

OPTIMAL PORTFOLIO SIZE UNDER PARAMETER UNCERTAINTY

Nathan Lassance, Rodolphe Vanderveken,
Frédéric Vrins

REPRINT • 2025 / 11

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>

Optimal Portfolio Size under Parameter Uncertainty*

Rodolphe Vanderveken¹ Nathan Lassance¹ Frédéric Vrins^{1,2}

Forthcoming at the
Journal of Financial and Quantitative Analysis

October 29, 2025

Abstract

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

Keywords: dimension reduction, estimation risk, out-of-sample performance, portfolio combination rules, portfolio selection.

JEL Classification: G11.

*R. Vanderveken is corresponding author. Affiliations: (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 Jean-François Boulier, Richard B. Evans, Mohamad Hassan Abou Daya, Aleksy Leeuwenkamp, Marco Saerens, Majeed Simaan, Guofu Zhou, and participants at the 12th General AMaMeF Conference, 17th Annual Meeting of the Society for Financial Econometrics (SoFiE), 6th Quantitative Finance and Financial Econometrics International Conference (QFFE), AF2I Academic Workshop, 2025 IFABS Oxford Conference, 17th International Risk Management Conference (IRMC), 40th International Conference of the French Finance Association (AFFI), 7th Quantitative Finance and Risk Analysis Symposium (QFRA), 18th Belgian Financial Research Forum (BFRF), 5th International Conference on Computational Finance (ICCF), and LFIN internal seminar for their comments. This work was supported by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under Grant Number J.0135.25 and T.0221.22.

1 Introduction

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

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

Sparse portfolio selection aims at alleviating parameter uncertainty precisely by reducing the portfolio size. This is typically achieved by imposing lasso-type constraints also known as soft-thresholding (DeMiguel et al., 2009a; Fan et al., 2012; Yen, 2016; Ao et al., 2019) or cardinality constraints (Gao and Li, 2013; Du et al., 2023). However, these methods suffer from five main

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

drawbacks. First, the optimization programs must be solved in dimension M , hence, face large estimation risk, even if the output portfolio is ultimately of lower dimension. Second, they can be computationally intensive in high dimension. For instance, cardinality constraints render the portfolio problem NP-hard. Third, they entangle the selection of the assets with the optimization of portfolio weights in one single step. Therefore, sparse methods do not allow for different rebalancing frequencies for portfolio weights and the asset selection, or for flexibility in the asset selection. Fourth, they generally feature hyperparameters that are often estimated by cross-validation, which is time consuming and adds an extra layer of estimation risk. Lastly, the resulting portfolio size N is implicit from the portfolio weights and does not have a clear meaning.

To address these drawbacks of sparse portfolio selection, we propose a sequential *three-stage process*. First, select a portfolio strategy. In this paper, we consider a class of portfolio strategies that consists of different combinations of sample mean-variance (MV), global-minimum-variance (GMV), and equally weighted (EW) portfolios. Although more sophisticated strategies exist, this class has the benefit of being simple, theoretically important, commonly considered in the literature, and allows analytical tractability in finite samples. Second, compute an *optimal* portfolio size $N \leq M$. Third, select which N assets to invest in among the M available ones and optimize the weights on the N assets with the strategy in step one. Interestingly, Ao et al. (2019) propose a similar sequential process for reasons of computational efficiency, selecting a chosen number of N assets to maximize the Sharpe ratio and then implementing their MAXSER strategy on this subset. Specifically, among the S&P500 stocks, they *arbitrarily* select 50 of them. In contrast, we show how to select an *optimal* N for several portfolio rules based on estimation-risk considerations.

Two key questions remain in our three-stage process: How do we find the optimal portfolio size N ? How do we then select the N assets? We focus on the first question. Specifically, we

introduce a methodology for finding the optimal N and leave flexibility to the investor as to the choice of the N assets. In our empirical analysis, we evaluate the performance of our optimal N on 10 simple and sensible asset selection rules as an illustration.

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

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

valuable in practice.² Fourth, our approach does not require the calibration of any hyperparameter such as a constraint threshold. Lastly, the optimal N has an intuitive meaning from trading off between investment opportunities and estimation risk when maximizing expected utility.

Our theoretical analysis starts with the sample MV (SMV) portfolio, which is optimal *in sample*, but not out of sample. Assuming that asset returns are equicorrelated as advocated by Engle and Kelly (2012) and Clements et al. (2015),³ we express the EU of this portfolio as a function of the portfolio size, the sample size, the correlation, the assets' Sharpe ratios, and three parameters that measure the impact of fat tails.⁴ We show that because SMV is highly sensitive to estimation errors, its optimal N is typically very small. As a remedy for this large estimation risk, Kan and Zhou (2007) introduce a *two-fund rule* that scales down the SMV portfolio by combining it with the risk-free asset to maximize the EU, which Kan and Lassance (2025) extend to the case with i.i.d. multivariate elliptical returns. Building on this, we derive the optimal N for the two-fund rule. Remarkably, for typical levels of rather high correlations in equity data, this optimal N is slightly below *half the sample size*. Given commonly used sample sizes (e.g., 120 months), this result means that, under the two-fund rule, a finite-sample setting, and our model assumptions, it is optimal to limit the portfolio size so as to optimize the out-of-sample performance.

In addition to the two-fund rule, we derive the optimal N for *three-fund rules* combining the

²In our empirical analysis, we rebalance the weights every month but the selected assets every year, which reduces asset turnover and outperforms selecting assets every month. Nonetheless, we show in Section OA.7.5 of the Online Appendix that our optimal N outperforms investing in all M assets also when the selected assets change every month. This disentanglement is not feasible under lasso or cardinality constraints because asset selection and weights optimization is simultaneous.

³Equicorrelation yields a satisfying tradeoff between specification and estimation error (Engle and Kelly, 2012) and allows us to express the EU of our different MV portfolio rules as an explicit function of N . We use this assumption only for finding the optimal N , not the portfolio weights. In Section OA.1 of the Online Appendix, we show that the optimal N found under equicorrelation is similar to that under a single-factor model. Moreover, we find that our method performs well also when returns are not equicorrelated.

⁴Most of the literature on portfolio choice with estimation risk assumes i.i.d. multivariate normally distributed returns. Our incorporation of elliptical fat tails builds on the recent work by Kan and Lassance (2025).

SMV portfolio, the sample GMV (SGMV) portfolio, and the risk-free asset as in Kan and Zhou (2007), DeMiguel et al. (2015), and Yuan and Zhou (2024), or combining the SMV portfolio, the equally weighted (EW) portfolio, and the risk-free asset as in Tu and Zhou (2011), Kan and Wang (2023), and Lassance et al. (2024b). Similar to the two-fund rule, the optimal N for these rules is slightly below half the sample size for typical equity-return correlations. These results extend the literature on portfolio choice with estimation risk by showing that it is beneficial to substantially reduce the portfolio size even after optimally combining the SMV portfolio with robust strategies.

We test our theory first in simulations where we assess the performance of our optimal N when (i) it is subject to estimation errors, (ii) asset returns are not equicorrelated, and (iii) the selection of the N assets out of the M available ones is random. The main conclusion is that the size-optimized two-fund and three-fund rules still deliver an EU close to the maximum and substantially larger than that when investing in all M assets or in too few assets. This shows that our method delivers a satisfying performance even when relaxing the theoretical assumptions, highlighting its practical relevance. Moreover, the simulation analysis allows us to illustrate that, for realistic values of the sample size, our size-optimized two-fund and three-fund rules can outperform more sophisticated benchmark MV portfolios that reduce the portfolio size using soft or hard-thresholding, as well as a factor-plus-alpha strategy built on PCA and the arbitrage portfolio of Da et al. (2024).

Finally, we turn to empirical data. We consider six datasets of characteristic and industry-sorted portfolios and one dataset of individual stocks, with M around 100. Using a rolling window of 120 months, we compare the net out-of-sample utility of 11 portfolio strategies, including our size-optimized two-fund and three-fund rules. For the latter, we propose 10 simple and sensible *selection rules* to decide which N assets to select. We find large benefits from optimizing N using our theory. In the vast majority of cases, the two-fund and three-fund rules implemented with the

10 asset selection rules outperform the same rules applied to all M assets. The improvement is particularly large and statistically significant when we shrink the covariance matrix. Moreover, for some asset selection rules, our size-optimized portfolio rules consistently outperform the EW and SGMV portfolios, a notoriously difficult task when the dataset dimension M is comparable to the sample size, as well as soft and hard-thresholding MV portfolios and the factor-plus-alpha strategy.

Our paper complements existing work that studies the effect of the portfolio size N using asymptotic theories. In particular, Ao et al. (2019) and Da et al. (2024) consider a class of more sophisticated portfolio strategies that treats estimation risk differently from ours via soft-thresholding and a factor model, respectively. They find, under specific assumptions, that they can asymptotically achieve the performance of the population optimal portfolio as both the portfolio size and the sample size go to infinity. That is, in their asymptotic theories, it is optimal to increase N . We complement these results by studying a simpler, but nonetheless important, class of portfolio strategies for which we can study the out-of-sample performance analytically in a finite-sample setting. In that case, we find that it is optimal to reduce the portfolio size N . Our simulation and empirical results signal that asymptotic theories like those of Ao et al. (2019) and Da et al. (2024), which suggest that increasing N is beneficial, might not kick in for realistic sample sizes and typically deliver a performance inferior to that of our size-optimized portfolios.

The paper is structured as follows. In Section 2, we study the optimal portfolio size for the MV portfolio with no parameter uncertainty. In Section 3, we show how to derive an optimal N for the SMV portfolio and the two-fund rule under parameter uncertainty. Sections 4 and 5 contain our simulation and empirical analysis, respectively. Section 6 concludes. The Appendix contains results and proofs that are central to the main text. The Online Appendix provides additional theoretical, simulation, and empirical results.

2 Optimal portfolio size without parameter uncertainty

In this section, we study the optimal portfolio size for an unconstrained MV investor who knows the parameters of asset returns without uncertainty. We suppose the investor starts from a universe of M assets and wishes to find an optimal number $N \leq M$ of assets.

Given a fixed number N of assets, let $\mathbf{r}_t$ be the $N \times 1$ vector of asset excess returns at time t , which has a mean $\boldsymbol{\mu}_N$ and a positive-definite covariance matrix $\boldsymbol{\Sigma}_N$. We denote by μ_i , σ_i^2 , and $s_i = \mu_i/\sigma_i$ the mean excess return, variance, and Sharpe ratio of asset i . An MV investor with risk-aversion coefficient $\gamma > 0$ selects the portfolio weights on the risky assets as

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

where $U(\mathbf{w})$ is the MV utility of portfolio $\mathbf{w}$. Solving (1) yields the MV portfolio satisfying

$$\mathbf{w}^* = \frac{1}{\gamma} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N \quad \text{and} \quad U(\mathbf{w}^*) = \frac{\theta_N^2}{2\gamma} \quad \text{with} \quad \theta_N^2 = \boldsymbol{\mu}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N, \quad (2)$$

where θ_N is the maximum Sharpe ratio. We have that $\mathbf{w}^*$ combines two funds: the fully invested tangency portfolio, $\mathbf{w} = (\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N)^{-1} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N$, and the risk-free asset, $\mathbf{w} = \mathbf{0}_N$. Clearly, $U(\mathbf{w}^*)$ is non-decreasing with N because one can set the weight of an additional asset to zero and keep the same utility. Thus, without parameter uncertainty, it is optimal to consider the whole investment universe and set $N = M$. To relate $U(\mathbf{w}^*)$ and N explicitly, we assume the following about $\boldsymbol{\Sigma}_M$.

Assumption 1 *The covariance matrix $\boldsymbol{\Sigma}_M$ takes the form $\boldsymbol{\Sigma}_M(\rho) = \mathbf{D}_M \mathbf{P}_M(\rho) \mathbf{D}_M$ with $\mathbf{D}_M$ the diagonal matrix of standard deviations and $\mathbf{P}_M(\rho)$ the equicorrelation matrix,*

$$\mathbf{P}_M(\rho) = (1 - \rho) \mathbf{I}_M + \rho \mathbf{1}_M \mathbf{1}_M', \quad (3)$$

where $\mathbf{1}_M$ and $\mathbf{I}_M$ are the unit vector and matrix, and $\rho \in \left(-\frac{1}{M-1}, 1\right)$ so that $\boldsymbol{\Sigma}_M$ is invertible.

Under Assumption 1, the dependence structure depends on a single parameter, ρ . It follows that for any $N \leq M$, Σ_N has the same form as in Assumption 1. This assumption is commonly used in portfolio selection to trade off between specification and estimation error.⁵ We use Assumption 1 *only* as a way to express the EU of our portfolio rules as an explicit function of N , which will allow us to find the optimal N . We do *not* need this assumption for the portfolio weights. In Section OA.1 of the Online Appendix, we find the optimal N under another commonly used approximation, the *single-factor model*, which we show is essentially equivalent to that under equicorrelation.

In the next proposition, we provide the analytical expression of $U(\mathbf{w}^*)$ under Assumption 1.⁶ This is a novel result to the best of our knowledge.

Proposition 1 *Under Assumption 1, the utility of the MV portfolio is $U(\mathbf{w}^*) = \theta_N^2/(2\gamma)$ with*

$$\theta_N^2 = \frac{N}{1-\rho} \delta_N(\bar{\boldsymbol{\theta}}_N), \quad \delta_N(\bar{\boldsymbol{\theta}}_N) = \bar{\theta}_{N,2} - \frac{N\rho}{1-\rho+N\rho} \bar{\theta}_{N,1}^2 \geq 0, \quad (4)$$

where $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ and $\bar{\theta}_{N,k} = \frac{1}{N} \sum_{i=1}^N s_i^k$. Moreover, $U(\mathbf{w}^*)$ is non-decreasing in N and is strictly increasing from N to $N+1$ whenever $s_{N+1} \neq \rho N \bar{\theta}_{N,1}/(1-\rho+N\rho)$.

Proposition 1 shows that the maximum utility is non-decreasing in N . In Figure 1, we depict $U(\mathbf{w}^*)$ as a function of $N \in \{1, \dots, M\}$ for $\rho \in \{0.2, 0.5, 0.8\}$ and assets' monthly Sharpe ratios s_i calibrated to a dataset of $M = 96$ decile portfolios⁷ sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.⁸

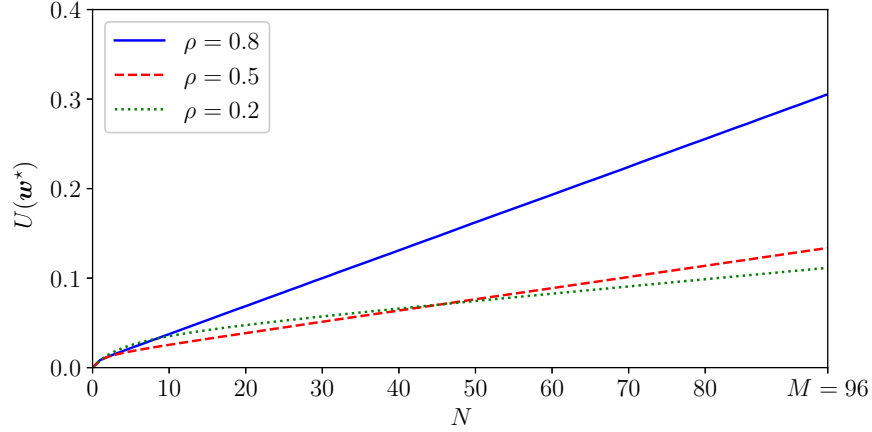
⁵Elton and Gruber (1973) find that assuming equicorrelation improves portfolio performance relative to a wide range of alternatives. Ledoit and Wolf (2003) shrink the sample covariance matrix toward an equicorrelation model. Boyle et al. (2012) study an ambiguity-averse portfolio and assume the parameters, including correlations, are homogeneous. Engle and Kelly (2012) propose a dynamic equicorrelation covariance matrix and find that it “[...] improves portfolio selection compared to an unrestricted dynamic correlation structure.” Clements et al. (2015) “[...] find evidence in favour of assuming equicorrelation across various portfolio sizes, particularly during times of crisis.” Finally, Chung et al. (2022) find that errors in correlations dominate those in means and variances for MV optimization.

⁶The proofs of all results in the main text are available in Section A.3 of the Appendix.

⁷We remove four portfolios for which there is missing data. We collect this dataset from Kenneth French's website.

⁸Figure 1 shows that $U(\mathbf{w}^*)$ is a non-monotonic function of ρ . In Section OA.2 of the Online Appendix, we demonstrate that $U(\mathbf{w}^*)$ is a convex function of ρ , extending results in Gandy and Veraart (2013).

Figure 1: Utility of the mean-variance portfolio as a function of N and ρ



Notes. This figure depicts $U(\mathbf{w}^*)$ in Equation (4) as a function of the portfolio size N under the assumption that asset returns are equicorrelated with a correlation $\rho \in \{0.2, 0.5, 0.8\}$. We calibrate the assets' monthly Sharpe ratios to a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023. Starting with $N = 1$ asset chosen randomly, we compute $U(\mathbf{w}^*)$. Then, we add a randomly selected asset not previously selected, and compute $U(\mathbf{w}^*)$ again. We continue this procedure until $N = M$. We repeat this procedure 10,000 times and depict the average $U(\mathbf{w}^*)$ over all draws. We consider a risk-aversion coefficient $\gamma = 1$.

3 Optimal portfolio size under parameter uncertainty

In this section, we show that it is no longer optimal for an MV investor to invest in all M assets when the assets' parameters are unknown and the investor relies on sample estimates.

