean-return estimators $\hat{\boldsymbol{\mu}}$. In the third column, SAMPLE denotes the sample mean estimator in (5) and JORION denotes the Bayes-Stein estimator of Jorion (1986). EUL and EUNL are the EU-optimal linear and nonlinear shrinkage covariance matrix estimators introduced in the paper. We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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 depict the best strategy for each dataset, sample size, and γ in bold, and we underline the second-best strategy.

the Ledoit and Wolf (2004) shrinkage covariance matrix, the median annualized net out-of-sample utility across datasets is -4.05 under the sample mean estimator and 7.36 under the Jorion mean estimator, respectively, for $\gamma = 3$. In comparison, considering our EUL shrinkage covariance matrix, the median annualized net out-of-sample utility across datasets is 9.61 under the sample mean estimator and 11.2 under the Jorion mean estimator, respectively. For EUNL, the medians are 9.93 under the sample mean and 8.61 under Jorion, respectively.

The second main observation we make from Table SM.3 is that for all estimators $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ except EUL and EUNL, the Jorion mean estimator systematically outperforms the sample estimator in (5), with only one exception across all configurations. This result underlines that, when using traditional estimators of the covariance matrix that do not account for estimation errors in sample mean returns, it is valuable to rely on a robust estimator of the mean vector. In contrast, under our proposed EUL and EUNL covariance matrix estimators, which account for estimation errors in sample mean returns, it is no longer systematically more favorable to use the Jorion mean estimator. In particular, under the EUNL covariance matrix estimator, the sample mean estimator outperforms the Jorion estimator in two thirds of the configurations, and when it underperforms, it does so by a thin margin. This shows that, with our EUNL shrinkage covariance matrix estimator, using the sample mean estimator is no longer a concern.

G Empirical performance under the normality assumption

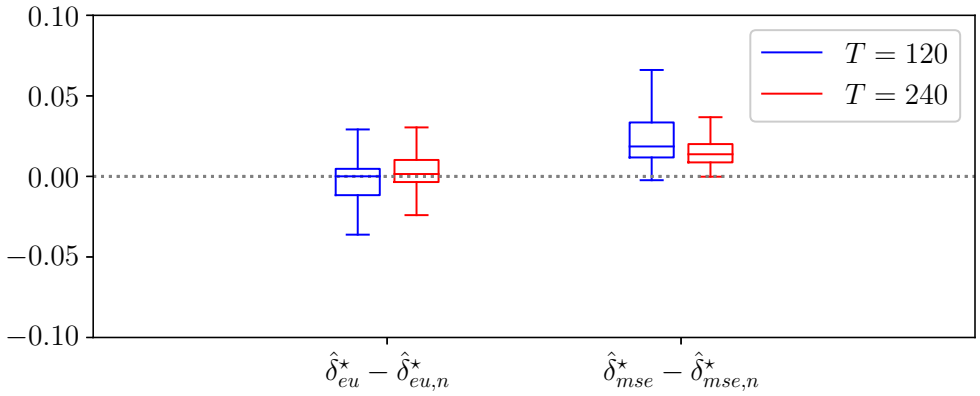
In this section, we test whether the MSE-optimal and EU-optimal linear and nonlinear shrinkage intensities derived under the multivariate elliptical distribution substantial improve the empirical performance compared to deriving them under the simpler multivariate normal distribution. Specifically, in Table SM.4, 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 2 and 3, as well as the nonlinear shrinkage intensities $\boldsymbol{\delta}^{\star}_{mse}$ and $\boldsymbol{\delta}^{\star}_{eu}$ studied in Propositions 4 and 5, under two estimation methods. First, the method introduced in Section 5.1, i.e., assuming a multivariate elliptical distribution for the asset returns (MSEL, EUL, MSENL, and EUNL). Second, an alternative method that assumes a multivariate normal distribution for the asset returns, i.e., we set $k_1 = k_2 = k_3 = 1$ and $\kappa = 0$ (MSEL-N, EUL-N, MSENL-N, and EUNL-N).

Table SM.4: Annualized net out-of-sample utility in empirical data (in percentage points) delivered by the multivariate elliptical versus normal assumption

Portfolio	$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$	$T = 120$						$T = 240$					
		Dataset						Dataset					
		25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV	25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV
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 $\delta^{\star}_{eu}$ in Propositions 2 and 3, as well as the nonlinear shrinkage intensities $\boldsymbol{\delta}^{\star}_{mse}$ and $\boldsymbol{\delta}^{\star}_{eu}$ studied in Propositions 4 and 5. For each shrinkage intensity, we use two estimation methods. First, the method introduced in Section 5.1 that assumes a multivariate elliptical distribution for the asset returns (MSEL, EUL, MSENL, and EUNL). Second, an alternative method that assumes a multivariate normal distribution for the asset returns, i.e., we set $k_1 = k_2 = k_3 = 1$ and $\kappa = 0$ (MSEL-N, EUL-N, MSENL-N, and EUNL-N). We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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 SM.6: Comparison of bona fide linear shrinkage intensities under the multivariate elliptical versus normal distributional assumption



Notes. This figure depicts boxplots of the difference between the bona fide linear shrinkage intensities estimated under the multivariate elliptical distribution, $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$, as detailed in Section 5.1, 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 $\kappa = 0$ for the latter. The bona fide shrinkage intensities are obtained using 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 1. The horizontal gray dashed line corresponds to the zero level.

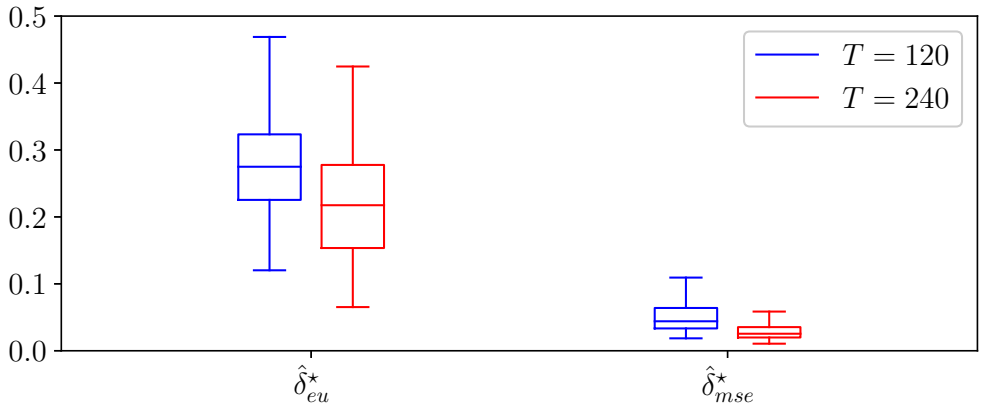
We observe from Table SM.4 that the MSE-optimal shrinkage intensities underperform systematically and significantly when they are estimated under the multivariate normal distribution relative to the elliptical distribution. This is because the MSE leads to rather small shrinkage intensities, and thus, 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.

To further test this distributional robustness result, Figure SM.6 depicts boxplots of the difference between the bona fide 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 bona fide 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 $\hat{\delta}^{\star}_{mse,n}$.

H Stability of shrinkage intensities and impact on turnover

In this section, we assess the empirical stability of our bona fide EU-optimal and MSE-optimal linear and nonlinear shrinkage intensities across different estimation windows, and its impact on

Figure SM.7: Comparison of bona fide EU and MSE-optimal linear shrinkage intensities



Notes. This figure depicts boxplots of the bona fide EU-optimal and MSE-optimal linear shrinkage intensities, $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$. The estimation methodology is detailed in Section 5.1. The boxplots are obtained using rolling windows of size $T = 120$ in blue and $T = 240$ in red, and they are averaged across the six datasets listed in Table 1.

portfolio turnover. In particular, because the bona fide EU-optimal intensities depend on the sample mean returns $\hat{\boldsymbol{\mu}}$, one concern is that they might be more unstable than the bona fide MSE-optimal intensities, which could increase portfolio turnover.

We start by comparing the EU-optimal and MSE-optimal linear shrinkage intensities across estimation windows. We focus on the linear intensities for conciseness because they suffice to make our point. Figure SM.7 depicts boxplots of the bona fide linear shrinkage intensities $\hat{\delta}^{\star}_{eu}$ and $\hat{\delta}^{\star}_{mse}$ following the methodology detailed in Section 5.1, which we average across our six datasets. We consider a sample size $T = 120$ and 240. We make two main observations. First, as expected, $\hat{\delta}^{\star}_{eu}$ is larger than $\hat{\delta}^{\star}_{mse}$, and both are larger for $T = 120$ than 240. Second, $\hat{\delta}^{\star}_{eu}$ is considerably more volatile 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 more volatile shrinkage intensities translates into a higher turnover. To that end, Table SM.5 reports the average monthly turnover of the 21 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. That is, 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. Table SM.5 further reveals that EUL (EUNL) achieves the third-lowest (second-lowest) turnover among the nine mean-variance portfolios $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ based on nine different covariance matrix estimators. As discussed in Section 6.4, the PCA estimator achieves the lowest turnover among the

Table SM.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 NMV	46 NMV	25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (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 (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 of the 21 portfolio strategies described in Section 6.2 and listed in Table 2. The first nine of them are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ 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 paper. We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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).

nine estimators because it completely removes low-variance PCs.

I Empirical out-of-sample Sharpe ratio

In Table 3, we report the out-of-sample utility as an empirical performance measure because it is the one we consider in our theory. Another important mean-variance portfolio performance measure is the Sharpe ratio. Therefore, in Table SM.6, we report the annualized out-of-sample Sharpe ratio net of proportional transaction costs of the 21 portfolio strategies considered in Table 3,

$$SR_k = \sqrt{12} \times \frac{\hat{\mu}_k}{\hat{\sigma}_k}, \tag{SM25}$$

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 (46), 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 Table SM.6 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 optimal linear and nonlinear shrinkage estimators of the sample covariance matrix that maximize the expected out-of-sample Sharpe ratio of the mean-variance portfolio.

J Proofs of theoretical results

Proof of Proposition 1

This proposition corresponds to Kan and Lassance (2025, Proposition 7).