3.1 Sample estimates and distributional assumption

Given a sample of asset excess returns of size T , $(\mathbf{r}_1, \dots, \mathbf{r}_T)$, let

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

be the sample estimates of $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$, and we require $T > N$ so that $\hat{\boldsymbol{\Sigma}}_N$ is almost surely invertible. Given a *sample portfolio* $\hat{\mathbf{w}}$ estimated with $\hat{\boldsymbol{\mu}}_N$ and $\hat{\boldsymbol{\Sigma}}_N$, we measure its performance with the *expected out-of-sample utility (EU)* pioneered by Kan and Zhou (2007),

$$EU(\hat{\mathbf{w}}) = \mathbb{E}[U(\hat{\mathbf{w}})] = \mathbb{E} \left[\hat{\mathbf{w}}' \boldsymbol{\mu}_N - \frac{\gamma}{2} \hat{\mathbf{w}}' \boldsymbol{\Sigma}_N \hat{\mathbf{w}} \right]. \quad (6)$$

We then define N^* as the optimal portfolio size N maximizing the EU of $\hat{w}$.

Kan and Zhou (2007) and follow-up papers like Tu and Zhou (2011), Kan et al. (2021), Lassance et al. (2024b,a), and Yuan and Zhou (2024), evaluate (6) under the i.i.d. multivariate normal distributional assumption. Instead, we follow Kan and Lassance (2025) who study the EU of various sample portfolios under the i.i.d. multivariate *elliptical* distribution. Like them, we use the stochastic representation of the elliptical distribution in El Karoui (2010, 2013).

Assumption 2 *Asset returns $\mathbf{r}_t$ are independent and identically distributed over time and follow a multivariate elliptical distribution. That is, $\mathbf{r}_{t_1} \perp \mathbf{r}_{t_2}$ for $t_1 \neq t_2$ and $\mathbf{r}_t \stackrel{d}{=} \boldsymbol{\mu}_M + (\tau \boldsymbol{\Sigma}_M)^{\frac{1}{2}} \mathbf{z}_M$, where $\mathbf{z}_M \sim \mathcal{N}(\mathbf{0}_M, \mathbf{I}_M)$, τ is a positive random variable satisfying $\mathbb{E}[\tau] = 1$, and $\mathbf{z}_M \perp \tau$. In particular, $\boldsymbol{\mu}_M$ and $\boldsymbol{\Sigma}_M$ are constant over time.*

We recover the normal distribution when $\tau = 1$, and the t -distribution when $\tau \sim (\nu - 2)/\chi_\nu^2$, where χ_ν^2 is a chi-square distribution with $\nu > 2$ degrees of freedom. We focus on the elliptical distribution because it is consistent with MV portfolios (Chamberlain, 1983; Schuhmacher et al., 2021). Specifically, since the multivariate elliptical distribution is closed under linear transformations, all portfolios have the same higher moments and only differ in their mean return and variance. Therefore, for all increasing and concave utility functions, the optimal portfolio under expected utility is MV efficient. We also opt for the elliptical distribution because Kan and Lassance (2025) show that accounting for fat tails is crucial in determining optimal portfolio combination rules.

3.2 Optimal portfolio size for sample portfolios

The first strategy we consider is the sample counterpart of the MV portfolio in (2), i.e.,

$$\hat{\mathbf{w}}^* = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N. \quad (7)$$

The sample MV (SMV) portfolio is not robust to parameter uncertainty because it is only optimal *in sample*. Therefore, we also consider the *two-fund rule* of Kan and Zhou (2007) that scales down the SMV portfolio to maximize its EU. This two-fund rule is of the form

$$\hat{\mathbf{w}}(\alpha) = \alpha \hat{\mathbf{w}}^* = \frac{\alpha}{\gamma} \hat{\Sigma}_N^{-1} \hat{\boldsymbol{\mu}}_N, \quad (8)$$

where $\alpha \in \mathbb{R}$ is the *combination coefficient*. The SMV in (7) corresponds to $\hat{\mathbf{w}}^* = \hat{\mathbf{w}}(1)$.

In the next proposition, we derive the EU of the two-fund rule, and thus also the SMV portfolio, the optimal combination coefficient α^* , and which EU it delivers, when asset returns are elliptically distributed. These results then allow us to determine the optimal portfolio size N . This Proposition follows, with minor adjustments, from Kan and Lassance (2025, Propositions 7 and 8). In Section A.1 of the Appendix, we follow the same approach for *three-fund rules* that extend the two-fund rule by incorporating a third fully invested fund, which is either the GMV or the EW portfolio. We refer to these as the *GMV-three-fund rule* and the *EW-three-fund rule*, respectively.

Proposition 2 *Let $T > N + 4$, $\mathbf{M} = \mathbf{I}_T - \mathbf{1}_T \mathbf{1}_T' / T$, $\mathbf{Z}_N$ be a $T \times N$ matrix of independent standard normal variables, $\boldsymbol{\Lambda}$ be a diagonal matrix of T independent copies of $\tau^{1/2}$, and $\boldsymbol{\Lambda} \perp \mathbf{Z}_N$. Then, under Assumption 2, the EU of the two-fund rule $\hat{\mathbf{w}}(\alpha)$ in (8) is*

$$EU(\hat{\mathbf{w}}(\alpha)) = \frac{T}{2\gamma(T - N - 2)} \left[\left(2\alpha\kappa_{N,1} - \frac{c_N\alpha^2\kappa_{N,2}T}{T - N - 2} \right) \theta_N^2 - \frac{c_N\alpha^2\kappa_{N,3}N}{T - N - 2} \right], \quad (9)$$

where θ_N^2 is the maximum squared Sharpe ratio given in (4), $c_N = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}$, and $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$, which we assume exist, are functions of only N , T , and the distribution of τ :

$$\kappa_{N,1} = \frac{T - N - 2}{N} \mathbb{E} \left[\text{tr}((\mathbf{Z}_N' \boldsymbol{\Lambda} \mathbf{M} \boldsymbol{\Lambda} \mathbf{Z}_N)^{-1}) \right], \quad (10)$$

$$\kappa_{N,2} = \frac{(T - N - 2)^2}{c_N N} \mathbb{E} \left[\text{tr}((\mathbf{Z}_N' \boldsymbol{\Lambda} \mathbf{M} \boldsymbol{\Lambda} \mathbf{Z}_N)^{-2}) \right], \quad (11)$$

$$\kappa_{N,3} = \frac{(T - N - 2)^2}{c_N NT} \mathbb{E} \left[\mathbf{1}'_T \mathbf{\Lambda} \mathbf{Z}_N (\mathbf{Z}'_N \mathbf{\Lambda} \mathbf{M} \mathbf{\Lambda} \mathbf{Z}_N)^{-2} \mathbf{Z}'_N \mathbf{\Lambda} \mathbf{1}_T \right]. \quad (12)$$

Moreover, the optimal combination coefficient α^* maximizing (9), and the resulting EU, are

$$\alpha^* = \frac{T - N - 2}{c_N T} \left(\frac{\kappa_{N,1} \theta_N^2}{\kappa_{N,2} \theta_N^2 + \kappa_{N,3} \frac{N}{T}} \right) \quad \text{and} \quad EU(\hat{\mathbf{w}}(\alpha^*)) = \frac{1}{2\gamma c_N} \left(\frac{\kappa_{N,1}^2 \theta_N^4}{\kappa_{N,2} \theta_N^2 + \kappa_{N,3} \frac{N}{T}} \right). \quad (13)$$

Three comments are in order. First, Proposition 2 is valid for any Σ_N , not only those of the form $\Sigma_N(\rho)$. Second, $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$ do not have a closed-form expression, but can be evaluated via Monte Carlo simulations given the distribution of τ . We explain how we estimate them in Section A.2 of the Appendix. Kan and Lassance (2025) show that they increase with estimation risk N/T and tail heaviness. Third, under parameter uncertainty in μ_N and Σ_N , a larger N may not always be favorable. Instead, the N maximizing (9) trades off between improving the population performance (i.e., increasing θ_N^2) and limiting estimation risk (i.e., decreasing N/T).

Next, we proceed as in Section 2 and relate the EU with N more explicitly by assuming that the covariance matrix Σ_N complies with Assumption 1, i.e., $\Sigma_N = \Sigma_N(\rho)$. We obtain this novel result by plugging the expression for θ_N^2 under equicorrelation in (4) into Equations (9) and (13).

Corollary 1 *Under Assumptions 1 and 2, the EU of the SMV portfolio $\hat{\mathbf{w}}^*$ and the optimal two-fund rule $\hat{\mathbf{w}}(\alpha^*)$ are given by*

$$EU(\hat{\mathbf{w}}^*) = \frac{NT}{2\gamma(T - N - 2)} \left[\frac{\delta_N(\bar{\boldsymbol{\theta}}_N)}{1 - \rho} \left(2\kappa_{N,1} - \frac{c_N \kappa_{N,2} T}{T - N - 2} \right) - \frac{c_N \kappa_{N,3}}{T - N - 2} \right] \quad (14)$$

$$=: EU_{smv}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N), \quad \text{and} \quad (15)$$

$$EU(\hat{\mathbf{w}}(\alpha^*)) = \frac{N}{2\gamma c_N(1 - \rho)} \times \frac{\kappa_{N,1}^2 \delta_N(\bar{\boldsymbol{\theta}}_N)^2}{\kappa_{N,2} \delta_N(\bar{\boldsymbol{\theta}}_N) + \kappa_{N,3} \frac{1-\rho}{T}} =: EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N). \quad (16)$$

EU_{smv} and EU_{2f} depend on several parameters.⁹ First, the sample size T , directly but also via

⁹The EU of the SMV portfolio and the two-fund rule is proportional to $1/\gamma$, and thus, the N optimizing the EU does not depend on γ . Therefore, we do not highlight the dependence on γ in EU_{smv} and EU_{2f} .

c_N and $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3})$. Second, the portfolio size N , directly but also via c_N , the function δ_N , and $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3})$. Third, *which* N assets are selected via $\bar{\theta}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$.

Now, to maximize $EU_{smv}(N, T, \rho, \tau, \bar{\theta}_N)$ or $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ with respect to N , one would also need to optimize the selection of assets because $\bar{\theta}_N$ depends on N via the chosen subset of assets. This is a difficult optimization problem because the number of possible subsets of N assets with N going from one to M grows very quickly with M .¹⁰ Moreover, this would introduce potentially severe estimation risk and time instability in the selection of assets, which is undesirable. This is why, to disentangle the determination of the optimal N from the asset selection, we replace $\bar{\theta}_N$ by its counterpart for the whole investment universe, i.e., $\bar{\theta}_M$. This approximation may not be accurate when N is small relative to M .¹¹ In this case, however, the estimated $\bar{\theta}_N$ will be particularly noisy, and thus estimating it from $\bar{\theta}_M$ can help too. All in all, the way we determine the optimal portfolio size N for the SMV portfolio and the optimal two-fund rule becomes¹²

$$N_{smv}^* = \underset{N \in \{1, \dots, \min(M, T-5)\}}{\operatorname{argmax}} EU_{smv}(N, T, \rho, \tau, \bar{\theta}_M), \text{ and} \quad (17)$$

$$N_{2f}^* = \underset{N \in \{1, \dots, \min(M, T-5)\}}{\operatorname{argmax}} EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M). \quad (18)$$

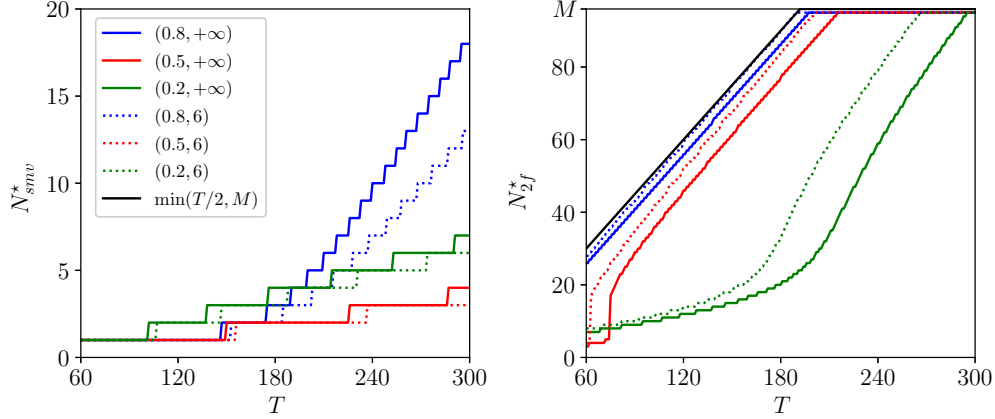
We now illustrate N_{smv}^* and N_{2f}^* . Using the full sample of the 96S-BM dataset introduced in Section 2, we obtain $\hat{\mu}_{96}$ and $\hat{\Sigma}_{96}$ using (5). Then, we set the population mean $\mu_M = \hat{\mu}_{96}$ and the population covariance matrix $\Sigma_M = \hat{\Sigma}_{96}(\rho) = \hat{D}_{96} P_{96}(\rho) \hat{D}_{96}$ with $\hat{D}_{96} = \operatorname{diag}(\hat{\Sigma}_{96})^{1/2}$, which yields $\bar{\theta}_{96} = (0.125, 0.0169)$, and we take $\rho \in \{0.2, 0.5, 0.8\}$. We consider either the normal

¹⁰The number of possible subsets of N assets out of $M = 100$ ones is equal to $\sum_{N=1}^M \binom{M}{N} \approx 1.26 \times 10^{30}$.

¹¹In Section OA.3 of the Online Appendix, we show that $\bar{\theta}_N$ quickly converges to $\bar{\theta}_M$ in practice.

¹²We search N up to $\min(M, T-5)$ because we assume $T > N+4$ in Proposition 2. In Section OA.3 of the Online Appendix, we compare, for the optimal two-fund rule, the dependence on N of the EU obtained with $\bar{\theta}_M$ and $\bar{\theta}_N$, i.e., $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$ and $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ in (16). $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ depends on N via the sequence of assets considered, and thus, we generate a large number of random asset sequences, evaluate $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ for each sequence, and compare it to $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$. We find that $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$ provides a good approximation and that the dispersion of $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ across different asset sequences is reasonable.

Figure 2: Optimal portfolio size for the SMV portfolio and the optimal two-fund rule



Notes. This figure depicts the optimal portfolio size for the SMV portfolio, N_{smv}^* in (17) (left panel), and for the optimal two-fund rule, N_{2f}^* in (18) (right panel), as a function of the sample size T . Each line represents a different choice of the correlation, $\rho = \{0.2, 0.5, 0.8\}$, and the degrees of freedom of the t -distribution, $\nu = \{6, \infty\}$. We calibrate $\bar{\theta}_M = (0.125, 0.0169)$ to a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023.

distribution, i.e., $\tau = 1$ and $\kappa_{N,1} = \kappa_{N,2} = \kappa_{N,3} = 1$, or the t -distribution, i.e., $\tau \sim (\nu - 2)/\chi_\nu^2$, with $\nu = 6$. Figure 2 depicts N_{smv}^* and N_{2f}^* as a function of T . We make several observations. First, N_{smv}^* and N_{2f}^* increase with T because, as T increases, the estimated portfolios become better estimates of the true MV portfolio for which the optimal $N = M$.

Second, N_{smv}^* is very small compared to T and M . For instance, with $T = 240$ and $\rho = 0.5$, $N_{smv}^* = 3$ both for the normal and t -distributions. For a larger $\rho = 0.8$, N_{smv}^* gets larger but is still small. For example, when $\rho = 0.8$ and $T = 120, 180$ and 240 , we have $N_{smv}^* = 1, 3$, and 10 for the normal distribution, and $N_{smv}^* = 1, 2$, and 7 for the t -distribution, respectively.

Third, N_{smv}^* and N_{2f}^* tend to increase with ρ . This is because, as we show in Section OA.2 of the Online Appendix, the maximum utility without parameter uncertainty, $U(\mathbf{w}^*)$, is a convex function of ρ , and thus, is increasing in ρ for ρ large enough. Therefore, given that $U(\mathbf{w}^*)$ is an increasing function of N , a larger ρ tends to increase the optimal N .

Fourth, N_{2f}^* is higher than N_{smv}^* , below $T/2$, equal to M if $T/2$ is enough above M , and gets closer to $T/2$ as ρ increases. These results can be explained based on (16) and (18). Specifically,

we can show that in the normal case, i.e., $\kappa_{N,1} = \kappa_{N,2} = \kappa_{N,3} = 1$, N_{2f}^* is below $T/2$ (and equal to M if M is sufficiently below $T/2$), and close to $T/2$ when ρ is close to zero or one. To see this, consider the N maximizing the first term of EU_{2f} , i.e., N/c_N . We have

$$\frac{N}{c_N} = \frac{N(T - N - 1)(T - N - 4)}{(T - 2)(T - N - 2)} \approx \frac{N(T - N - 4)}{(T - 2)}, \quad (19)$$

and the latter is maximized by

$$\operatorname{argmax}_{N \in \{1, \dots, \min(M, T-5)\}} \frac{N(T - N - 4)}{(T - 2)} = \min([T/2 - 2], M). \quad (20)$$

Moreover, when $\kappa_{N,1} = \kappa_{N,2} = \kappa_{N,3} = 1$, the second term of EU_{2f} in (16) decreases with N and is maximized by $N = 1$, which moves N_{2f}^* below (20). However, the sensitivity of this second term to N depends on how sensitive $N\rho/(1 - \rho + N\rho)$ is to N , and it is less so as ρ approaches zero or one. In equity data where ρ is typically large, we thus expect N_{2f}^* to stay close to (20). Finally, in the elliptical case, $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$ also depend on N , but Figure 2 and later simulations suggest that N_{2f}^* still remains close to that under normality.

Finally, fat tails positively impact N_{2f}^* , while we observe the opposite for N_{smv}^* . This can be explained because the two-fund rule optimally scales down the SMV portfolio according to tail heaviness via $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$. Therefore, as shown by Kan and Lassance (2025, Proposition 5), the two-fund rule often performs better when asset returns are elliptically instead of normally distributed, and thus we find a larger N_{2f}^* under elliptical returns.