Proof of Proposition 2

The linear shrinkage covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (14) depends on the shrinkage intensity δ as $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = (1-\delta)\hat{\boldsymbol{\Sigma}} + \delta\hat{\mu}_\lambda \boldsymbol{I}_N$, where $\hat{\boldsymbol{\Sigma}}$ in (5) is the sample covariance matrix and $\hat{\mu}_\lambda = \frac{1}{N}\text{tr}(\hat{\boldsymbol{\Sigma}})$. Therefore, the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$, as defined in (12), is

$$N \times \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^2)\big] + \text{tr}(\boldsymbol{\Sigma}^2) - 2\mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}\boldsymbol{\Sigma})\big]$$

Table SM.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 NMV	46 NMV	25 SBM	10 MOM	25 OPIN	13 JKP	16 NMV	46 NMV
Panel (a): $\gamma = 3$													
$\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ in (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 (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-of-cost out-of-sample Sharpe ratio, in percentage points, for the 21 portfolio strategies described in Section 6.2 and listed in Table 2. The first nine of them are mean-variance portfolios of the form $\hat{\boldsymbol{w}}_{\boldsymbol{D}} = \gamma^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\mu}}$ 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 EU-optimal linear and nonlinear shrinkage estimators introduced in this paper. We report results across the six datasets described in Table 1 and we follow the out-of-sample methodology detailed in Section 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 net 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.

$$
\begin{aligned}
&= (1-\delta)^2\mathbb{E}\big[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + \delta^2 N\mathbb{E}\big[\hat{\mu}_\lambda^2\big] + 2\delta(1-\delta)N\mathbb{E}\big[\hat{\mu}_\lambda^2\big] \\
&\quad + \mathrm{tr}(\boldsymbol{\Sigma}^2) - 2(1-\delta)\mathrm{tr}(\boldsymbol{\Sigma}^2) - 2\delta N\mu_\lambda^2 \\
&= (1-\delta)^2\mathbb{B}\big[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + \delta^2\big(\mathrm{tr}(\boldsymbol{\Sigma}^2) - N\mu_\lambda^2\big) + \delta(2-\delta)N\mathbb{B}\big[\hat{\mu}_\lambda^2\big], \\
&= (1-\delta)^2\mathbb{B}\big[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + \delta^2 N(N-1)s\mu_\lambda^2 + \delta(2-\delta)N\mathbb{B}\big[\hat{\mu}_\lambda^2\big], \qquad \text{(SM26)}
\end{aligned}
$$

where $\mathbb{B}$ is the bias operator and $s \in [0,1)$ is the sphericity parameter defined in (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 $\mathrm{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$ in (SM26) is

$$
\delta^\star_{mse} = \underset{\delta\in\mathbb{R}}{\mathrm{argmin}}\,\mathrm{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \frac{\mathbb{B}\big[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] - N\mathbb{B}\big[\hat{\mu}_\lambda^2\big]}{\mathbb{B}\big[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] + N(N-1)s\mu_\lambda^2 - N\mathbb{B}\big[\hat{\mu}_\lambda^2\big]}. \qquad \text{(SM27)}
$$

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

$$
\mathbb{B}\big[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2)\big] = N\mu_\lambda^2\left(\frac{N+1+(N-1)s}{T-1} + \frac{\kappa(N+2+2(N-1)s)}{T}\right), \qquad \text{(SM28)}
$$

$$
N\mathbb{B}\big[\hat{\mu}_\lambda^2\big] = \mu_\lambda^2\left(\frac{2+2(N-1)s}{T-1} + \frac{\kappa(N+2+2(N-1)s)}{T}\right). \qquad \text{(SM29)}
$$

Plugging (SM28)–(SM29) into (SM27) yields Equation (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 3

Under the assumptions of the proposition, we have from Kan and Lassance (2025, Proposition 7) that, for $T \geq N + 3$,

$$
\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}\big] = \frac{k_1(T-1)}{\gamma(T-N-2)}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\mu} = \frac{k_1(T-1)}{\gamma(T-N-2)}\boldsymbol{V}[(1-\delta)\boldsymbol{\Lambda} + \delta\mu_\lambda\boldsymbol{I}_N]^{-1}\boldsymbol{V}'\boldsymbol{\mu}, \qquad \text{(SM30)}
$$

where k_1 is defined in (8). Therefore, the expected out-of-sample mean return of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ is

$$
\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\mu}\big] = \frac{k_1(T-1)}{\gamma(T-N-2)}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2, \qquad \text{(SM31)}
$$

where the vector $\boldsymbol{f}(\delta)$ is defined in Proposition 3 and $\theta^2_{PC_i} = (\boldsymbol{v}_i'\boldsymbol{\mu})^2/\lambda_i$. Next, from Kan and Lassance (2025, Proposition 7), we have that the expected out-of-sample variance of $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ is

$$\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D}}\big] = \frac{1}{\gamma^2}\left(k_2\boldsymbol{\mu}'\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\mu} + \frac{k_3}{T}\mathrm{tr}\big(\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\Sigma}\big)\right), \tag{SM32}$$

where k_2 and k_3 are defined in (9)–(10) 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 $T \geq N+5$,

$$\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big] = \frac{(T-1)^2\big[(T-N-2)\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\mathrm{tr}\big(\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\big)\big]}{(T-N-1)(T-N-2)(T-N-4)}, \tag{SM33}$$

where

$$\mathrm{tr}\big(\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\big) = \sum_{i=1}^{N} f_i(\delta). \tag{SM34}$$

Therefore, the two terms composing $\mathbb{E}\big[\hat{\boldsymbol{w}}_{\boldsymbol{D}}'\boldsymbol{\Sigma}\hat{\boldsymbol{w}}_{\boldsymbol{D}}\big]$ in (SM32) become

$$\boldsymbol{\mu}'\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\mu} = \frac{(T-1)^2}{(T-N-1)(T-N-4)}\sum_{i=1}^{N}\theta^2_{PC_i}\left(f_i(\delta)^2 + \frac{f_i(\delta)\sum_{j=1}^{N} f_j(\delta)}{T-N-2}\right), \tag{SM35}$$

$$\mathrm{tr}\big(\mathbb{E}_{\mathcal{W}}\big[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^{-1}\big]\boldsymbol{\Sigma}\big) = \frac{(T-1)^2}{(T-N-1)(T-N-4)}\sum_{i=1}^{N}\left(f_i(\delta)^2 + \frac{f_i(\delta)\sum_{j=1}^{N} f_j(\delta)}{T-N-2}\right). \tag{SM36}$$

Plugging (SM35)–(SM36) into (SM32), and then using (SM31) (SM32), the EU delivered by $\hat{\boldsymbol{w}}_{\boldsymbol{D}}$ becomes

$$\begin{aligned}\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = {} & \frac{k_1(T-1)}{\gamma(T-N-2)}\sum_{i=1}^{N} f_i(\delta)\theta^2_{PC_i} \\ & - \frac{1}{2\gamma}\frac{(T-1)^2}{(T-N-1)(T-N-4)}\sum_{i=1}^{N}\left(f_i(\delta)^2 + \frac{f_i(\delta)\sum_{j=1}^{N} f_j(\delta)}{T-N-2}\right)\left(k_2\theta^2_{PC_i} + \frac{k_3}{T}\right),\end{aligned} \tag{SM37}$$

which, given the matrix $\boldsymbol{A}$ defined in Proposition 3, is equivalent to (17). This completes the proof.

Proof of Proposition 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 $\boldsymbol{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_N)$, find the optimal δ_i's, and infer the shrinkage intensities on the inverse

eigenvalues using $1 - 1/(1 - \delta_i)$. We can start from the sample eigenvalues instead of their inverse without loss of generality because each sample eigenvalue $\hat{\lambda}_i$ has its own shrinkage intensity δ_i.

Now, the MSE of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}} = \hat{\boldsymbol{V}}(\boldsymbol{I}_N - \boldsymbol{\Delta})\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}'$ is

$$N \times \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) = \mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^2)\big] + \text{tr}(\boldsymbol{\Sigma}^2) - 2\text{tr}\big(\mathbb{E}[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}]\boldsymbol{\Sigma}\big). \tag{SM38}$$

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),

$$\mathbb{E}\big[\text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}^2)\big] = \left(\frac{1}{T-1} + \frac{\kappa}{T}\right)\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}})^2 + \left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}}^2). \tag{SM39}$$

Letting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)'$, $\boldsymbol{\lambda}^2 = (\lambda_1^2, \ldots, \lambda_N^2)'$, and $\boldsymbol{\Lambda}^2 = \text{diag}(\boldsymbol{\lambda}^2)$, we have

$$\text{tr}\big(\mathbb{E}[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}]\boldsymbol{\Sigma}\big) = (\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}^2, \tag{SM40}$$

$$\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}})^2 = (\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}), \tag{SM41}$$

$$\text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{D}}^2) = (\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}). \tag{SM42}$$

Therefore, $\text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}})$ in (SM38) becomes

$$\begin{aligned}
&N \times \text{MSE}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}) \\
&= \left(\frac{1}{T-1} + \frac{\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}) + \left(\frac{T}{T-1} + \frac{2\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}) + \mathbf{1}_N'\boldsymbol{\lambda}^2 \\
&\quad - 2(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}^2 \\
&= \left(\frac{1}{T-1} + \frac{\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}) + \left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}) + \boldsymbol{\delta}'\boldsymbol{\Lambda}^2\boldsymbol{\delta}.
\end{aligned} \tag{SM43}$$

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,

$$\begin{aligned}
\boldsymbol{\delta}^\star = \underset{\boldsymbol{\delta} \in \mathbb{R}^N}{\text{argmin}} &\left(\frac{1}{T-1} + \frac{\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'(\mathbf{1}_N - \boldsymbol{\delta}) + \left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\mathbf{1}_N - \boldsymbol{\delta})'\boldsymbol{\Lambda}^2(\mathbf{1}_N - \boldsymbol{\delta}) + \boldsymbol{\delta}'\boldsymbol{\Lambda}^2\boldsymbol{\delta} \\
&\text{subject to } \boldsymbol{\delta}'\boldsymbol{\lambda} = 0.
\end{aligned} \tag{SM44}$$

We can show that the solution to Problem (SM44) is

$$\boldsymbol{\delta}^\star = \frac{\frac{1}{T-1} + \frac{2\kappa}{T}}{1 + \frac{1}{T-1} + \frac{2\kappa}{T}}\big(\mathbf{1}_N - \mu_\lambda \boldsymbol{\lambda}^{-1}\big). \tag{SM45}$$

Finally, the shrinkage intensity on the i-th inverse sample eigenvalue is

$$\delta^{\star}_{mse,i} = 1 - \frac{1}{1-\delta^{\star}_i} = \frac{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)(\mu_\lambda - \lambda_i)}{\left(\frac{1}{T-1} + \frac{2\kappa}{T}\right)\mu_\lambda + \lambda_i}. \tag{SM46}$$

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 $\lambda_i \to 0$, $\delta^{\star}_{mse,i} \geq 0$ if $\lambda_i \leq \mu_\lambda$, and $\delta^{\star}_{mse,i} \leq 0$ if $\lambda_i \geq \mu_\lambda$. This completes the proof.