4 Simulation analysis

In Section 4.1, we study the estimated optimal N for the SMV portfolio and two-fund and three-fund rules. In Section 4.2, we test whether the EU is close to optimal under the *estimated* optimal

N and when Assumption 1, i.e., equicorrelation, is not satisfied. In Section 4.3, we analyze the effect of N on the EU delivered by alternative benchmark portfolio strategies.

To run this analysis, we draw monthly excess returns from a t -distribution, i.e., $\tau \sim (\nu - 2)/\chi_\nu^2$, with $\nu \in \{6, \infty\}$. As in Section 3.2, we calibrate the mean and covariance matrix to the full sample of 96S-BM dataset introduced in Section 2. Specifically, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and consider $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, where $\bar{\rho} = 0.74$ is the average of the correlations in $\hat{\boldsymbol{\Sigma}}_{96}$. We also consider varying ρ . Setting $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$ allows us to compare the estimated optimal values of N with the oracle optimal one that, for the SMV portfolio and the two-fund and three-fund rules, is known theoretically under Assumption 1, i.e., when $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$.

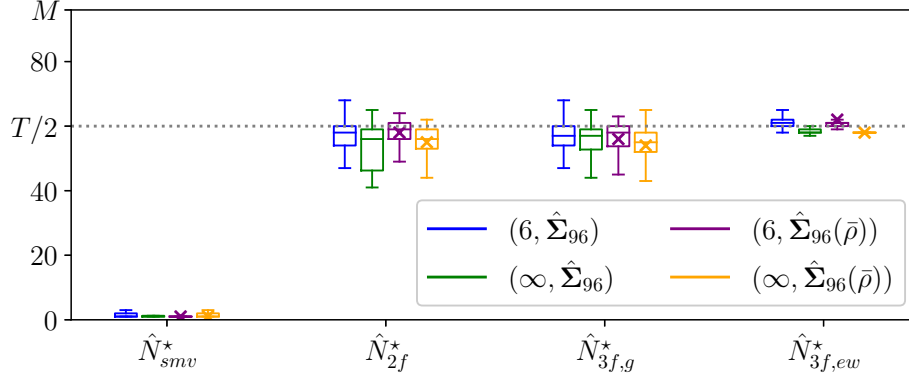
Throughout this simulation analysis, the EU gains from reducing portfolio size originate from the benefits of a reduction in parameter dimensionality on out-of-sample performance. In-sample, with $\boldsymbol{\mu}_M$ and $\boldsymbol{\Sigma}_M$ being known, it is optimal to have $N = M$.

4.1 Estimated optimal portfolio size

We now analyze the estimated optimal portfolio size for the SMV portfolio ($\hat{N}_{smv}^*$), the two-fund rule ($\hat{N}_{2f}^*$), the GMV-three-fund rule ($\hat{N}_{3f,g}^*$), and the EW-three-fund rule ($\hat{N}_{3f,ew}^*$). These are the estimated counterparts of N_{smv}^* in (17), N_{2f}^* in (18), $N_{3f,g}^*$ in (A10), and $N_{3f,ew}^*$ in (A19) following the estimation methodology in Section A.2 of the Appendix.

We simulate $K = 10,000$ times $T = 120$ returns and estimate the optimal N in each simulation. Figure 3 depicts boxplots of $\hat{N}^*$ for the different strategies and choices of $(\nu, \boldsymbol{\Sigma}_M)$. We make several observations. First, when $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, $\hat{N}^*$ is close to the oracle N^* , highlighting the quality of our estimation procedure. Second, $\hat{N}^*$ is similar under $\hat{\boldsymbol{\Sigma}}_{96}$ and $\hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, meaning that drawing

Figure 3: Estimated optimal portfolio size in simulated data



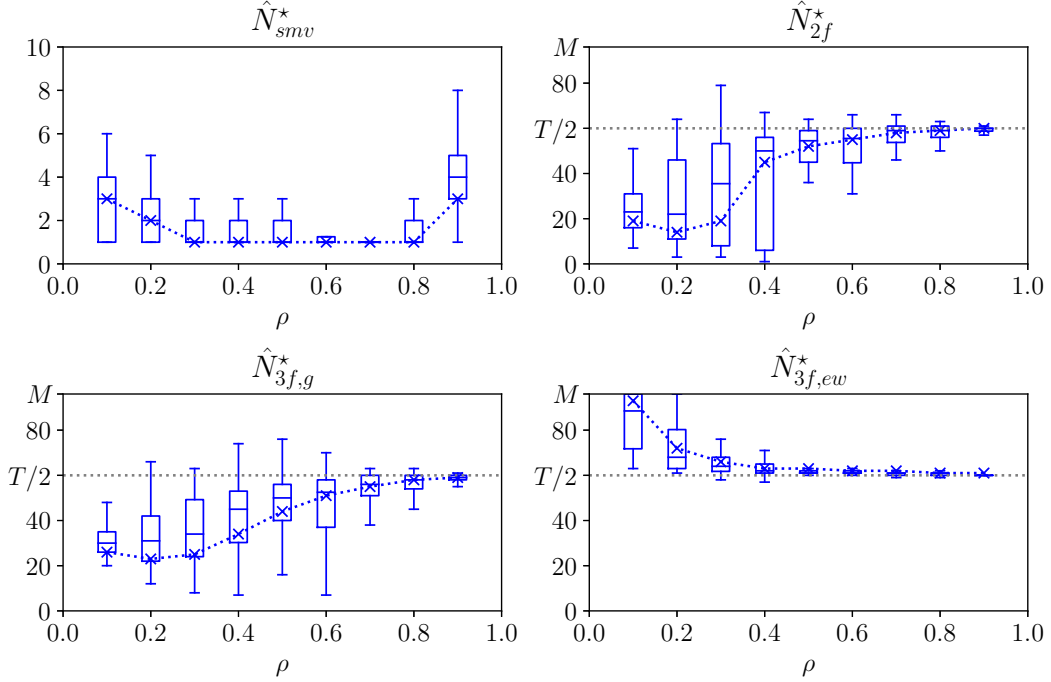
Notes. This figure depicts boxplots of the estimated optimal portfolio size N for the SMV portfolio ($\hat{N}_{smv}^*$), two-fund rule ($\hat{N}_{2f}^*$), GMV-three-fund rule ($\hat{N}_{3f,g}^*$), and EW-three-fund rule ($\hat{N}_{3f,ew}^*$). These are the estimated counterparts of N_{smv}^* in (17), N_{2f}^* in (18), $N_{3f,g}^*$ in (A10), and $N_{3f,ew}^*$ in (A19) following the estimation methodology in Section A.2 of the Appendix. The boxplots are obtained by simulating 10,000 times $T = 120$ t -distributed returns. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M \in \{\hat{\Sigma}_{96}, \hat{\Sigma}_{96}(\bar{\rho})\}$, where $\hat{\Sigma}_{96}(\bar{\rho})$ is an equicorrelation covariance matrix and $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\Sigma}_{96}$. We consider $\nu \in \{6, \infty\}$ degrees of freedom. Each boxplot corresponds to a different choice of (ν, Σ_M) . We depict with crosses the oracle value of the optimal N that is known under Assumption 1, i.e., when Σ_M is of the form $\Sigma_M(\rho)$.

returns from a distribution violating Assumption 1 does not materially affect the estimation of N .

Third, $\hat{N}_{smv}^*$ is very small, in line with Figure 2. Fourth, $\hat{N}^*$ is similar across the two-fund and three-fund rules and is typically slightly below $T/2$, which is because $\bar{\rho}$ is rather large; see Section 3.2. Finally, the lower the ν the higher the $\hat{N}^*$ in the two-fund and three-fund rules, in line with Figure 2.

Next, in Figure 4, we let $\Sigma_M = \hat{\Sigma}_{96}(\rho)$ and depict the oracle N^* and boxplots of $\hat{N}^*$ for ρ varying between 0.1 and 0.9. We set $\nu = 6$; the results are similar for $\nu = \infty$. We observe the following. First, the oracle N^* is slightly below $T/2$ when ρ is close enough to one, in line with Figure 2 where $\rho = \bar{\rho} = 0.74$. In this case of high correlation, typical for equity data, $\hat{N}^*$ is remarkably robust to parameter variability as the boxplots are thin. Second, for the SMV, 2F, and 3FGMV portfolios, the oracle N^* first decreases with ρ and then reincreases, attaining its maximum as ρ approaches one. This is because we want to invest in more assets when the maximum Sharpe ratio, θ_N in (4), is larger, which also first decreases then reincreases with ρ ; see Section OA.2 of the Online Appendix. For ρ away from zero and one, the boxplots widen as $\hat{N}^*$ becomes more sensitive

Figure 4: Impact of correlation ρ on optimal portfolio size in simulated data



Notes. This figure depicts boxplots of the estimated optimal portfolio size N for the SMV portfolio ($\hat{N}_{smv}^*$), two-fund rule ($\hat{N}_{2f}^*$), GMV-three-fund rule ($\hat{N}_{3f,g}^*$), and EW-three-fund rule ($\hat{N}_{3f,ew}^*$). These are the estimated counterparts of N_{smv}^* in (17), N_{2f}^* in (18), $N_{3f,g}^*$ in (A10), and $N_{3f,ew}^*$ in (A19) following the estimation methodology in Section A.2 of the Appendix. The boxplots are obtained by simulating 10,000 times $T = 120$ t -distributed returns. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}(\rho)$, where $\hat{\Sigma}_{96}(\rho)$ is an equicorrelation covariance matrix and ρ varies between 0.1 and 0.9 with a step size of 0.1. We consider $\nu = 6$ degrees of freedom. We depict with dotted lines and crosses the oracle value N^* of the optimal N that is known under Assumption 1, i.e., when Σ_M is of the form $\Sigma_M(\rho)$.

to the parameters. Third, ρ has a different effect for the 3FEW portfolio: N^* decreases with ρ . This can be explained because 3FEW combines SMV with EW, which is not subject to estimation risk. Thus, when ρ decreases and θ_N decreases too, 3FEW mostly invests in the EW portfolio and it is optimal to have a large N even under estimation risk. Finally, $\hat{N}^*$ is overall close to N^* on average.

4.2 Performance of size-optimized portfolios

We now evaluate the performance of the size-optimized SMV, two-fund, and three-fund rules using $N = \hat{N}^*$. For each choice of (ν, Σ_M) , where $\nu \in \{6, \infty\}$ and $\Sigma_M \in \{\hat{\Sigma}_{96}, \hat{\Sigma}_{96}(\bar{\rho})\}$, we draw $K = 10,000$ times $L = 300$ t -distributed returns, which we use in a rolling-window exercise.

Specifically, in each simulation k , given the sample size $T = 120$ and a chosen N , we randomly select in each rolling window a subset of N assets,¹³ estimate the portfolio on these N assets (see our estimation methodology in Section A.2 of the Appendix), and evaluate the out-of-sample portfolio return $r_{t,k}$ on the next month. We also implement the EW portfolio, $\mathbf{w}_{ew} = \mathbf{1}_N/N$, on the same N assets for comparison. We then roll the window by one month and proceed similarly until we reach the end of the sample of L returns. This gives us, for any N , a time series of out-of-sample portfolio returns $r_{t,k}$, $t = 1, \dots, L - T$ and $k = 1, \dots, K$, and we compute the EU as

$$EU = \frac{1}{K} \sum_{k=1}^K \left(\hat{\mu}_k - \frac{\gamma}{2} \hat{\sigma}_k^2 \right) \quad \text{with} \quad \hat{\mu}_k = \frac{1}{L-T} \sum_{t=1}^{L-T} r_{t,k}, \quad \hat{\sigma}_k^2 = \frac{1}{L-T} \sum_{t=1}^{L-T} (r_{t,k} - \hat{\mu}_k)^2. \quad (21)$$

In Figure 5, we vary N from five to $M = 96$ and depict the results for $\nu = 6$ and $\Sigma_M = \hat{\Sigma}_{96}$, which is the most relevant case. In Section OA.6.1 of the Online Appendix, we set $\nu = \infty$ and $\Sigma_M = \hat{\Sigma}_{96}$ and, in Section OA.6.2, $\nu \in \{6, \infty\}$ and $\Sigma_M = \hat{\Sigma}_{96}(\bar{\rho})$. The four panels in Figure 5 consider the SMV, 2F, 3FGMV, and 3FEW portfolios. Each panel depicts the EU as a function of N with a solid line using $\gamma = 1$, and also of the EW portfolio with a dash-dotted red line. In addition, we depict with an horizontal dashed line the EU obtained by using the estimated optimal portfolio size, $N = \hat{N}_{t,k}^*$, which varies across rolling windows and simulations.¹⁴

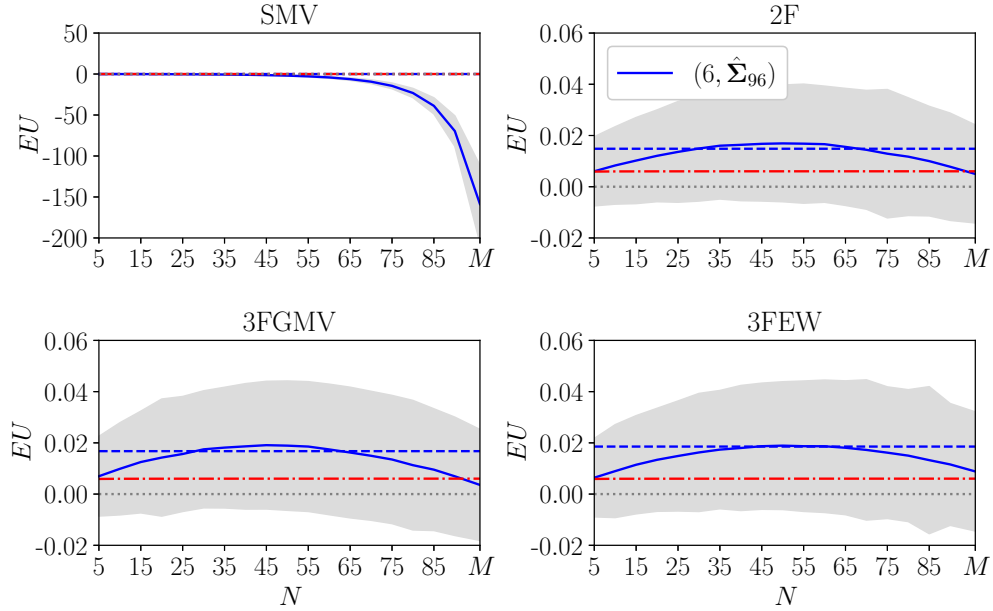
We make several observations. First, $\hat{N}^*$ delivers an EU close to the maximum, even though equicorrelation does not hold, there is estimation error in $\hat{N}^*$, and the asset selection is random.¹⁵ Second, for 2F, 3FGMV, and 3FEW, the EU is maximized for N slightly below $T/2$, in line with Section 3.2 and the average correlation being large ($\bar{\rho} = 0.74$). In contrast, the optimal N is very

¹³In the empirical analysis of Section 5, we use different selection rules to improve upon a random selection. In the current simulation exercise we are interested in general conclusions, independent of one specific asset selection rule, averaged over a large number of random selections.

¹⁴In Section OA.6.3 of the Online Appendix, we implement a similar exercise where the correlation matrix has a block structure. Even in this setup that violates Assumption 1, there are significant gains from reducing the portfolio size from M to $\hat{N}^*$ for the SMV, 2F, 3FGMV, and 3FEW portfolio rules.

¹⁵In Section OA.6.2 of the Online Appendix, we show that this result is robust under equicorrelation to considering different values of ρ and, in particular, low values for which $\hat{N}^*$ has a large variance as shown in Figure 4.

Figure 5: Expected out-of-sample utility in simulated data ($\nu = 6, \Sigma_M = \hat{\Sigma}_{96}$)



Notes. This figure depicts the expected out-of-sample utility (EU) of the sample mean-variance portfolio (SMV, top-left panel), two-fund rule (2F, top right), GMV-three-fund rule (3FGMV, bottom left), EW-three-fund rule (3FEW, bottom right) as a function of the portfolio size N . We simulate 10,000 samples of t -distributed returns with $\nu = 6$ degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}$. In each panel, the solid blue line depicts the EU in (21), the shaded gray area depicts the one-sigma interval around the EU across simulations, and the dashed horizontal blue line depicts the EU obtained when using the estimated optimal N , i.e., $\hat{N}_{smv}^*$ for the SMV portfolio, $\hat{N}_{2f}^*$ for 2F, $\hat{N}_{3f,g}^*$ for 3FGMV, and $\hat{N}_{3f,ew}^*$ for 3FEW. The dash-dotted red line depicts the EU of the equally weighted portfolio. The dotted horizontal gray line depicts the zero EU level. The risk-aversion coefficient is $\gamma = 1$.

small for SMV, as in Figure 2. Third, for 2F, 3FGMV, and 3FEW, $\hat{N}^*$ delivers an EU substantially larger than that when N is too small or too close to M , highlighting the value of choosing the right portfolio size and the cost of blindly investing in all assets. This EU obtained with $\hat{N}^*$ substantially outperforms the naive EW portfolio, which is a challenging benchmark for M close to T .

4.3 Comparison to benchmarks

We now consider benchmark portfolio strategies that we also consider in the empirical analysis of Section 5. First, we consider two natural ways of reducing portfolio size, which are soft and hard-thresholding versions of the SMV portfolio, called SMV-ST and SMV-HT.