Proof of Proposition 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 (22), with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$ in (19), is obtained in the same way as for the case of linear shrinkage in Equation (17).

Part 2. Given the expression of $\mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (22), the optimal vector of shrinkage intensities is

$$\boldsymbol{\delta}^{\star}_{eu} = \underset{\boldsymbol{\delta}\in\mathbb{R}}{\operatorname{argmax}}\, \mathrm{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \mathbf{1}_N - \frac{T-N-4}{T-1} k_1 \boldsymbol{A}^{-1}\boldsymbol{\theta}^2_{PC}. \tag{SM47}$$

To derive a closed-form expression for $\boldsymbol{A}^{-1}$, we use the following lemma.

Lemma 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 $\boldsymbol{B}$ whose entries are $B_{ij} = b_i(\mathbb{1}_{i=j} + c\mathbb{1}_{i\neq j})$. Then, the inverse of $\boldsymbol{B}$ exists provided $c \notin \{-\frac{1}{N-1}, 1\}$ and $b_i \neq 0$ for all i, and its entries are given by*

$$(\boldsymbol{B}^{-1})_{ij} = \begin{cases} \frac{1+(N-2)c}{(1-c)(1+(N-1)c)} b_i^{-1} & \text{if } i = j, \\ -\frac{c}{(1-c)(1+(N-1)c)} b_j^{-1} & \text{if } i \neq j. \end{cases} \tag{SM48}$$

Proof of Lemma 1. *We can rewrite $\boldsymbol{B}$ as*

$$\boldsymbol{B} = (1-c)\boldsymbol{D}_b + c\boldsymbol{b}\mathbf{1}'_N, \tag{SM49}$$

where $\boldsymbol{D}_b = \mathrm{diag}(\boldsymbol{b})$. Using the Sherman–Morrison formula, we have

$$\boldsymbol{B}^{-1} = \frac{1}{1-c}\left(\boldsymbol{D}_b^{-1} - \frac{c\boldsymbol{D}_b^{-1}\boldsymbol{b}\mathbf{1}'_N\boldsymbol{D}_b^{-1}}{1-c+c\mathbf{1}'_N\boldsymbol{D}_b^{-1}\boldsymbol{b}}\right), \tag{SM50}$$

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*

$$\boldsymbol{B}_{ii}^{-1} = \frac{b_i^{-1}}{1-c}\left(1 - \frac{c}{1+(N-1)c}\right) = \frac{1+(N-2)c}{(1-c)(1+(N-1)c)}b_i^{-1}, \tag{SM51}$$

$$\boldsymbol{B}_{ij}^{-1} = -\frac{c}{(1-c)(1+(N-1)c}b_j^{-1}, \tag{SM52}$$

which corresponds to (SM48)*. This completes the proof of Lemma 1.*

Using Lemma 1, the entries of the inverse of $\boldsymbol{A}$ defined in Proposition 3 are

$$(\boldsymbol{A}^{-1})_{ij} = \begin{cases} \frac{(T-3)(T-N-1)}{(T-2)(T-N-2)}(k_2\theta_{PC_i}^2 + k_3/T)^{-1} & \text{if } \ i=j, \\ -\frac{(T-N-1)}{(T-2)(T-N-2)}(k_2\theta_{PC_j}^2 + k_3/T)^{-1} & \text{if } \ i\neq j. \end{cases} \tag{SM53}$$

Plugging (SM53) into (SM47) yields the desired result in (23), where $R_i \in [0,1]$ in (21) because $\theta_{PC_i}^2 \geq 0$, $k_2 \geq k_1$, and $k_1, k_2, k_3 > 0$. We have $\delta_{eu,i}^\star \geq 0$ for all i, the lower bound 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

$$\delta_{eu,i}^\star = 1 + \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\sum_{j\neq i}^{N} R_j - \frac{(T-3)(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}R_i. \tag{SM54}$$

Therefore, an upper bound on $\delta_{eu,i}^\star$ is

$$\delta_{eu,i}^\star \leq 1 + \frac{(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\sum_{j\neq i}^{N} R_j \leq 1 + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}\bar{R}, \tag{SM55}$$

which is achieved when $R_i = 0$.

Part 3. Plugging $\boldsymbol{\delta}_{eu}^\star$ in (SM47) into (22) gives

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1^2(T-N-4)}{2\gamma(T-N-2)}(\boldsymbol{\theta}_{PC}^2)'\boldsymbol{A}^{-1}\boldsymbol{\theta}_{PC}^2. \tag{SM56}$$

Given the expression for $\boldsymbol{A}^{-1}$ in (SM53), we have

$$\begin{aligned}(\boldsymbol{\theta}_{PC}^2)'\boldsymbol{A}^{-1}\boldsymbol{\theta}_{PC}^2 &= \frac{T-N-1}{k_1(T-N-2)}\sum_{i=1}^{N}\left(R_i - \frac{N}{T-2}\bar{R}\right)\theta_{PC_i}^2 \\ &= \frac{T-1}{k_1(T-N-4)}\sum_{i=1}^{N}(1-\delta_{eu,i}^\star)\theta_{PC_i}^2, \end{aligned} \tag{SM57}$$

and plugging this into (SM56) yields (25). 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 (22) yields

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-1)(1-\delta)\theta^2}{\gamma(T-N-2)} - \frac{(T-1)^2(1-\delta)^2 \mathbf{1}_N' \boldsymbol{A} \mathbf{1}_N}{2\gamma(T-N-2)(T-N-4)}, \tag{SM58}$$

where $\theta^2 = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ is the maximum squared Sharpe ratio and $\mathbf{1}_N' \boldsymbol{A} \mathbf{1}_N = \frac{T-2}{T-N-1}(k_2\theta^2 + k_3 N/T)$. Therefore, the optimal $\delta^\star_{eu,cst}$ maximizing (SM58) is given by (26). Plugging $\delta^\star_{eu,cst}$ into (SM58) then yields

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{(T-N-1)(T-N-4)}{2\gamma(T-2)(T-N-2)} \left(\frac{k_1^2 \theta^4}{k_2\theta^2 + k_3 N/T} \right) \geq 0, \tag{SM59}$$

which is equivalent to the lower bound in (25) because $\theta^2 = \sum_{i=1}^N \theta^2_{PC_i}$.

Part 4. From (23), the condition $\delta^\star_{eu,i} < 1$ is equivalent to

$$\begin{aligned} \delta^\star_{eu,i} < 1 \;&\Leftrightarrow\; \frac{k_1 \theta^2_{PC_i}}{k_2 \theta^2_{PC_i} + k_3/T} > \frac{N}{T-2} \bar{R} \\ &\Leftrightarrow\; \theta^2_{PC_i} \left(k_1 - \frac{k_2 N}{T-2} \bar{R} \right) > \frac{k_3 \frac{N}{T}}{T-2} \bar{R}. \end{aligned} \tag{SM60}$$

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 (SM60) simplifies to $\theta^2_{PC_i} > \bar{\theta}^2_{PC}$ in (27). Otherwise, $\delta^\star_{eu,i} \geq 1$. This completes the proof.

Proof of Proposition SM.1

Under the i.i.d. multivariate elliptical distributional assumption, we have

$$\begin{aligned} \text{tr}(\mathbb{V}[\text{vec}(\hat{\boldsymbol{\Sigma}})]) &= \mathbb{E}[\text{tr}(\hat{\boldsymbol{\Sigma}}^2)] - \text{tr}(\boldsymbol{\Sigma}^2) \\ &= \left(\frac{1}{T-1} + \frac{\kappa}{T} \right) \text{tr}(\boldsymbol{\Sigma})^2 + \left(\frac{1}{T-1} + \frac{2\kappa}{T} \right) \text{tr}(\boldsymbol{\Sigma}^2), \end{aligned} \tag{SM61}$$

where the second equality follows from from Ollila and Raninen (2019, Lemma 2). Moreover, for the case of $\hat{\mu}_\lambda \boldsymbol{I}_N$, we have

$$\begin{aligned} \text{tr}(\mathbb{V}[\text{vec}(\hat{\mu}_\lambda \boldsymbol{I}_N)]) &= N \mathbb{V}[\hat{\mu}_\lambda] = \frac{1}{N} \mathbb{V}[\text{tr}(\hat{\boldsymbol{\Sigma}})] = \frac{1}{N} \left(\mathbb{E}[\text{tr}(\hat{\boldsymbol{\Sigma}})^2] - \text{tr}(\boldsymbol{\Sigma})^2 \right) \\ &= \frac{1}{N} \left(\frac{\kappa}{T} \text{tr}(\boldsymbol{\Sigma})^2 + 2 \left(\frac{1}{T-1} + \frac{\kappa}{T} \right) \text{tr}(\boldsymbol{\Sigma}^2) \right), \end{aligned} \tag{SM62}$$

where the last equality follows from Ollila and Raninen (2019, Lemma 2). Finally,

$$\begin{aligned}&\operatorname{tr}(\mathbb{V}[\operatorname{vec}(\hat{\boldsymbol{\Sigma}})]) - \operatorname{tr}(\mathbb{V}[\operatorname{vec}(\hat{\mu}_\lambda \boldsymbol{I}_N)]) \\ &= \frac{1}{N}\left[\left(\frac{N}{T-1} + \frac{(N-1)\kappa}{T}\right)\operatorname{tr}(\boldsymbol{\Sigma})^2 + \left(\frac{N-2}{T-1} + \frac{2(N-1)\kappa}{T}\right)\operatorname{tr}(\boldsymbol{\Sigma}^2)\right] \geq 0,\end{aligned} \tag{SM63}$$

and it is easy to check that the lower bound of zero is achieved if and only if $N = 1$ (in which case $\operatorname{tr}(\boldsymbol{\Sigma})^2 = \operatorname{tr}(\boldsymbol{\Sigma}^2)$) or $T \to \infty$.

Proof of Proposition SM.2

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 (SM5), is

$$\begin{aligned}\operatorname{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) &= \frac{1}{\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\mu} - \frac{1}{2\gamma}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\mu} + \frac{1}{T}\operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{D}}^{-1}\boldsymbol{\Sigma}\right)\right) \\ &= \frac{1}{\gamma}\sum_{i=1}^{N} f_i(\delta)\theta_{PC_i}^2 - \frac{1}{2\gamma}\sum_{i=1}^{N} f_i(\delta)^2\left(\theta_{PC_i}^2 + \frac{1}{T}\right),\end{aligned} \tag{SM64}$$

where $\boldsymbol{f}(\delta)$ is defined in Proposition 3, which corresponds to (SM6). Then, we establish when $\frac{\partial \widetilde{\operatorname{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})}{\partial \delta}\big|_{\delta=0} > 0$. We have