SMV-ST solves the mean-variance portfolio problem subject to an L_1 -norm constraint:

$$\hat{\mathbf{w}}_{SMV-ST} = \underset{\mathbf{w} \in \mathbb{R}^M}{\operatorname{argmax}} \quad \mathbf{w}' \hat{\boldsymbol{\mu}}_M - \frac{\gamma}{2} \mathbf{w}' \hat{\boldsymbol{\Sigma}}_M \mathbf{w} \quad \text{subject to} \quad \|\mathbf{w}\|_1 = \sum_{i=1}^M |w_i| \leq \delta. \quad (22)$$

The L_1 -norm constraint performs asset selection and limits short-selling, which helps improve out-of-sample performance (DeMiguel et al., 2009a). We calibrate the threshold δ using cross-validation with out-of-sample utility as a decision criterion. Specifically, we search the optimal δ in an interval of 1,000 equally spaced values ranging from zero to $\delta_{\max}$, defined as the norm of the unconstrained solution to (22), i.e., $\delta_{\max} = \|\frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\boldsymbol{\mu}}_M\|_1$.

SMV-HT selects the portfolio size by keeping only the assets whose absolute SMV portfolio weight is above a threshold $\bar{w}$. In that case, the value of N is determined by $\bar{w}$ as

$$\hat{N}_{SMV-HT} = \sum_{i=1}^M \mathbb{1}_{\{|\hat{w}_i^*| \geq \bar{w}\}} \quad \text{where} \quad \hat{\mathbf{w}}^* = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\boldsymbol{\mu}}_M. \quad (23)$$

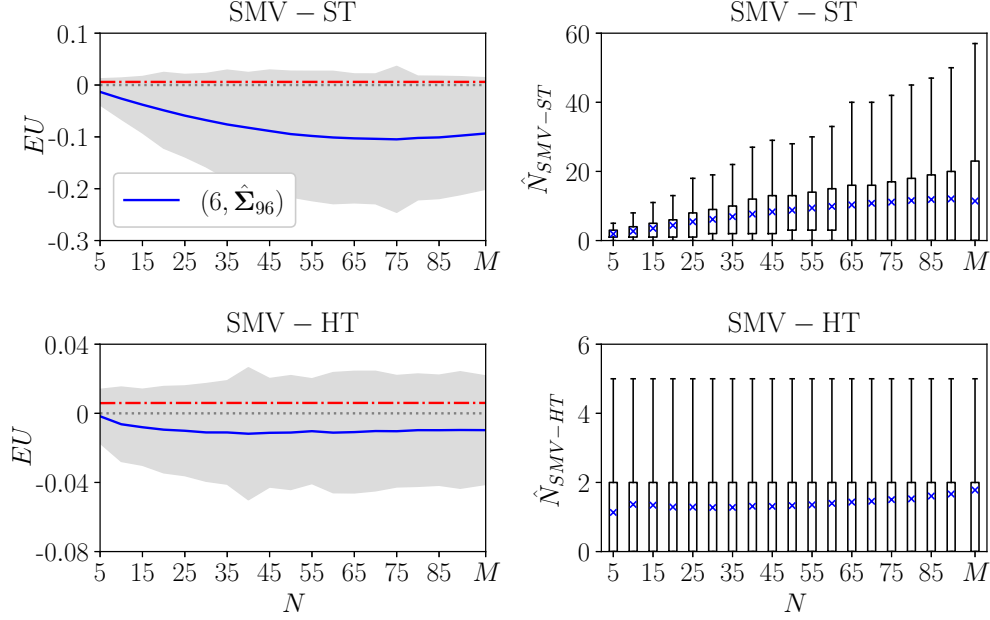
Once the $\hat{N}_{SMV-HT}$ assets are selected, we recompute the SMV portfolio on these assets, which delivers better performance than keeping the original weights. We select the threshold $\bar{w}$ via cross-validation with out-of-sample utility as a decision criterion. Specifically, we search $\bar{w}$ in an interval of 10 equally spaced values ranging from zero (all assets selected) to $\bar{w}_{\max}$ (no asset selected), where $\bar{w}_{\max} = \max|\frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\boldsymbol{\mu}}_M|$ is the largest absolute weight of the full SMV portfolio.¹⁶

In Figure 6, we depict the EU of SMV-ST and SMV-HT as a function of the number of assets on which they are estimated using the same setup as in Figure 5, with $\nu = \infty$.¹⁷ We also depict boxplots of the estimated optimal N under SMV-ST and SMV-HT. We make several observations. First, SMV-ST and SMV-HT substantially improve upon the plain SMV portfolio in Figure 5. Second, SMV-ST and SMV-HT deliver an EU that is consistently negative. Therefore, they are

¹⁶Given the large estimation risk of the SMV portfolio, using more than 10 values yields similar results because cross-validation typically yields $\bar{w}$ close to $\bar{w}_{\max}$ such that we keep only a few assets.

¹⁷In Section OA.6.1 of the Online Appendix, we obtain similar conclusions with $\nu = \infty$ and $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}$.

Figure 6: Expected out-of-sample utility and estimated portfolio size of soft and hard-thresholding of the sample mean-variance portfolio in simulated data ($\nu = 6$, $\Sigma_M = \hat{\Sigma}_{96}$)

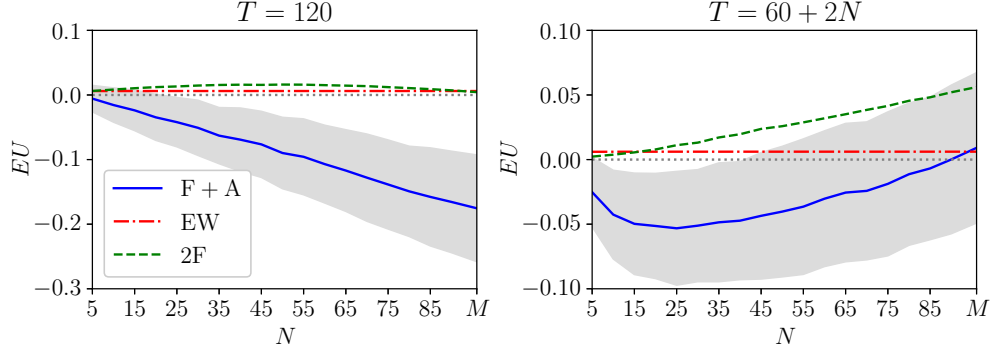


Notes. This figure depicts the expected out-of-sample utility (EU) and the estimated optimal portfolio size of the soft and hard-thresholding versions of the sample mean-variance portfolio, SMV-ST and SMV-HT, described in Section 4.3. We depict these as a function of the number of assets N on which SMV-ST and SMV-HT are estimated. We simulate 10,000 samples of t -distributed returns with $\nu = 6$ degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}$. In the left panels, the solid blue line depicts the EU of SMV-ST or SMV-HT in (21), the shaded gray area depicts the one-sigma interval around the EU across simulations, the dash-dotted red line depicts the EU of the equally weighted portfolio, and the dotted horizontal gray line depicts the zero EU level. The risk-aversion coefficient is $\gamma = 1$. In the right panels, the blue crosses depict the average estimated optimal portfolio size.

less effective approaches to reducing portfolio size than our size-optimized two-fund and three-fund rules in Figure 5. Third, SMV-ST and SMV-HT set a small optimal N , in line with the small EU-optimal portfolio size for the SMV portfolio that we obtain theoretically; see Figure 2.

The final benchmark imposes a small-dimensional factor model in the estimation. Specifically, as detailed in Section OA.4 of the Online Appendix, we construct a factor-plus-alpha (F+A) strategy, where the factor part is estimated via PCA, the alpha part via the arbitrage portfolio methodology of Da et al. (2024), and the factor and alpha portfolios are combined to maximize expected utility. This is a relevant benchmark for two reasons. First, although it does not reduce the portfolio size, it still reduces the number of parameters to estimate due to the factor model. Second, Da

Figure 7: Expected out-of-sample utility of factor-plus-alpha strategy in simulated data ($\nu = 6, \Sigma_M = \hat{\Sigma}_{96}$)



Notes. This figure depicts the expected out-of-sample utility (EU) of the factor-plus-alpha (F+A) strategy, described in Section OA.4 of the Online Appendix, as well as the two-fund rule (2F), as a function of the portfolio size N . We simulate 10,000 samples of t -distributed returns with $\nu = 6$ degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}$. In each panel, the solid blue and green lines depict the EU in (21) of F+A and 2F, respectively, the shaded gray area depicts the one-sigma interval around the EU of F+A across all simulations, and the dash-dotted red line depicts the EU of the equally weighted portfolio. The dotted horizontal gray line depicts the zero EU level. In the left panel, the sample size $T = 120$ is fixed. In the right panel, the sample size increases with N as $T = 60 + 2N$. The risk-aversion coefficient is $\gamma = 1$.

et al. (2024, Theorem 4) demonstrate that in the limit as N and T get large, and under specific assumptions, their arbitrage portfolio attains the maximum achievable Sharpe ratio over all portfolios having zero factor exposure. Therefore, the F+A strategy might work well also for a large N .

In Figure 7, we consider the same setup as in Figures 5 and 6 and depict the EU delivered by the F+A strategy as a function of N . In the left panel, the sample size $T = 120$. In the right panel, $T = 60 + 2N$ increases with N to capture the asymptotics in Da et al. (2024). For comparison, we also depict the EU of the two-fund rule; we find a similar conclusion for three-fund rules. We observe that when T is fixed, the EU of F+A decreases with N and substantially underperforms the 2F and EW portfolios. However, when T increases with N , the EU of F+A tends to improve with N and is maximized at $N = M$, where it slightly outperforms the EW portfolio. Nonetheless, the EU of the 2F strategy also improves with N in that case, and outperforms F+A for all N . Overall, these results suggest that, in our simulation setting, the 2F strategy delivers a better out-of-sample performance under estimation risk than the F+A strategy constructed from Da et al. (2024).

5 Empirical analysis

In Section 5.1 and 5.2, we present the datasets and portfolio strategies, respectively. In Section 5.3, we propose different rules to select in which assets to invest. In Section 5.4, we detail the out-of-sample methodology and performance measures. Finally, we discuss the results in Section 5.5. In this empirical analysis, the out-of-sample performance gains from reducing portfolio size originate from parameter uncertainty, but also from the inclusion of transaction costs.

5.1 Datasets

Our empirical analysis is based on six datasets of characteristic and industry portfolios and one dataset of 100 individual stocks, which we list in Table 1 with the time period considered for each. Given that our objective is to reduce the portfolio size, the full investment universe must not be too small in the first place, and thus, we consider datasets with M around 100.

The first dataset already introduced in Section 2, 96S-BM, is composed of decile portfolios sorted on size and book-to-market. The second dataset, 108CHA, is composed of the long and short legs of the 54 characteristics considered in Lassance and Martín-Utrera (2023), collected from the authors. The third dataset, 100S-OP, is composed of decile portfolios sorted on size and operating profitability. The fourth dataset, 94IN-NV, is composed of 48 industry portfolios and the long and short legs obtained from the 23 characteristics in Novy-Marx and Velikov (2015), available on Robert Novy-Marx’s website. The fifth dataset, 107IN-CHA, is composed of 47 industry portfolios,¹⁸ quintile portfolios sorted on size and book-to-market and on operating profitability and investment, and decile portfolios sorted on momentum. The sixth dataset, 98IN-CHA-NV, is

¹⁸We consider 47 portfolios instead of 48 because of missing data in the healthcare portfolio before July 1969.

Table 1: List of datasets considered in the empirical analysis

#	Name	Description	M	Time period
1	96S-BM	96 portfolios sorted on size and book-to-market	96	07/1963–08/2023
2	108CHA	108 characteristic portfolios built on long and short legs of 54 characteristics in Lassance and Martín-Utrera (2023)	108	09/1966–12/2020
3	100S-OP	100 portfolios sorted on size and operating profitability	100	07/1963–08/2023
4	94IN-NV	48 industry portfolios and 46 characteristic portfolios built on long and short legs of 23 characteristics in Novy-Marx and Velikov (2015)	94	07/1973–12/2013
5	107IN-CHA	47 industry portfolios, 25 portfolios sorted on size and book-to-market, 25 portfolios sorted on operating profitability and investment, 10 portfolios sorted on momentum	107	07/1963–12/2022
6	98IN-CHA-NV	47 industry portfolios, 25 portfolios sorted on operating profitability and investment, 10 portfolios sorted on momentum, 16 characteristic portfolios built on long and short legs of eight low-turnover characteristics in Novy-Marx and Velikov (2015)	98	07/1963–12/2013
7	100STO	100 random datasets of 100 U.S. individual stocks	100	01/1998–03/2022

Notes. This table lists the datasets of monthly excess returns considered in the empirical analysis of Section 5, which we detail in Section 5.1. The columns report, in order, the dataset number, name, description, size M , and time period. The industry portfolios and the portfolios sorted on size, book-to-market, operating profitability, investment, and momentum are downloaded from Kenneth French’s website. The characteristic portfolios in Novy-Marx and Velikov (2015) are downloaded from Robert Novy-Marx’s website. In the construction of these datasets, we have removed assets with missing data over the time period considered. The 108CHA dataset is the same as that used by Lassance and Martín-Utrera (2023). In the construction of the 108CHA dataset, Lassance and Martín-Utrera (2023) drop characteristics with more than five percent of missing observations for more than five percent of firms with CRSP returns available for the entire sample. For the 100STO dataset, we follow Lassance et al. (2024b) and collect adjusted returns from CRSP for the 235 stocks traded on the three major U.S. stock exchanges that have an history of returns between 1998 and 2022, and we report the portfolio performance across 100 datasets of 100 stocks, randomly drawn from the total pool of 235 stocks, after merging the out-of-sample portfolio returns across the 100 datasets.

composed of 47 industry portfolios, quintile portfolios sorted on size and book-to-market, decile portfolios sorted on momentum, and the long and short legs obtained from the eight low-turnover characteristics in Novy-Marx and Velikov (2015). For the seventh and final dataset, 100STO, we follow Lassance et al. (2024b) and collect adjusted returns from CRSP for the 235 stocks traded on the three major U.S. stock exchanges that traded between 1998 and 2022. We consider 100 datasets of size $M = 100$, randomly drawn from the total pool of 235 stocks, and report the portfolio performance after merging the out-of-sample portfolio returns across the 100 datasets.

In Section OA.7.1 of the Online Appendix, we look at how Assumption 1 (equicorrelated re-

turns) and Assumption 2 (elliptical returns) transpire in the seven datasets. First, the equicorrelation assumption is reasonable for the 96S-BM, 100S-OP, and 108CHA datasets, with a root mean squared error (RMSE) between the observed and equicorrelation matrices of around 5-10%, but less so for the four other datasets, with an RMSE around 10-20%. Thus, we can assess the performance of our optimal portfolio size across different correlation structures. Second, the parameters $(\kappa_{M,1}, \kappa_{M,2}, \kappa_{M,3})$ defined in (10)–(12), which control the impact of the elliptical fat tails of returns on the EU, substantially depart from one, i.e., from normality. Thus, it is crucial to account for fat tails when implementing combination rules given their impact on combination coefficients.

5.2 Portfolio strategies

We evaluate the performance of the 11 portfolio strategies listed in Table 2. The first three are the optimal combination rules in Section 3.2 and Section A.1 of the Appendix: the two-fund rule (2F), GMV-three-fund rule (3FGMV), and EW-three-fund-rule (3FEW). We apply these strategies to all M assets or a subset of N assets, with N chosen optimally as $\hat{N}_{2f}^*$ in (18) for 2F, $\hat{N}_{3f,g}^*$ in (A10) for 3FGMV, and $\hat{N}_{3f,ew}^*$ in (A19) for 3FEW, estimated following Section A.2 of the Appendix.

The next two portfolio strategies are individual portfolios: the EW portfolio and the SGMV portfolio in (A1). We then consider two portfolio strategies that combine EW and SGMV with the risk-free asset to maximize the EU, EWRF and GMVRF, which might be preferred to 2F, 3FGMV, and 3FEW because they disregard the SMV portfolio that faces high estimation risk. We explain how we estimate the EWRF portfolio in Section OA.5 of the Online Appendix. Regarding the GMVRF portfolio, we follow Kan and Lassance (2025, Section VI.C.2.) and estimate it as

$$\hat{\mathbf{w}}_{gmvr f} = \frac{\kappa_{M,1}}{\kappa_{M,2}} \times \frac{T - M - 2}{T c_M} \times \frac{\hat{\mu}_{g,M} \hat{\Sigma}^{-1} \mathbf{1}}{\gamma}, \quad (24)$$

Table 2: List of portfolio strategies considered in the empirical analysis

Name	Description
2F	Two-fund rule that combines the SMV portfolio and the risk-free asset. See Equation (8) and Proposition 2.
3FGMV	Three-fund rule that combines the SMV portfolio, the SGMV portfolio, and the risk-free asset. See Equation (A3) and Proposition A.1.
3FEW	Three-fund rule that combines the SMV portfolio, the EW portfolio, and the risk-free asset. See Equation (A12) and Proposition A.3.
EW	Equally weighted portfolio
EWRF	Combination of EW and the risk-free asset. See Equation (OA15).
SGMV	Sample global minimum-variance portfolio. See Equation (A1).
GMVRF	Combination of SGMV and the risk-free asset. See Equation (24).
SMV	Sample mean-variance portfolio. See Equation (7).
SMV-ST	SMV portfolio with soft-thresholding and the L_1 -norm-constraint threshold selected via cross-validation. See Equation (22).
SMV-HT	SMV portfolio with hard-thresholding and the weight threshold selected via cross-validation.
F+A	Factor-plus-alpha strategy from Da et al. (2024). See Section OA.4 of the Online Appendix.

Notes. This table lists the portfolio strategies that we consider in the empirical analysis, which we detail in Section 5.2. We apply the 2F, 3FGMV, 3FEW, and EW portfolio strategies on a subset of N assets out of the M available ones, and the asset selection rules we implement for that purpose are described in Section 5.3.

where $\hat{\mu}_{g,M}$ is defined in (A27) and the estimation of $\kappa_{M,1}$ and $\kappa_{M,2}$ is described in Section A.2.1 of the Appendix.