$$\frac{\partial \widetilde{\operatorname{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})}{\partial \delta} = \frac{1}{\gamma}\left[\sum_{i=1}^{N} f_i'(\delta)\theta_{PC_i}^2 - \sum_{i=1}^{N} f_i'(\delta)f_i(\delta)\left(\theta_{PC_i}^2 + \frac{1}{T}\right)\right], \tag{SM65}$$

where $f_i(0) = 1$ and $f_i'(0) = (\lambda_i - \mu_\lambda)/\lambda_i$. Therefore, at $\delta = 0$, we have

$$\frac{\partial \widetilde{\operatorname{EU}}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})}{\partial \delta}\bigg|_{\delta=0} = \frac{1}{\gamma T}\sum_{i=1}^{N}\frac{\mu_\lambda - \lambda_i}{\lambda_i} \geq 0, \tag{SM66}$$

where the inequality holds from the fact that the arithmetic mean of eigenvalues, μ_λ, is larger than the harmonic mean, $\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i^{-1}\right)^{-1}$. The lower bound of zero is attained if and only if $T \to \infty$ or all the λ_i's are equal. This completes the proof.

Proof of Proposition SM.3

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 (SM10), with $\boldsymbol{\Sigma}_{\boldsymbol{D}}$ in (SM9), is obtained in the same way as for the case of linear shrinkage in Equation (SM6).

Part 2. Given the expression for $\operatorname{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (SM10), it is easy to show that the optimal intensities are given by (SM11), and $\tilde{\delta}_{eu,i}^{\star} \in [0, 1]$ because $\theta_{PC_i}^2 \geq 0$. To show that $\tilde{\delta}_{eu,i}^{\star}$ in (SM11) is lower than

its counterpart under the sample covariance matrix, $\delta^{\star}_{eu,i}$ in (23), we must show that

$$1-\frac{\theta^2_{PC_i}}{\theta^2_{PC_i}+1/T}\leq 1-\frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i-\frac{N}{T-2}\bar{R}\right), \tag{SM67}$$

which holds because $\frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\leq 1$ and $\frac{\theta^2_{PC_i}}{\theta^2_{PC_i}+1/T}\geq R_i=\frac{k_1\theta^2_{PC_i}}{k_2\theta^2_{PC_i}+k_3/T}$ given that $k_1,k_2,k_3\geq 1$ and $k_2\geq k_1$.

Part 3. Plugging the $\tilde{\delta}^{\star}_{eu,i}$ in (SM11) into (SM10), we directly obtain that the EU delivered by the $\tilde{\delta}^{\star}_{eu,i}$ is given by (SM12). This completes the proof.

Proof of Proposition SM.4

Following the proof of Proposition 4, the solution is obtained by finding the vector $\boldsymbol{\delta}^{\star}$ minimizing MSE($\hat{\boldsymbol{\Sigma}}_{\boldsymbol{D}}$) in (SM43), and then, taking $1-1/(1-\delta^{\star}_i)$. We find that the $\delta^{\star}_i$'s minimizing (SM43) are given by

$$\delta^{\star}_i=\frac{\left(\frac{N}{T-1}+\frac{\kappa N}{T}\right)\mu_{\lambda}+\left(\frac{1}{T-1}+\frac{2\kappa}{T}\right)\left(\frac{T}{T-1}+\frac{2\kappa}{T}\right)\left(1+\frac{N}{T-1}+\frac{\kappa N}{T}\right)\lambda_i}{\left(\frac{T}{T-1}+\frac{2\kappa}{T}\right)^2\left(1+\frac{N}{T-1}+\frac{\kappa N}{T}\right)\lambda_i}, \tag{SM68}$$

which yields $\delta^{\star}_{mse,unc,i}=1-1/(1-\delta^{\star}_i)$ in (SM13). Using the formula for the constrained intensities $\delta^{\star}_{mse,i}$ in (20) and the unconstrained intensities $\delta^{\star}_{mse,unc,i}$ in (SM13), it is easy to check that $\lim_{\lambda_i\to 0}\delta^{\star}_{mse,i}=\lim_{\lambda_i\to 0}\delta^{\star}_{mse,unc,i}=1$ and $\lim_{\lambda_i\to\infty}\delta^{\star}_{mse,i}=\lim_{\lambda_i\to\infty}\delta^{\star}_{mse,unc,i}=-\left(\frac{1}{T-1}+\frac{2\kappa}{T}\right)$. We can also verify that the denominator of (SM13) is equal to zero when $\lambda_i/\mu_{\lambda}=\zeta$ defined in (SM16). It follows that $\delta^{\star}_{mse,unc,i}\to\pm\infty$ as $\lambda_i/\mu_{\lambda}\to\zeta$ from below (above). This completes the proof.

Proof of Proposition SM.5

Part 1. The difference between $\tilde{\delta}^{\star}_{eu,i}$ in (SM18) and $\delta^{\star}_{eu,i}$ in (23) is

$$\begin{aligned}
&\tilde{\delta}^{\star}_{eu,i}-\delta^{\star}_{eu,i}\\
&=1-\frac{(T-N-1)(T-N-4)}{(T-1)(T-2)}R_i-\left(1-\frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i-\frac{N}{T-2}\bar{R}\right)\right)\\
&=\frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}(R_i-\bar{R}),
\end{aligned} \tag{SM69}$$

which corresponds to (SM19). Therefore, it is clear that $\tilde{\delta}^{\star}_{eu,i}\geq\delta^{\star}_{eu,i}$ if $R_i\geq\bar{R}$, and $\tilde{\delta}^{\star}_{eu,i}\leq\delta^{\star}_{eu,i}$ otherwise.