The next portfolio strategy we consider is the SMV portfolio, which we implement in three different ways. In the first case, we compute the SMV portfolio on all M assets using Equation (7). In the second and third case, we reduce the portfolio size with soft and hard-thresholding using the SMV-ST and SMV-HT portfolios introduced in Section 4.3.¹⁹

The last portfolio strategy we implement is the factor-plus-alpha (F+A) strategy introduced in Section 4.3. We refer to Section OA.4 of the Online Appendix for more details.

Finally, we consider three different estimates of the covariance matrix. First, the sample estimator in (5) for consistency with our theory. Second, the linear shrinkage estimator of Ledoit and

¹⁹In Section OA.7.3 of the Online Appendix, we also report the performance of the size-optimized SMV portfolio with N determined using our theory in Section 3.2, i.e., using $\hat{N}_{smv}^*$. We find that the size-optimized SMV portfolio outperforms the full SMV on all M assets. We don't report these results in the main text because, even with a reduced N , the SMV portfolio largely underperforms 2F, 3FGMV, and 3FEW.

Wolf (2004). Third, the nonlinear shrinkage estimator of Ledoit and Wolf (2020), for which we report the results in Section OA.7.2 of the Online Appendix for conciseness.²⁰

Note that whereas we implement the EWRF, SGMV, GMVRF, SMV, SMV-ST, SMV-HT, and F+A benchmark portfolios on all M assets, we study the impact of the portfolio size N on their empirical performance in Section OA.7.8 of the Online Appendix.

5.3 Asset selection rules

Given a portfolio size $N \leq M$, we consider several *selection rules* to decide which N assets to select. Our purpose is not to find an optimal asset selection rule. Instead, our main objective is to illustrate the added value of optimally reducing the portfolio size, and that even simple but sensible selection rules can deliver substantial gains over choosing all M assets.

For comparison purposes, we apply the proposed selection rules not just to 2F, 3FGMV, and 3FEW but also to the EW portfolio, so as to evaluate their intrinsic value without being affected by estimation errors in optimized portfolio weights. In that case, we take $N = T/2$, where $T = 120$ months as explained in Section 5.4, because the optimal N is close to $T/2$ for typical levels of equity correlations as explained in our theory and simulations.

We propose 11 asset selection rules, listed in Table 3, satisfying two criteria: they are *simple* to understand and implement and, apart from the random rule, they are *sensible*, i.e., they have an economic intuition. The first selection rule consists in choosing all assets and setting $N = M$.

The 10 remaining selection rules select a subset of $N \leq M$ assets. Among those, the first one is

²⁰Analytical expressions for the EU of portfolio strategies implemented with shrinkage estimators of Σ_N are not available. Therefore, we cannot obtain the optimal combination coefficients and optimal N in those cases. As an alternative, we follow Kan et al. (2021) and implement the portfolios by relying on the obtainable combination coefficients and optimal N found under the sample estimator of the covariance matrix Σ_N , but use shrinkage estimators of Σ_N to compute the portfolio weights.

Table 3: List of asset selection rules considered in the empirical analysis

Name	Description
All	Include all assets: $N = M$.
Rand	Select N assets randomly.
MaxSR	Select N assets with the maximum Sharpe ratios.
MinSR	Select N assets with the minimum Sharpe ratios.
BWSR	Select $\lceil N/2 \rceil$ assets with the best Sharpe ratios and $\lfloor N/2 \rfloor$ assets with the worst Sharpe ratios.
MaxVar	Select N assets with the maximum variances.
MinVar	Select N assets with the minimum variances.
BWVar	Select $\lceil N/2 \rceil$ assets with the best variances and $\lfloor N/2 \rfloor$ assets with the worst variances.
MinPC	Select the N assets with the minimum correlations with the first principal component.
Best θ_N^2	Select the N assets that deliver the best portfolio Sharpe ratio as in Ao et al. (2019).
MaxW	Select the N assets with the maximum absolute weights.

Notes. This table lists the asset selection rules that we consider in the empirical analysis of Section 5, which we detail in Section 5.3. That is, in an investment universe of M assets, which N assets we select for a given $N \leq M$. The notation $\lceil N/2 \rceil$ ($\lfloor N/2 \rfloor$) means $N/2$ rounded above (below) to the nearest integer.

a random selection, the next six are based solely on marginal information about the assets, and the last three rules also consider information about their dependence. Moreover, the first nine rules select the assets before computing the portfolio weights, whereas the last rule selects the subset of assets after optimizing the weights on all assets. We now describe these 10 asset selection rules.²¹

Random selection (Rand). We randomly select the N assets among all M assets and report the performance across 100 repetitions of the out-of-sample analysis after merging all portfolio returns. This random rule serves as a benchmark to which other more sensible selection rules can be compared, and it is consistent with the simulation analysis of Section 4.

Maximum Sharpe ratio (MaxSR). We select the N assets with the maximum in-sample Sharpe ratios in the estimation window, because the EU of the SMV portfolio in (14) and the two-fund rule in (16) both depend on the assets' Sharpe ratios. This selection rule is expected to work particularly

²¹Our three-stage approach, as described in Section 1, is flexible and allows any asset selection rule. We propose a specific set of simple rules as an illustration. One may for instance consider more sophisticated rules based on asset clustering, time-series models, or machine learning, which may further improve performance.

well when there is momentum in the assets.²² We can anticipate several issues with this rule. First, it ranks the assets according to average returns that are notoriously noisy (Merton, 1980). Second, the selected assets have the best Sharpe ratios, and thus, we might short good assets. Third, we do not include assets with the worst Sharpe ratios although shorting them might deliver gains.

Minimum Sharpe ratio (MinSR). We select the N assets with the minimum in-sample Sharpe ratios. This rule will work well under mean-reversion in the assets, or from using short positions. We expect MinSR to face the same issues as MaxSR, i.e., not having access to the assets with the best Sharpe ratios to form long positions, and ending up with long positions on bad assets.

Best-worst Sharpe ratio (BWSR). We blend the MaxSR and MinSR rules by selecting the $\lceil N/2 \rceil$ assets with the best in-sample Sharpe ratios and the $\lfloor N/2 \rfloor$ assets with the worst in-sample Sharpe ratios, which can be combined with long and short positions. This is in line with the standard way of constructing anomaly portfolios as long-short portfolios of stocks in the top and bottom quantiles of a given firm characteristic. However, BWSR may lead to more extreme portfolio weights that do not generalize well out of sample.

The MaxSR, MinSR, and BWSR selection rules rank assets based on their in-sample Sharpe ratios.²³ Given that average returns are notoriously noisy, these rankings are bound to be unstable.²⁴ As a result, the set of selected assets is likely to change substantially from one period to another, increasing portfolio turnover and transaction costs. To address this point, the next three asset selection rules in Table 3 are the following.

Maximum, minimum, and best-worst variance (MaxVar, MinVar, BWVar). These selection

²²Six of our datasets are composed of characteristic and industry-sorted portfolios, which exhibit more momentum (i.e., positive return autocorrelation) than individual stocks (Campbell et al., 1997, p.74).

²³In unreported results, we implement selection rules that rank assets according to their in-sample utilities. We find that these rules deliver an out-of-sample performance close to the rules based on Sharpe ratios.

²⁴In unreported results, we implement selection rules that rank assets according to average returns. We find that they typically underperform selection rules based on the Sharpe ratio and variance.

rules rank assets based on the in-sample variance instead of Sharpe ratio. The resulting rankings of assets are expected to be more stable because the variance is quite persistent over time. The MaxVar rule selects assets with the maximum variances, which are bound to also have larger expected returns if a positive risk-return tradeoff stands. However, the low-risk anomaly implies that assets and portfolios with lower variances often have larger expected returns (Ang et al., 2006; Moreira and Muir, 2017), which the MinVar rule can exploit.

Minimum correlation with the first principal component (MinPC). We select the assets whose returns have minimum correlations with the first principal component (PC), which are least explained by the remaining assets, and thus, offer higher diversification potential.

Best portfolio Sharpe ratio (Best θ_N^2). This selection rule follows Ao et al. (2019) and selects the assets that deliver a high maximum squared Sharpe ratio, θ_N^2 in (2). It works in three steps. First, we form 1,000 random selections of N assets. Second, for each selection, we estimate θ_N^2 as detailed in Section A.2 of the Appendix. Third, we pick the selection that corresponds to the 95% quantile of all estimated values of θ_N^2 , to avoid outliers.

Maximum absolute weights (MaxW). The above selection rules are independent of the portfolio strategy. However, the in-sample portfolio weights on all M assets are a valuable information. Therefore, MaxW selects the assets based on the portfolio strategy that will be applied to the selected assets. It works in three steps. First, compute $\hat{\mathbf{w}} \in \mathbb{R}^M$, the portfolio weights on all M assets for a given strategy (2F, 3FGMV, or 3FEW). Second, select the N assets with the maximum absolute weights $|\hat{w}_i|$. Third, recompute the strategy on the N assets to obtain $\hat{\mathbf{w}} \in \mathbb{R}^N$.²⁵ MaxW selects the assets whose weights are large, and thus, identified as relevant. Note that MaxW does

²⁵We find that dropping this third step and keeping the portfolio weights computed on the M assets performs poorly because these weights are computed under high estimation risk, i.e., a large M/T .

not apply to the EW portfolio whose weights are equal. One issue we anticipate is that MaxW inherits large estimation errors as it is based on portfolio weights optimized under a large M/T .

In Section OA.7.7, we introduce a measure that determines how much more or less overlap there is between two asset selection rules relative to a pair of independent random selection rules. For essentially all pairs of selection rules we consider, we find that there is either around the same or less overlap than under a random selection. That is, our selection rules select dissimilar assets, which allows us to test our portfolio strategies on different investment universes.

5.4 Out-of-sample methodology

We evaluate the out-of-sample portfolio performance with a standard monthly rebalancing. Specifically, at the end of month t , we estimate portfolio k over the T previous months, and we compute its out-of-sample return in month $t + 1$. We consider a fixed sample size of $T = 120$ months.²⁶ We repeat this process iteratively, resulting in $T_{tot} - T$ out-of-sample gross returns $r_{gross,k,t}$, where T_{tot} is the total number of months in the dataset. We then compute the net out-of-sample returns, $r_{net,k,t} = r_{gross,k,t}$ if $t = T + 1$ and

$$r_{net,k,t} = (1 + r_{gross,k,t}) (1 - \kappa \times \text{turnover}_{k,t-1}) - 1 \quad \text{if } t = T + 2, \dots, T_{tot}, \quad (25)$$

where κ is the proportional transaction cost parameter and

$$\text{turnover}_{k,t} = \sum_{i=1}^N |w_{i,k,t} - w_{i,k,(t-1)+}|, \quad t = T + 1, \dots, T_{tot}, \quad (26)$$

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 $\kappa = 10$ basis points in line with the average bid-ask spreads reported by Engle

²⁶Given that our theoretical results require $T > M + 4$, and that the largest $M = 108$, we need $T > 112$. If we choose a sample size that is too large, such as $T = 240$, then the optimal N for the 2F, 3FGMV, and 3FEW portfolios, which is close to $T/2$ as shown in Section 4.1, will often be larger than the total size of the investment universe, M . In that case, we cannot assess the effect of reducing the portfolio size.

et al. (2012) and Frazzini et al. (2018). Finally, we compare the portfolio strategies in terms of *annualized out-of-sample utility net of transaction costs*,

$$U_k = 12 \times \left(\hat{\mu}_k - \frac{\gamma}{2} \hat{\sigma}_k^2 \right), \quad (27)$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the sample mean and variance of $r_{net,k,t}$. We set a risk-aversion coefficient of $\gamma = 1$.²⁷ As explained in Section 5.3, we repeat this out-of-sample analysis 100 times under the random asset selection rule and compute U_k after merging all portfolio returns.²⁸ In Section OA.7.4 of the Online Appendix, we also report the out-of-sample Sharpe ratio.

For 2F, 3FGMV, and 3FEW, the assets change as we re-estimate the optimal N and apply an asset selection rule. Because our methodology disentangles the determination of the optimal N from the choice of the N assets, we can also disentangle the rebalancing of portfolio weights from the asset selection. Therefore, to make our method less costly and more practical, we rebalance the portfolio weights every month but change the optimal N and selected assets every year.²⁹

Finally, we compute two-sided p -values for the statistical test of the difference between the net utility of the 2F, 3FGMV, and 3FEW portfolios computed on all assets versus under each of the 10 other selection rules in Table 3. For SMV-HT and SMV-ST, we report p -values relative to SMV. 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 p -values. We use the symbols $\bigcirc$, $\bullet$, and $\bullet$ to indicate that the p -value is less than 10%, 5%, and 1%, respectively.

²⁷The out-of-sample utility net of costs of the 2F, 3FGMV, 3FEW, GMVRF, EWRF, SMV, and F+A portfolios is proportional to $1/\gamma$, and thus, their ranking is not affected by the chosen γ . However, the ranking of the SGMV and EW portfolios, which are fully invested in risky assets, is affected by the chosen γ .

²⁸We obtain similar utility values when computing one utility per repetition and averaging them.

²⁹In Section OA.7.5 of the Online Appendix, we also consider changing the selected assets every 1, 6, and 24 months. We find that setting a lower frequency than monthly substantially reduces turnover and improves the net performance. Moreover, frequencies of 6, 12, or 24 months yield similar results.

5.5 Discussion of results

In Table 4, we report the annualized net out-of-sample utility, in percentage points, of the 11 portfolio strategies listed in Table 2. For the 2F, 3FGMV, 3FEW, and EW portfolio strategies, we report the performance across the 11 asset selection rules in Table 3.

The results in Table 4 show that portfolio strategies estimated with the shrinkage covariance matrix generally outperform those estimated with the sample one. Therefore, albeit our theory considers the sample covariance matrix for tractability, it is beneficial to shrink it. The ranking of portfolio strategies and asset selection rules is overall similar in both cases, and thus, unless stated otherwise, the discussion below essentially applies to either estimator.

First, we address our main objective, which is to assess the value of optimally reducing the portfolio size. For either of 2F, 3FGMV, and 3FEW, reducing the portfolio size delivers substantial performance gains. Specifically, the asset selection rule ‘All’, which sets $N = M$, is almost always outperformed by the remaining selection rules.³⁰ The differences are often statistically significant, particularly under the shrinkage covariance matrix. Moreover, for the six datasets of characteristic and industry portfolios, the differences are systematically positive for the Rand, MinSR, BWSR, MaxVar, BWVar, and MinPC asset selection rules, and are often substantial. For instance, under the shrinkage covariance matrix, on average across these six datasets, the 2F strategy on all M assets delivers an annualized net utility of 11.30 percentage points, versus 58.17 with the BWSR selection rule. Turning to the 100STO dataset, asset selection rules based on average returns (MaxSR, MinSR, BWSR) do not work as well because there is less predictability in average returns among

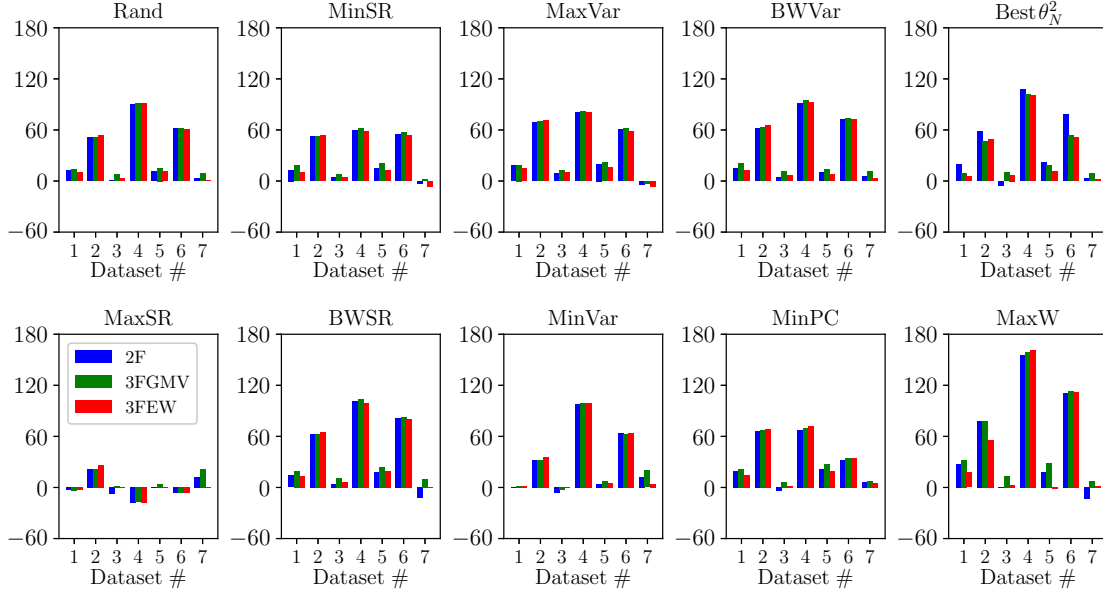
³⁰Out of the 210 configurations (10 selection rules $\times$ 7 datasets $\times$ 3 portfolio strategies), the size-optimized portfolios outperform the ‘All’ selection rule in 172, 183, and 187 cases under the sample, linear shrinkage, and nonlinear shrinkage covariance matrix, respectively. If we restrict to cases where the p -value for the performance gain is below 10%, these numbers reduce to 115, 168, and 173, respectively.

Table 4: Annualized net out-of-sample utility in empirical data (in percentage points)

Port- folio strat.	Asset select. rule	Sample covariance matrix							Linear shrinkage covariance matrix						
		Dataset							Dataset						
		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO
2F	All	-0.31	-59.6	-15.3	99.4	-18.8	56.8	-5.47	5.11	1.20	1.09	41.6	1.09	17.7	0.16
2F	Rand	12.3 [●]	14.2 [●]	-7.17 [●]	145 [●]	1.46 [●]	149 [●]	-1.77 [●]	17.0 [●]	52.4 [●]	1.60	132 [●]	12.8 [●]	79.5 [●]	3.48 [●]
2F	MaxSR	1.38	11.6 [●]	-7.48	23.6 [○]	-10.5	-27.1 [○]	7.75 [●]	1.93 [○]	23.1 [●]	-6.10	23.1 [●]	0.69	10.8 [○]	12.2 [●]
2F	MinSR	25.4 [●]	73.2 [●]	7.10 [●]	146	24.9 [●]	139 [●]	-3.86 [●]	18.0 [●]	53.2 [●]	5.21 [●]	101 [●]	16.0 [●]	72.1 [●]	-2.49 [●]
2F	BWSR	7.39	84.4 [●]	-4.42	133	11.7 [○]	242 [●]	-29.9 [●]	19.2 [●]	63.8 [●]	4.74	143 [●]	18.8 [●]	99.5 [●]	-11.8 [●]
2F	MaxVar	25.0 [●]	124 [●]	17.9 [●]	114	21.7 [●]	83.8	-7.25 [●]	23.2 [●]	70.5 [●]	9.99 [●]	122 [●]	21.1 [●]	77.7 [●]	-3.62 [●]
2F	MinVar	2.47	12.5 [●]	-13.1	219 [●]	-1.88	225 [●]	-0.42 [●]	6.08	33.1 [●]	-5.25 [○]	140 [●]	5.40 [○]	81.4 [●]	12.4 [●]
2F	BWVar	13.0	89.3 [●]	-6.76	182 [○]	0.14	180 [●]	-0.53 [●]	19.8 [●]	63.4 [●]	5.39	133 [●]	10.6 [●]	89.9 [●]	5.38 [●]
2F	MinPC	27.4 [●]	112 [●]	-6.23	115	17.9 [●]	94.1	-1.01 [●]	24.9 [●]	67.7 [●]	-2.55	109 [●]	22.8 [●]	50.5 [●]	6.68 [●]
2F	Best θ_N^2	-1.22	14.7 [●]	-8.77	182 [○]	4.73	137 [●]	-9.71 [●]	24.4 [●]	59.2 [●]	-4.03	149 [●]	23.0 [●]	95.7 [●]	3.33 [●]
2F	MaxW	-127 [●]	-305 [●]	-179 [●]	-581 [●]	-129 [●]	-366 [●]	-77.2 [●]	32.3 [●]	79.0 [●]	2.19	197 [●]	19.8 [○]	129 [●]	-13.5 [●]
3FGMV	All	0.47	-60.5	-12.5	99.1	-19.7	55.4	-8.35	5.77	1.21	1.79	41.9	1.31	17.9	1.99
3FGMV	Rand	16.8 [●]	14.5 [●]	2.65 [●]	145 [●]	6.11 [●]	148 [●]	-3.74 [●]	19.8 [●]	52.7 [●]	9.42 [●]	133 [●]	16.6 [●]	80.0 [●]	10.7 [●]
3FGMV	MaxSR	-4.06	10.8 [●]	-1.93 [○]	15.6 [●]	-6.02	-27.4 [●]	13.5 [●]	1.63	23.3 [●]	3.04	24.6 [●]	5.32	10.8 [○]	23.3 [●]
3FGMV	MinSR	36.0 [●]	71.6 [●]	12.9 [●]	150	31.5 [●]	141 [●]	-2.49 [●]	24.3 [●]	53.4 [●]	8.94 [●]	103 [●]	21.5 [●]	74.8 [●]	3.93 [●]
3FGMV	BWSR	10.9	82.4 [●]	1.02	143	19.9 [●]	253 [●]	-7.29	24.7 [●]	64.0 [●]	13.0 [●]	146 [●]	25.2 [●]	101 [●]	11.7 [●]
3FGMV	MaxVar	25.1 [●]	126 [●]	24.9 [●]	117	21.7 [●]	85.0	-13.4 [●]	24.6 [●]	70.9 [●]	14.3 [●]	123 [●]	23.1 [●]	79.6 [●]	-0.62 [●]
3FGMV	MinVar	6.13	13.3 [●]	-14.9	222 [●]	0.35 [○]	219 [●]	9.13 [●]	7.94	33.2 [●]	-0.72	141 [●]	8.69 [●]	80.7 [●]	22.0 [●]
3FGMV	BWVar	27.4 [●]	89.0 [●]	2.74	183 [○]	11.2 [●]	190 [●]	0.08 [●]	25.8 [●]	63.5 [●]	12.9 [●]	137 [●]	15.4 [●]	90.8 [●]	13.4 [●]
3FGMV	MinPC	33.3 [●]	113 [●]	6.47 [●]	115	25.9 [●]	91.3	-10.1	27.0 [●]	68.2 [●]	8.74 [○]	112 [●]	28.7 [●]	52.0 [●]	10.1 [●]
3FGMV	Best θ_N^2	12.3	7.39 [●]	-0.08	122	1.33 [○]	165 [●]	-4.21 [●]	15.1 [○]	48.2 [●]	11.9 [●]	144 [●]	19.6 [●]	71.0 [●]	11.0 [●]
3FGMV	MaxW	-136 [●]	-300 [●]	-147 [●]	-650 [●]	-110 [●]	-325 [●]	-82.9 [●]	38.2 [●]	79.4 [●]	15.7	201 [●]	30.4 [●]	131 [●]	9.34 [●]
3FEW	All	1.83	-62.8	-10.8	102	-16.8	60.8	14.1	7.21	-2.56	3.01	40.3	2.38	18.5	19.1
3FEW	Rand	13.3 [●]	13.6 [●]	-0.44 [●]	146 [●]	2.76 [●]	150 [●]	15.3 [●]	17.1 [●]	51.2 [●]	6.39 [●]	131 [●]	13.7 [●]	79.3 [●]	19.8 [●]
3FEW	MaxSR	3.68	12.6 [●]	-0.61	23.6 [●]	-12.1	-20.4 [●]	17.6 [●]	4.05	23.4 [●]	2.23	22.4 [●]	1.79	12.0	18.1
3FEW	MinSR	24.8 [●]	70.8 [●]	11.3 [●]	145	24.8 [●]	146 [●]	9.96 [●]	17.6 [●]	51.0 [●]	7.15	98.2 [●]	15.3 [●]	72.0 [●]	12.1 [●]
3FEW	BWSR	8.43	84.4 [●]	3.41	135	17.9 [●]	237 [●]	12.6	20.6 [●]	62.6 [●]	9.48 [○]	140 [●]	22.0 [●]	98.8 [●]	19.8
3FEW	MaxVar	26.0 [●]	122 [●]	25.3 [●]	124	16.0 [●]	98.9	8.08 [●]	22.2 [●]	68.7 [●]	12.8 [●]	121 [●]	18.0 [●]	77.0 [●]	13.0 [●]
3FEW	MinVar	4.93	12.3 [●]	-5.87	220 [●]	0.40	230 [●]	19.2 [●]	8.94	33.5 [●]	2.79	140 [●]	8.08 [○]	82.2 [●]	23.5 [●]
3FEW	BWVar	14.2	88.6 [●]	-3.30	195 [○]	-0.15	187 [●]	18.9 [●]	19.0 [●]	62.4 [●]	9.50 [○]	132 [●]	9.65 [○]	90.4 [●]	21.6 [●]
3FEW	MinPC	26.0 [●]	114 [●]	2.26	120	15.9 [●]	103	19.1 [●]	22.1 [●]	66.3 [●]	5.27	113 [●]	21.4 [●]	52.8 [●]	23.9 [●]
3FEW	Best θ_N^2	12.5	4.58 [●]	-4.45	136	-8.83	166 [●]	15.9 [○]	12.5	46.2 [●]	10.1 [●]	140 [●]	13.2 [●]	69.5 [●]	20.5 [○]
3FEW	MaxW	-124 [●]	-305 [●]	-157 [●]	-650 [●]	-139 [●]	-361 [●]	-7.46 [●]	24.8 [●]	52.9 [●]	5.80	202 [●]	0.52	131 [●]	20.4 [○]
EW	All	8.11	2.80	8.00	3.98	7.13	5.61	13.5	8.11	2.80	8.00	3.98	7.13	5.61	13.5
EW	Rand	8.04 [●]	2.73 [●]	7.90 [●]	3.89 [●]	7.04 [●]	5.54 [●]	13.4 [●]	8.04 [●]	2.73 [●]	7.90 [●]	3.89 [●]	7.04 [●]	5.54 [●]	13.4 [●]
EW	MaxSR	8.77 [○]	4.72 [●]	8.49	5.97 [●]	7.40	6.25 [●]	11.2 [●]	8.77 [○]	4.72 [●]	8.49	5.97 [●]	7.40	6.25 [●]	11.2 [●]
EW	MinSR	7.39 [●]	0.83 [●]	7.55 [○]	2.75 [●]	6.77	5.18	14.8 [●]	7.39 [●]	0.83 [●]	7.55 [○]	2.75 [●]	6.77	5.18	14.8 [●]
EW	BWSR	7.94	2.42 [●]	7.72	2.86 [●]	7.35	5.28	14.7 [●]	7.94	2.42 [●]	7.72	2.86 [●]	7.35	5.28 [○]	14.7 [●]
EW	MaxVar	7.95	1.38 [●]	7.97	3.52	6.68	5.31	14.8 [●]	7.95	1.38 [●]	7.97	3.52	6.68	5.31	14.8 [●]
EW	MinVar	8.60	4.23 [●]	8.18	4.36	7.44	5.86	11.2 [●]	8.60	4.23 [●]	8.18	4.36	7.44	5.86	11.2 [●]
EW	BWVar	7.60 [●]	2.34 [●]	7.52 [●]	3.99	6.88	5.58	14.1 [●]	7.60 [●]	2.34 [●]	7.52 [●]	3.99	6.88	5.58	14.1 [●]
EW	MinPC	7.87	2.48 [●]	7.20 [●]	5.00 [●]	6.97	6.22 [●]	12.4 [●]	7.87	2.48 [●]	7.20 [●]	5.00 [●]	6.97	6.22 [●]	12.4 [●]
EW	Best θ_N^2	7.88 [○]	2.72	7.92	3.82	7.02	6.03 [●]	13.5	8.27	2.75	7.82	3.95	6.90	5.92 [○]	13.5
EWRF	All	0.63	-1.55	0.59	-1.84	1.00	0.59	16.4	0.63	-1.55	0.59	-1.84	1.00	0.59	16.4
SGMV		1.76	-49.7	4.25	-0.34	-2.28	-4.67	1.25	10.3	2.33	7.75	4.02	8.48	6.65	6.89
GMVRF		1.81	-26.3	-0.98	-11.4	-12.3	-5.34	-5.70	4.01	-0.00	1.47	0.14	0.66	0.49	2.10
SMV		-4E+05	-3E+11	-1E+06	-6E+07	-8E+08	-8E+07	-9E+04	-1443	-107	-1399	-2360	-1457	-1143	-1844
SMV-ST		15.8 [●]	-108 [●]	11.7 [●]	-579 [●]	-2.91 [●]	0.88 [●]	0.77 [●]	23.9 [●]	67.8	5.13 [●]	-195 [●]	10.4 [●]	16.4 [●]	1.43 [●]
SMV-HT		-35.3 [●]	-302 [●]	-16.8 [●]	-105 [●]	-25.9 [●]	-143 [●]	-7.62 [●]	-29.9 [●]	-52.1	-17.8 [●]	-376 [●]	-33.6 [●]	-349 [●]	-21.1 [●]
F+A		-232	-158	-368	-338	-138	-138	-110	-113	19.7	-227	-214	-113	-37.5	-111

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 11 portfolio strategies in Table 2, the 11 asset selection rules in Table 3, across the seven datasets in Table 1. The table is constructed following the out-of-sample methodology described in Section 5.4. We estimate the portfolios either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p -values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under each selection rule relative to the 'All' selection rule as a benchmark, using the block bootstrap methodology described in Section 5.4. We also report p -values for SMV-HT and SMV-ST, using SMV as benchmark. The symbols ○, ●, and ● indicate that the p -value is less than 10%, 5%, and 1%, respectively.

Figure 8: Difference in net out-of-sample utility relative to investing in all assets



Notes. This figure depicts, for three different portfolio strategies, the difference between the annualized net out-of-sample utility in percentage points obtained when implementing the portfolios on a subset of N assets, where the optimal N is estimated using our theory, using 10 asset selection rules relative to the case where the portfolios are implemented on all M assets. The three portfolio strategies are the two-fund rule (2F, blue), the GMV-three-fund rule (3FGMV, green), and the EW-three-fund rule (3FEW, red). Each panel considers one of the 10 asset selection rules described in Table 3. The figure is constructed following the methodology described in Section 5.4. We consider seven datasets described in Table 1. We estimate the portfolios with the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

individual stocks. However, the MinVar and BWVar rules based on the variance, as well as the Rand rule, consistently outperform selecting all M assets. We depict these findings in Figure 8 which shows, for the shrinkage covariance matrix, the difference between the annualized net out-of-sample utility of 2F, 3FGMV, and 3FEW implemented on a subset of $\hat{N}^*$ assets versus all M assets. Having addressed our main objective, we now discuss other findings from Table 4.

Second, we turn to the EW portfolio. Under the random asset selection rule, reducing N from M to $T/2$ is detrimental. This is because the EW portfolio faces no estimation risk, and thus, benefits from the diversification gains coming from a larger N . However, when the selection of the N assets is done using a sensible rule, such as MaxSR or MinVar, reducing N often improve the EU of the EW portfolio by removing undesirable assets. We also observe that when $N = M$, it is difficult for the 2F, 3FGMV, and 3FEW portfolios to outperform the EW portfolio. In particular,

EW outperforms in five out of seven datasets. However, when reducing the portfolio size to $N = \hat{N}^*$ with our theory, there are a number of asset selection rules for which 2F, 3FGMV, and 3FEW systematically outperform EW, particularly when we shrink the covariance matrix, no matter if EW is implemented on all M assets or on $T/2$ assets with the same asset selection rule. This key observation highlights the practical importance of our theoretically guided optimal portfolio size.

Third, for the six datasets of characteristic and industry portfolios, applying the MinSR and MaxVar asset selection rules to the 2F, 3FGMV, and 3FEW strategies delivers consistently positive utilities almost systematically larger than those of the seven other benchmarks. Focusing on the shrinkage covariance matrix, the BWSR, BWVar, MinPC, and Best θ_N^2 selection rules also consistently deliver positive utilities generally above the benchmarks. Turning to the 100STO dataset, under the shrinkage covariance matrix, the MaxSR, MinVar, and BWVar selection rules yield consistently positive net utilities generally above the benchmarks. This is particularly true when applying these selection rules to 3FGMV and 3FEW because, for individual stocks, adding an exposure to SGMV or EW is valuable given the lower predictability in expected returns. These results are remarkable because the EW portfolio, in particular, is difficult to outperform with mean-variance portfolios when M/T is large, which is the case in our datasets where M/T ranges from 0.78 to 0.9. These results are due to our optimal portfolio size because, under the ‘All’ selection rule, the 2F, 3FGMV, and 3FEW portfolios underperform the EW portfolio in five out of seven datasets.

Fourth, we turn to the Sharpe ratio selection rules (MaxSR, MinSR, BWSR). Consider first the six datasets of characteristic and industry portfolios. For the EW portfolio, MaxSR consistently outperforms selecting all M assets. In contrast, MinSR and BWSR underperform. Specifically, on average across these six datasets, the EW portfolio of all assets delivers an annualized net out-of-sample utility of 5.94 percentage points, versus 6.93, 5.08, and 5.60 under MaxSR, MinSR, and

BWSR, respectively. These results indicate momentum in characteristic and industry portfolios as an EW portfolio of assets with maximum in-sample Sharpe ratios outperforms the full EW portfolio. In contrast, for the 100STO dataset, it is the MinSR and BWSR selection rules that deliver a better EW portfolio than that applied to all assets, indicating reversal in individual stocks.

Although MaxSR works well among characteristic and industry portfolios for constructing an EW portfolio, we find the opposite for the 2F, 3FGMV, and 3FEW portfolios. Specifically, MinSR consistently outperforms MaxSR and selecting all M assets, and MaxSR is one the worst selection rules. These results mean that when we use optimized weights, the performance gains do not originate from selecting assets with the best marginal performance, but assets that can be best exploited by the portfolio, e.g., with short positions for the MinSR rule. Finally, BWSR, which capitalizes on long and short positions by selecting both good and bad assets, achieves its objective as it consistently outperforms MaxSR and MinSR under the shrinkage covariance matrix.³¹

Fifth, we turn to the variance asset selection rules (MaxVar, MinVar, BWVar). Consider first the six datasets of characteristic and industry portfolios. For the EW portfolio, MinVar consistently outperforms selecting all M assets. Specifically, on average across the six datasets, the EW portfolio of all assets delivers an annualized net out-of-sample utility of 5.94 percentage points, versus 5.47, 6.44, and 5.65 under MaxVar, MinVar, and BWVar, respectively. This finding indicates a low-risk anomaly in these six datasets. However, as noted above, this does not necessarily mean that MinVar is also best for the 2F, 3FGMV, and 3FEW strategies. Indeed, Figure 8 shows that for these portfolios, MinVar delivers a worse performance overall than MaxVar and BWVar. This can be explained because assets with maximum variances selected by MaxVar and BWVar often have

³¹However, BWSR often underperforms MinSR under the sample covariance matrix because, in that case, the long-short positions are too extreme, which increases turnover, and thus, hurts the net performance.

low average returns as the low-risk anomaly predicts, which the 2F, 3FGMV, and 3FEW portfolios can leverage using small or short positions.³² The insights drawn from the 100STO dataset are similar, with the asset selection rules favored when considering the EW portfolio, i.e., MaxVar and BWVar, underperforming the MinVar rule when turning to 2F, 3FGMV, and 3FEW.

Sixth, the MinPC asset selection rule, which accounts for asset dependence by selecting those least correlated with the first PC, performs consistently well. In particular, under the shrinkage covariance matrix, the resulting 2F, 3FGMV, and 3FEW portfolios have positive net out-of-sample utilities larger than those under the ‘All’ selection rule in all cases but one.