Part 2. Plugging $\delta_i = \tilde{\delta}^{\star}_{eu,i}$ into (22), the EU delivered by the $\tilde{\boldsymbol{\delta}}^{\star}_{eu}$ is

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-N-1)(T-N-4)}{\gamma(T-2)(T-N-2)} \sum_{i=1}^{N} R_i \theta^2_{PC_i} - \frac{(T-N-1)^2(T-N-4)}{2\gamma(T-2)^2(T-N-2)} \boldsymbol{R}' \boldsymbol{A} \boldsymbol{R}, \quad \text{(SM70)}$$

where $\boldsymbol{R} = (R_1, \ldots, R_N)'$. Given the matrix $\boldsymbol{A}$ defined in Proposition 3, we have

$$\boldsymbol{R}' \boldsymbol{A} \boldsymbol{R} = \frac{k_1(T-2)}{T-N-1} \left(\left(1 - \frac{N}{T-2}\right) \sum_{i=1}^{N} R_i \theta^2_{PC_i} + \frac{N}{T-2} \bar{R} \theta^2 \right), \quad \text{(SM71)}$$

and thus, $\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}})$ in (SM70) simplifies to

$$\text{EU}(\hat{\boldsymbol{w}}_{\boldsymbol{D}}) = \frac{k_1(T-N-1)(T-N-4)(T+N-2)}{2\gamma(T-2)^2(T-N-2)} \sum_{i=1}^{N} R_i \left(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2} \right), \quad \text{(SM72)}$$

which corresponds to (SM20). 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 (25), we have that

$$\sum_{i=1}^{N} (1 - \delta^{\star}_{eu,i}) \theta^2_{PC_i} \geq 0 \;\Leftrightarrow\; \sum_{i=1}^{N} R_i \left(\theta^2_{PC_i} - \frac{\theta^2}{T-2} \right) \geq 0 \;\Leftrightarrow\; \sum_{i=1}^{N} R_i \left(\theta^2_{PC_i} - \frac{\theta^2}{T+N-2} \right) \geq 0. \quad \text{(SM73)}$$

This completes the proof.

Proof of Proposition SM.6

Let $\boldsymbol{\kappa} = \mathbf{1}_N - \boldsymbol{\delta}$. Then, Problem (SM21) is equivalent to

$$\begin{aligned} \boldsymbol{\kappa}^{+}_{eu} = \underset{\boldsymbol{\kappa} \in \mathbb{R}^N}{\text{argmax}} \; & \frac{k_1(T-1)}{\gamma(T-N-2)} \boldsymbol{\kappa}' \boldsymbol{\theta}^2_{PC} - \frac{(T-1)^2}{2\gamma(T-N-2)(T-N-4)} \boldsymbol{\kappa}' \boldsymbol{A} \boldsymbol{\kappa} \\ & \text{subject to} \;\; \kappa_i \geq 0 \; \forall i, \end{aligned} \quad \text{(SM74)}$$

whose solution is

$$\boldsymbol{\kappa}^{+}_{eu} = \frac{T-N-4}{T-1} k_1 \boldsymbol{A}^{-1} (\boldsymbol{\theta}^2_{PC} + \boldsymbol{\eta}) \;\Leftrightarrow\; \boldsymbol{\delta}^{+}_{eu} = \mathbf{1}_N - \frac{T-N-4}{T-1} k_1 \boldsymbol{A}^{-1} (\boldsymbol{\theta}^2_{PC} + \boldsymbol{\eta}), \quad \text{(SM75)}$$

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 (SM53), $\boldsymbol{\delta}^{+}_{eu}$ takes a form similar to (23) with R_i replaced by R^{+}_i in (SM22), which gives the solution in (SM23). 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 $\delta^{\star}_{eu,i}$

in (23) as

$$\begin{aligned}\delta_{eu,i}^{+} &= 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i^{+} - \frac{N}{T-2}\bar{R}^{+}\right) \\ &= 1 - \frac{(T-N-1)(T-N-4)}{(T-1)(T-N-2)}\left(R_i - \frac{N}{T-2}\bar{R}^{+}\right) \\ &= \delta_{eu,i}^{\star} + \frac{N(T-N-1)(T-N-4)}{(T-1)(T-2)(T-N-2)}(\bar{R}^{+} - \bar{R}),\end{aligned} \tag{SM76}$$

which corresponds to (SM24), where $\bar{R}^{+} \geq \bar{R}$ because $R_j^{+} \geq R_j$ for all j. This completes the proof.

References in Supplementary Material

Haff, L. R. (1979), "An identity for the Wishart distribution with applications." *Journal of Multivariate Analysis*, 9(4):531–544.

Jorion, P. (1986), "Bayes-Stein estimation for portfolio analysis." *Journal of Financial and Quantitative Analysis*, 21(3):279–292.

Kan, R. and Lassance, N. (2025), "Optimal portfolio choice with fat tails and parameter uncertainty." *Journal of Financial and Quantitative Analysis*, forthcoming.

Ledoit, O. and Wolf, M. (2004), "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, 88(2):365–411.

Ollila, E. and Raninen, E. (2019), "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions." *IEEE Transactions on Signal Processing*, 67(10):2707–2719.