Seventh, the Best θ_N^2 selection rule, which maximizes the portfolio Sharpe ratio, outperforms investing in all assets for the 2F, 3FGMV, and 3FEW portfolios under the shrinkage covariance matrix, in all cases but one.

Eighth, the MaxW selection rule outperforms on average all other rules under the shrinkage covariance matrix. For instance, considering the 3FGMV portfolio, MaxW delivers an annualized net utility of 72.15 percentage points on average across datasets, versus 55.09 for the second-best selection rule, BWSR. However, MaxW underperforms all other selection rules under the sample covariance matrix. This can be explained because, in that case, 2F, 3FGMV, and 3FEW are more subject to estimation risk, and thus, selecting assets with large absolute weights is less desirable.

Finally, we consider the SMV, SMV-ST, SMV-HT, and F+A benchmarks. First, as expected, SMV is by far the worst strategy. Second, F+A improves upon SMV but is the second-worst strategy. This suggests that although the Da et al. (2024) method is optimal under specific assumptions as N and T get large, it does not work as well as competitors in a finite-sample setting, in

³²It seems that the low-risk anomaly is strong enough in the 94IN-NV and 98IN-CHA-NV datasets for the MinVar selection rule to outperform MaxVar and BWVar when applied to the 2F, 3FGMV, and 3FEW portfolios. This can be explained because Moreira and Muir (2017) show that the low-risk anomaly does not apply to the size factor, and this factor is used to construct all datasets except 94IN-NV and 98IN-CHA-NV.

line with Figure 7. Third, SMV-ST, which maximizes the MV utility subject to an L_1 -norm constraint, greatly improves upon SMV and SMV-HT. However, in most cases, it underperforms the size-optimized 2F, 3FGMV, and 3FEW portfolios. Fourth, SMV-HT, which keeps only the assets whose SMV portfolio weights are above a given threshold, outperforms the plain SMV but delivers negative net utilities. In Section OA.7.3 of the Online Appendix, we compare the portfolio size obtained with SMV-ST and SMV-HT to that obtained with our theory, $\hat{N}_{smv}^*$. We find that the three methods set a small N on average, and that our $\hat{N}_{smv}^*$ has much less variability across windows.

6 Conclusion

We offer a novel perspective on the optimal *portfolio size* and challenge the conventional wisdom that investors should invest in as many assets as possible to eliminate idiosyncratic risk. Specifically, because of *parameter uncertainty*, we show that it is optimal to invest in a limited number N of assets. We consider a class of portfolio strategies that consists of different combinations of the sample mean-variance (MV), global-minimum-variance (GMV), and equally weighted (EW) portfolios. Within this class, we derive the optimal N^* in a finite-sample setting that maximizes the expected *out-of-sample* utility under the assumption of equicorrelated i.i.d. multivariate elliptical returns. This N^* strikes a tradeoff between accessing additional investment opportunities and limiting the number of parameters and weights to estimate. For typical levels of equity return correlations, we find that N^* is slightly below *half the sample size*. Our approach is flexible because it disentangles the computation of the optimal N from the choice of which N assets to select.

To test our theory empirically, we propose an estimator $\hat{N}^*$ of N^* and suggest a set of simple, sensible and intrinsically different selection rules to determine which $\hat{N}^*$ assets to select. We show

that our size-optimized portfolios outperform their counterparts that invest in all assets in nearly all cases, across different portfolio strategies, datasets, and asset selection rules, and under transaction costs. Our size-optimized portfolios can also outperform EW and GMV portfolios, which are hard to beat in high dimension, and more sophisticated strategies that build on asymptotic theories.

Overall, our methodology renders portfolio theory valuable in the challenging but practically relevant case where the sample size is close to the size of the investment universe.

Appendix

This Appendix contains three sections. In Section A.1, we present the theory we use to determine the optimal portfolio size for the three-fund rules. In Section A.2, we explain how we estimate the different parameters on which the combination coefficients and the optimal N depend. In Section A.3, we provide the proofs of all theoretical results in the main text.

A.1 Optimal portfolio size for three-fund rules

In Section 3, we study the optimal N for the SMV portfolio and two-fund rule. We now study the optimal N for *three-fund rules*. Specifically, we consider in Section A.1.1 a three-fund rule based on the GMV portfolio, and in Section A.1.2 a three-fund rule based on the EW portfolio.

A.1.1 Three-fund rule with the global minimum-variance portfolio

We include the GMV portfolio as an additional robust portfolio rule to invest in,

$$\mathbf{w}_g = \frac{\boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N}{\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N}. \quad (\text{A1})$$

We introduce the following parameters:

$$\mu_{g,N} = \frac{\boldsymbol{\mu}' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N}{\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N}, \quad \sigma_{g,N}^2 = \frac{1}{\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N}, \quad \lambda_{g,N} = \frac{\mu_{g,N}}{\sigma_{g,N}^2}, \quad \theta_{g,N}^2 = \frac{\mu_{g,N}^2}{\sigma_{g,N}^2}, \quad \psi_{g,N}^2 = \theta_N^2 - \theta_{g,N}^2, \quad (\text{A2})$$

which stand for the mean return, variance, price of risk, squared Sharpe ratio, and inefficiency of the GMV portfolio on N assets. Under parameter uncertainty, the three-fund rule that invests in the SMV portfolio, the sample GMV (SGMV) portfolio, and the risk-free asset is

$$\hat{\mathbf{w}}(\boldsymbol{\alpha}) = \frac{\alpha_1}{\gamma} \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N + \frac{\alpha_2}{\gamma} \hat{\boldsymbol{\Sigma}}_N^{-1} \mathbf{1}_N, \quad (\text{A3})$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{R}^2$ is the vector of combination coefficients. We call $\hat{\mathbf{w}}(\boldsymbol{\alpha})$ the *GMV-three-fund rule*. In the next proposition, we derive the EU of $\hat{\mathbf{w}}(\boldsymbol{\alpha})$ when asset returns are i.i.d. multivariate elliptically distributed, the resulting optimal combination coefficients $\boldsymbol{\alpha}^*$, and which EU they deliver.³³ This Proposition follows, with minor adjustments, from Kan and Lassance (2025, Propositions 7 and 8).

Proposition A.1 *Let $T > N + 4$ and Assumption 2 hold. Then, the expected out-of-sample utility of the GMV-three-fund rule $\hat{\mathbf{w}}(\boldsymbol{\alpha})$ in (A3) is*

$$\begin{aligned} EU(\hat{\mathbf{w}}(\boldsymbol{\alpha})) &= \frac{1}{2\gamma} \frac{T}{T - N - 2} \left[2\kappa_{N,1} (\alpha_1 \theta_N^2 + \alpha_2 \lambda_{g,N}) - \frac{c_N T}{T - N - 2} \right. \\ &\quad \left. \times \left(\alpha_1^2 \left(\kappa_{N,2} \theta_N^2 + \kappa_{N,3} \frac{N}{T} \right) + \frac{\alpha_2^2 \kappa_{N,2}}{\sigma_{g,N}^2} + 2\alpha_1 \alpha_2 \kappa_{N,2} \lambda_{g,N} \right) \right]. \end{aligned} \quad (\text{A4})$$

Moreover, the optimal combination coefficients $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*)$ maximizing (A4) are

$$(\alpha_1^*, \alpha_2^*) = \frac{T - N - 2}{c_N T} \left(\frac{\kappa_{N,1} \psi_{g,N}^2}{\kappa_{N,2} \psi_{g,N}^2 + \kappa_{N,3} \frac{N}{T}}, \frac{\kappa_{N,3}}{\kappa_{N,2}} \times \frac{\kappa_{N,1} \frac{N}{T} \mu_{g,N}}{\kappa_{N,2} \psi_{g,N}^2 + \kappa_{N,3} \frac{N}{T}} \right), \quad (\text{A5})$$

and the resulting expected out-of-sample utility is

$$EU(\hat{\mathbf{w}}(\boldsymbol{\alpha}^*)) = \frac{\kappa_{N,1}^2}{2\gamma c_N} \left(\frac{\theta_N^2 \psi_{g,N}^2 + \frac{\kappa_{N,3}}{\kappa_{N,2}} \frac{N}{T} \theta_{g,N}^2}{\kappa_{N,2} \psi_{g,N}^2 + \kappa_{N,3} \frac{N}{T}} \right). \quad (\text{A6})$$

³³The proofs of all results given in the Appendix are available in Section OA.8 of the Online Appendix.

Proposition A.1 shows that for the optimal GMV-three-fund rule, the optimal N is found by maximizing the EU in (A6). We proceed by expressing θ_N^2 and $\theta_{g,N}^2$ as explicit functions of N by assuming that the covariance matrix Σ_N complies with Assumption 1. This is done in Proposition 1 for θ_N^2 , and we now derive the expression for $\theta_{g,N}^2$.

Proposition A.2 *Under Assumption 1, the squared Sharpe ratio of the GMV portfolio is*

$$\theta_{g,N}^2 = \frac{N}{1-\rho} r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N), \quad r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N) = \frac{(\bar{\lambda}_N - \frac{N\rho}{1-\rho+N\rho} \bar{\theta}_{N,1} \bar{\sigma}_{N,-1})^2}{\bar{\sigma}_{N,-2} - \frac{N\rho}{1-\rho+N\rho} \bar{\sigma}_{N,-1}^2}, \quad (\text{A7})$$

where $\bar{\theta}_{N,1}$ is defined in Proposition 1, $\bar{\sigma}_N = (\bar{\sigma}_{N,-1}, \bar{\sigma}_{N,-2})$, $\bar{\lambda}_N = \frac{1}{N} \sum_{i=1}^N \lambda_i$, and $\bar{\sigma}_{N,k} = \frac{1}{N} \sum_{i=1}^N \sigma_i^k$, with $\lambda_i = \mu_i / \sigma_i^2$ the price of risk of asset i .

Using Propositions 1 and A.2, the EU of the GMV-three-fund rule in (A6) becomes

$$\begin{aligned} & EU(\hat{\mathbf{w}}(\boldsymbol{\alpha}^*)) \\ &= \frac{N\kappa_{N,1}^2}{2\gamma c_N(1-\rho)} \frac{\delta_N(\bar{\boldsymbol{\theta}}_N) (\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N)) + \frac{\kappa_{N,3}}{\kappa_{N,2}} \frac{1-\rho}{T} r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N)}{\kappa_{N,2} (\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N)) + \kappa_{N,3} \frac{1-\rho}{T}} \end{aligned} \quad (\text{A8})$$

$$=: EU_{3f,g}(N, T, \rho, \tau, \bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N). \quad (\text{A9})$$

Then, as before, we replace $(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\sigma}_N)$ by $(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\sigma}_M)$ and find the optimal N as

$$N_{3f,g}^* = \underset{N \in \{1, \dots, \min(M, T-5)\}}{\operatorname{argmax}} EU_{3f,g}(N, T, \rho, \tau, \bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\sigma}_M). \quad (\text{A10})$$

Note that, similar to the two-fund rule, the objective function $EU_{3f,g}$ is proportional to N/c_N , which as shown in (19)–(20) is maximized by N slightly below $T/2$.³⁴ The simulations in Section 4 show that $N_{3f,g}^*$ is similar to N_{2f}^* and is slightly below $T/2$ when ρ is not too small.

³⁴That the EU of both the optimal two-fund rule and the GMV-three-fund rule is proportional to N/c_N can be explained as follows: the proportionality to $1/c_N$ is because both the SMV portfolio and the SGMV portfolio are proportional to $\hat{\Sigma}_N^{-1}$, and the proportionality to N is because both the squared Sharpe ratio of the MV portfolio (θ_N^2) and the GMV portfolio ($\theta_{g,N}^2$) are proportional to N under Assumption 1.

A.1.2 Three-fund rule with the equally weighted portfolio

We now consider the EW portfolio as an additional portfolio rule to invest in, $\mathbf{w}_{ew} = \mathbf{1}_N/N$. We introduce the following parameters similar to the case with the GMV-three-fund rule:

$$\mu_{ew,N} = \mathbf{w}_{ew}' \boldsymbol{\mu}_N, \sigma_{ew,N}^2 = \mathbf{w}_{ew}' \boldsymbol{\Sigma}_N \mathbf{w}_{ew}, \lambda_{ew,N} = \frac{\mu_{ew,N}}{\sigma_{ew,N}^2}, \theta_{ew,N}^2 = \frac{\mu_{ew,N}^2}{\sigma_{ew,N}^2}, \psi_{ew,N}^2 = \theta_N^2 - \theta_{ew,N}^2. \quad (\text{A11})$$

Under parameter uncertainty, the three-fund rule that invests in the SMV portfolio, the EW portfolio, and the risk-free asset is

$$\hat{\mathbf{w}}(\boldsymbol{\beta}) = \frac{\beta_1}{\gamma} \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N + \frac{\beta_2}{\gamma} \mathbf{w}_{ew}, \quad (\text{A12})$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbb{R}^2$ is the vector of combination coefficients. We call $\hat{\mathbf{w}}(\boldsymbol{\beta})$ the *EW-three-fund rule*. In the next proposition, we derive the EU of the EW-three-fund rule (A3) when asset returns are i.i.d. multivariate elliptically distributed, the resulting optimal combination coefficients $\boldsymbol{\alpha}^*$, and which EU they deliver. This is a novel result relative to Kan and Lassance (2025) who do not consider the combination of the SMV and EW portfolios.

Proposition A.3 *Let $T > N + 4$ and Assumption 2 hold. Then, the expected out-of-sample utility of the EW-three-fund rule $\hat{\mathbf{w}}(\boldsymbol{\beta})$ in (A12) is*

$$EU(\hat{\mathbf{w}}(\boldsymbol{\beta})) = \frac{1}{2\gamma} \times \left[\frac{2\beta_1 \kappa_{N,1} \theta_N^2 T}{T - N - 2} + 2\beta_2 \mu_{ew,N} - \frac{\beta_1^2 c_N T^2}{(T - N - 2)^2} \left(\kappa_{N,2} \theta_N^2 + \kappa_{N,3} \frac{N}{T} \right) - \beta_2^2 \sigma_{ew,N}^2 - \frac{2\beta_1 \beta_2 \kappa_{N,1} \mu_{ew,N} T}{T - N - 2} \right]. \quad (\text{A13})$$

Moreover, the optimal combination coefficients $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*)$ maximizing (A13) are

$$(\beta_1^*, \beta_2^*) = \left(\frac{T - N - 2}{T} \frac{\kappa_{N,1} \psi_{ew,N}^2}{\kappa_{N,1}^2 \psi_{ew,N}^2 + d_N}, \frac{d_N \lambda_{ew,N}}{\kappa_{N,1}^2 \psi_{ew,N}^2 + d_N} \right), \quad (\text{A14})$$

where $d_N = c_N \kappa_{N,3} \frac{N}{T} + (c_N \kappa_{N,2} - \kappa_{N,1}^2) \theta_N^2$, and the resulting expected out-of-sample utility is

$$EU(\hat{\mathbf{w}}(\boldsymbol{\beta}^*)) = \frac{1}{2\gamma} \left(\frac{\kappa_{N,1}^2 \theta_N^2 \psi_{ew,N}^2 + d_N \theta_{ew,N}^2}{\kappa_{N,1}^2 \psi_{ew,N}^2 + d_N} \right). \quad (\text{A15})$$

The EU of the optimal EW-three-fund rule depends on θ_N^2 and $\theta_{ew,N}^2$. As usual, we express θ_N^2 and $\theta_{ew,N}^2$ as explicit functions of N by assuming that the covariance matrix Σ_N complies with Assumption 1. We derive the expression for $\theta_{ew,N}^2$ in the next proposition.

Proposition A.4 *Under Assumption 1, the squared Sharpe ratio of the EW portfolio is*

$$\theta_{ew,N}^2 = \frac{N}{1-\rho} r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}), \quad r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) = \frac{(1-\rho)\bar{\mu}_N^2}{(1-\rho)\bar{\sigma}_{N,2} + \rho N \bar{\sigma}_{N,1}^2}, \quad (\text{A16})$$

where $\bar{\sigma}_{N,1}$ and $\bar{\sigma}_{N,2}$ are defined in Proposition A.2 and $\bar{\mu}_N = \frac{1}{N} \sum_{i=1}^N \mu_i$.

Using Propositions 1 and A.4, the EU of the EW-three-fund rule in (A15) becomes

$$EU(\hat{\mathbf{w}}(\boldsymbol{\alpha}^*)) = \frac{N}{2\gamma(1-\rho)} \times \frac{\kappa_{N,1}^2 \delta_N(\bar{\boldsymbol{\theta}}_N) \left(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) \right) + \left(\frac{(1-\rho)c_N \kappa_{N,3}}{T} + (c_N \kappa_{N,2} - \kappa_{N,1}^2) \delta_N(\bar{\boldsymbol{\theta}}_N) \right) r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})}{\kappa_{N,1}^2 \left(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) \right) + \left(\frac{(1-\rho)c_N \kappa_{N,3}}{T} + (c_N \kappa_{N,2} - \kappa_{N,1}^2) \delta_N(\bar{\boldsymbol{\theta}}_N) \right)} \quad (\text{A17})$$

$$=: EU_{3f,ew}(N, T, \rho, \tau, \bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}). \quad (\text{A18})$$

Finally, we replace $(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})$ by $(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ and find the optimal N as

$$N_{3f,ew}^* = \underset{N \in \{1, \dots, \min(M, T-5)\}}{\operatorname{argmax}} EU_{3f,ew}(N, T, \rho, \tau, \bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2}). \quad (\text{A19})$$

Although it does not appear as clearly as for the two-fund rule and the GMV-three-fund rule via N/c_N as a proportionality factor in the objective function, the simulations in Section 4 also show that $N_{3f,ew}^*$ typically hovers around $T/2$ for ρ not too small.

A.2 Estimation of parameters

In this section, we explain how we estimate the different parameters that are needed as inputs in our different portfolio rules to determine the optimal portfolio size (i.e., N_{smv}^* , N_{2f}^* , $N_{3f,g}^*$, $N_{3f,ew}^*$) and the optimal combination coefficients (i.e., α^* , α_1^* , α_2^* , β_1^* , β_2^*).

A.2.1 Estimation of elliptical fat tails

The first set of parameters are those that determine the impact of the fat tails of the elliptical distribution, i.e., $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$ in (10)–(12). To speed up the computation, we use the high-dimensional approximation of these parameters. Specifically, from El Karoui (2010, 2013), we have that as $N, T \rightarrow \infty$ and $N/T \rightarrow \phi \in (0, 1)$, $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3}) \rightarrow (\tilde{\kappa}_{N,1}, \tilde{\kappa}_{N,2}, \tilde{\kappa}_{N,1})$, where $\tilde{\kappa}_{N,1} \geq 1$ is the unique positive solution to

$$\mathbb{E} \left[(1 - \phi + \phi \tilde{\kappa}_{N,1} \tau)^{-1} \right] = 1, \quad (\text{A20})$$

and $\tilde{\kappa}_{N,2} \geq \tilde{\kappa}_{N,1}^2$ is given by

$$\tilde{\kappa}_{N,2} = (1 - \phi) \left(\tilde{\kappa}_{N,1}^{-2} - \mathbb{E} \left[\frac{\phi \tau^2}{(1 - \phi + \phi \tilde{\kappa}_{N,1} \tau)^2} \right] \right)^{-1}. \quad (\text{A21})$$

Kan and Lassance (2025) show that this high-dimensional approximation is accurate for typical values of N and T .

Given this result, we need to estimate $\tilde{\kappa}_{N,1}$ and $\tilde{\kappa}_{N,2}$. We estimate them in two different ways as in Kan and Lassance (2025). First, we assume that the asset returns follow a multivariate t -distribution, i.e., $\tau \sim (\nu-2)/\chi_\nu^2$, and we estimate the number of degrees of freedom ν by maximum likelihood from a sample of T historical returns $(\mathbf{r}_1, \dots, \mathbf{r}_T)$, giving us $\hat{\nu}$. Then, we can use the

closed-form expression for $\tilde{\kappa}_{N,1}$ and $\tilde{\kappa}_{N,2}$ when asset returns are t -distributed in Kan and Lassance (2025, Proposition 6), which yields that the estimate of $\tilde{\kappa}_{N,1}$, denoted $\tilde{\kappa}_{N,1}^\nu$, is the unique positive solution to

$$ye^y E_{\hat{\nu}/2}(y) = \phi_N \quad \text{with} \quad y = \frac{(\hat{\nu} - 2)\phi_N \tilde{\kappa}_{N,1}^\nu}{2(1 - \phi_N)} \quad \text{and} \quad \phi_N = \frac{N}{T}, \quad (\text{A22})$$

where $E_n(x) = \int_1^\infty t^{-n} e^{-xt} dt$ is the exponential integral, and the estimate of $\tilde{\kappa}_{N,2}$ is

$$\tilde{\kappa}_{N,2}^\nu = \frac{2(\tilde{\kappa}_{N,1}^\nu)^2(1 - \phi_N)}{\hat{\nu} - \tilde{\kappa}_{N,1}^\nu(\hat{\nu} - 2)}. \quad (\text{A23})$$

We use this estimation method in Section 4 where we simulate returns from a t -distribution.

The second method we use to estimate $\tilde{\kappa}_{N,1}$ and $\tilde{\kappa}_{N,2}$ relies on the following sample estimate of the distribution of τ proposed by El Karoui (2010, 2013):

$$\hat{\tau}_{N,t} = \frac{(\mathbf{r}_t - \hat{\boldsymbol{\mu}}_N)'(\mathbf{r}_t - \hat{\boldsymbol{\mu}}_N)}{\frac{1}{T} \sum_{i=1}^T (\mathbf{r}_i - \hat{\boldsymbol{\mu}}_N)'(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_N)}, \quad t = 1, \dots, T, \quad (\text{A24})$$

which is consistent as $N \rightarrow \infty$. Using $\hat{\tau}_{N,t}$, we estimate $\tilde{\kappa}_{N,1}$ and $\tilde{\kappa}_{N,2}$ with their sample counterparts, i.e., the estimate of $\tilde{\kappa}_{N,1}$, denoted $\tilde{\kappa}_{N,1}^s$, is the unique positive solution to

$$\frac{1}{T} \sum_{t=1}^T (1 - \phi_N + \phi_N \tilde{\kappa}_{N,1}^s \hat{\tau}_{N,t})^{-1} = 1, \quad (\text{A25})$$

and the estimate of $\tilde{\kappa}_{N,2}$ is

$$\tilde{\kappa}_{N,2}^s = (1 - \phi_N) \left((\tilde{\kappa}_{N,1}^s)^{-2} - \frac{1}{T} \sum_{t=1}^T \frac{\phi_N \hat{\tau}_{N,t}^2}{(1 - \phi_N + \phi_N \tilde{\kappa}_{N,1}^s \hat{\tau}_{N,t})^2} \right)^{-1}. \quad (\text{A26})$$

We use this second estimation method in Section 5 to have more freedom in describing the tails of empirical data that may not be t -distributed.

A.2.2 Estimation of portfolios' performance

The second set of parameters are those that determine the performance of the MV, GMV, and EW portfolios and on which the optimal combination coefficients depend: θ_N^2 in (2), $\psi_{g,N}^2$ in (A2),

$\psi_{ew,N}^2$ in (A11), $\mu_{g,N}$ in (A2), and $\lambda_{ew,N}$ in (A11).

For $\mu_{g,N}$ and $\lambda_{ew,N}$, we rely on the estimates that are unbiased when asset returns are i.i.d. multivariate normally distributed³⁵, i.e.,

$$\hat{\mu}_{g,N} = \frac{\mathbf{1}'_N \hat{\Sigma}_N^{-1} \hat{\boldsymbol{\mu}}_N}{\mathbf{1}'_N \hat{\Sigma}_N^{-1} \mathbf{1}_N}, \quad (\text{A27})$$

$$\hat{\lambda}_{ew,N} = \frac{T-3}{T} \frac{\mathbf{w}'_{ew} \hat{\boldsymbol{\mu}}_N}{\mathbf{w}'_{ew} \hat{\Sigma}_N \mathbf{w}_{ew}}. \quad (\text{A28})$$

For θ_N^2 , $\psi_{g,N}^2$, and $\psi_{ew,N}^2$, the sample estimates, obtained by plugging $(\hat{\boldsymbol{\mu}}_N, \hat{\Sigma}_N)$, are severely biased. Therefore, we estimate them using the adjusted estimates in Kan and Zhou (2007) and Kan and Wang (2023) that correct the unbiased estimates to ensure they are positive. Specifically, let $\hat{\theta}_N^2 = \hat{\boldsymbol{\mu}}'_N \hat{\Sigma}_N^{-1} \hat{\boldsymbol{\mu}}_N$, $\hat{\psi}_{g,N}^2 = \hat{\theta}_N^2 - \hat{\theta}_{g,N}^2$, and $\hat{\psi}_{ew,N}^2 = \hat{\theta}_N^2 - \hat{\theta}_{ew,N}^2$ be the sample estimates of θ_N^2 , $\psi_{g,N}^2$, and $\psi_{ew,N}^2$, where $\hat{\theta}_{g,N}^2 = (\mathbf{1}'_N \hat{\Sigma}_N^{-1} \hat{\boldsymbol{\mu}}_N)^2 / (\mathbf{1}'_N \hat{\Sigma}_N^{-1} \mathbf{1}_N)$ and $\hat{\theta}_{ew,N}^2 = (\mathbf{w}'_{ew} \hat{\boldsymbol{\mu}}_N)^2 / (\mathbf{w}'_{ew} \hat{\Sigma}_N \mathbf{w}_{ew})$.

Then, the adjusted estimates are

$$\hat{\theta}_{N,a}^2 = \frac{(T-N-2)\hat{\theta}_N^2 - N}{T} + \frac{2(\hat{\theta}_N^2)^{\frac{N}{2}}(1+\hat{\theta}_N^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\theta}_N^2/(1+\hat{\theta}_N^2)}\left(\frac{N}{2}, \frac{T-N}{2}\right)}, \quad (\text{A29})$$

$$\hat{\psi}_{g,N,a}^2 = \frac{(T-N-1)\hat{\psi}_{g,N}^2 - (N-1)}{T} + \frac{2(\hat{\psi}_{g,N}^2)^{\frac{N-1}{2}}(1+\hat{\psi}_{g,N}^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\psi}_{g,N}^2/(1+\hat{\psi}_{g,N}^2)}\left(\frac{N-1}{2}, \frac{T-N+1}{2}\right)}, \quad (\text{A30})$$

$$\hat{\psi}_{ew,N,a}^2 = \frac{(T-N-2)\hat{\psi}_{ew,N}^2 - (N-1)(1+\hat{\theta}_{ew,N}^2)}{T} + \frac{2(1+\hat{\theta}_{ew,N}^2)^{\frac{T-N}{2}}(\hat{\psi}_{ew,N}^2)^{\frac{N-1}{2}}(1+\hat{\theta}_N^2)^{\frac{3-T}{2}}}{T \times B_{\hat{\psi}_{ew,N}^2/(1+\hat{\theta}_N^2)}\left(\frac{N-1}{2}, \frac{T-N}{2}\right)}, \quad (\text{A31})$$

where $B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ is the incomplete beta function.

A.2.3 Estimation of assets' correlation and marginal performance

The third and last set of parameters are those that control the dependence between the assets, i.e., ρ under Assumption 1, and the marginal performance of the assets, i.e., the three functions $\delta_N(\bar{\boldsymbol{\theta}}_M)$

³⁵When asset returns are i.i.d. multivariate elliptically distributed as in Assumption 2, $\hat{\mu}_{g,N}$ and $\hat{\lambda}_{ew,N}$ are also unbiased in the high-dimensional asymptotic regime as $N, T \rightarrow \infty$ and $N/T \rightarrow \phi \in (0, 1)$.

in (4), $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\sigma}_M)$ in (A7), and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ in (A16). Recall that the parameters in these functions are computed on all M assets to find the optimal N . Following the advice of Adams et al. (2017), we use an estimator $\hat{\rho}$ given by the average of all sample correlations $\hat{\rho}_{ij}$ obtained from the sample covariance matrix $\hat{\Sigma}_M$,

$$\hat{\rho} = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \hat{\rho}_{ij}. \quad (\text{A32})$$

Regarding the functions $\delta_N(\bar{\theta}_M)$, $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\sigma}_M)$, and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$, we rely on the following proposition.

Proposition A.5 *Let the covariance matrix Σ_M be known and satisfy Assumption 1. Then, the biases of the estimators of $\delta_N(\bar{\theta}_M)$, $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\sigma}_M)$, and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ obtained by plugging the sample mean $\hat{\mu}_M$ are*

$$\mathbb{E}[\hat{\delta}_N] - \delta_N = \frac{1-\rho}{T} \times \frac{1 - \frac{N}{M}\rho + N\rho}{1 - \rho + N\rho}, \quad (\text{A33})$$

$$\mathbb{E}[\hat{r}_{g,N}] - r_{g,N} = \frac{1-\rho}{MT} \times \frac{\bar{\sigma}_{M,-2} - \frac{N\rho}{1-\rho+N\rho} \left(1 + \frac{(1-\rho)(1-\frac{M}{N})}{1-\rho+N\rho}\right) \bar{\sigma}_{M,-1}^2}{\bar{\sigma}_{M,-2} - \frac{N\rho}{1-\rho+N\rho} \bar{\sigma}_{M,-1}^2}, \quad (\text{A34})$$

$$\mathbb{E}[\hat{r}_{ew,N}] - r_{ew,N} = \frac{1-\rho}{MT} \times \frac{\bar{\sigma}_{M,2} + \frac{\rho}{1-\rho} M \bar{\sigma}_{M,1}^2}{\bar{\sigma}_{M,2} + \frac{\rho}{1-\rho} N \bar{\sigma}_{M,1}^2}. \quad (\text{A35})$$

Building on Proposition A.5, we estimate the function $\delta_N(\bar{\theta}_M)$, $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\sigma}_M)$, and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ by plugging the sample mean $\hat{\mu}_M$ and the sample covariance matrix $\hat{\Sigma}_M$ and by removing the bias, which we estimate from $\hat{\mu}_M$ and $\hat{\Sigma}_M$ too.

A.3 Proofs of results in the main text

In this section, we provide the proofs for all theoretical results given in the main text. Proofs for the theoretical results given in this Appendix are available in the Online Appendix.

A.3.1 Proof of Proposition 1

Denoting $\mathbf{D}_N = \text{diag}(\sigma_1, \dots, \sigma_N)$, $\Sigma_N = \mathbf{D}_N \mathbf{P}_N(\rho) \mathbf{D}_N$ and its inverse is given by

$$\Sigma_N^{-1} = \mathbf{D}_N^{-1} \mathbf{P}_N(\rho)^{-1} \mathbf{D}_N^{-1}, \quad (\text{A36})$$

where $\mathbf{D}_N^{-1} = \text{diag}(1/\sigma_1, \dots, 1/\sigma_N)$ and $\mathbf{P}_N(\rho)^{-1}$ exists if and only if $\rho \in (-\frac{1}{N-1}, 1)$ and is equal to

$$\mathbf{P}_N(\rho)^{-1} = \frac{1}{1-\rho} \left(\mathbf{I}_N - \frac{\rho}{1-\rho+N\rho} \mathbf{1}_N \mathbf{1}_N' \right). \quad (\text{A37})$$

Combining (A36) and (A37) yields

$$\Sigma_N^{-1} = \frac{1}{1-\rho} \left(\mathbf{D}_N^{-2} - \frac{\rho}{1-\rho+N\rho} \mathbf{D}_N^{-1} \mathbf{1}_N \mathbf{1}_N' \mathbf{D}_N^{-1} \right), \quad (\text{A38})$$

and thus the maximum utility becomes

$$U(\mathbf{w}^*) = \frac{\boldsymbol{\mu}_N' \Sigma_N^{-1} \boldsymbol{\mu}_N}{2\gamma} = \frac{1}{2\gamma(1-\rho)} \left[\boldsymbol{\mu}_N' \mathbf{D}_N^{-2} \boldsymbol{\mu}_N - \frac{\rho}{1-\rho+N\rho} (\boldsymbol{\mu}_N' \mathbf{D}_N^{-1} \mathbf{1}_N)^2 \right], \quad (\text{A39})$$

where $\boldsymbol{\mu}_N' \mathbf{D}_N^{-2} \boldsymbol{\mu}_N = N\bar{\theta}_{N,2}$ and $\boldsymbol{\mu}_N' \mathbf{D}_N^{-1} \mathbf{1}_N = N\bar{\theta}_{N,1}$, which yields the desired result in (4).

We then study how $U(\mathbf{w}^*)$ increases with N , which amounts to studying the difference $\theta_{N+1}^2 - \theta_N^2$. We have

$$\begin{aligned} & (1-\rho)(\theta_{N+1}^2 - \theta_N^2) \\ &= \left[\sum_{i=1}^{N+1} s_i^2 - \frac{\rho}{1-\rho+(N+1)\rho} \left(\sum_{i=1}^{N+1} s_i \right)^2 \right] - \left[\sum_{i=1}^N s_i^2 - \frac{\rho}{1-\rho+N\rho} \left(\sum_{i=1}^N s_i \right)^2 \right]. \quad (\text{A40}) \end{aligned}$$

Decomposing $\left(\sum_{i=1}^{N+1} s_i \right)^2$ as $\left(\sum_{i=1}^N s_i \right)^2 + s_{N+1}^2 + 2s_{N+1} \sum_{i=1}^N s_i$, (A40) becomes

$$(1-\rho)(\theta_{N+1}^2 - \theta_N^2) = \frac{1-\rho+N\rho}{1-\rho+(N+1)\rho} s_{N+1}^2 + \frac{\rho^2 \left(\sum_{i=1}^N s_i \right)^2}{(1-\rho+N\rho)(1-\rho+(N+1)\rho)}$$

$$\begin{aligned}
& - \frac{2\rho s_{N+1} \sum_{i=1}^N s_i}{1 - \rho + (N+1)\rho} \\
& = \frac{\left[\rho N \bar{\theta}_{N,1} - (1 - \rho + N\rho) s_{N+1} \right]^2}{(1 - \rho + N\rho)(1 - \rho + (N+1)\rho)},
\end{aligned} \tag{A41}$$

which is nonnegative, and strictly positive if and only if $s_{N+1} \neq \rho N \bar{\theta}_{N,1} / (1 - \rho + N\rho)$. This concludes the proof.

A.3.2 Proof of Proposition 2

Equation (9) is a direct extension of Kan and Lassance (2025, Proposition 7) for a general α instead of $\alpha = 1$. It is then easy to show that the α maximizing (9) is equal to (13), which also corresponds to Kan and Lassance (2025, Equation (50)). Finally, after some developments, plugging (13) into (9) yields the EU in (13), which concludes the proof.

A.3.3 Proof of Corollary 1

Under Assumption 1, the maximum squared Sharpe ratio θ_N^2 is given by (4). Plugging it into (9) yields the EU of the SMV portfolio $\hat{w}^*$ in Equation (14), and plugging it into (13) yields the EU of the optimal two-fund rule $\hat{w}(\alpha^*)$ in (16).

References

- Adams, Z.; R. Füss; and T. Glück. “Are correlations constant? Empirical and theoretical results on popular correlation models in finance.” *Journal of Banking and Finance*, 84 (2017), 9–24.
- Ang, A.; R. Hodrick; Y. Xing; and X. Zhang. “The cross-section of volatility and expected returns.” *Journal of Finance*, 61 (2006), 259–299.
- Ao, M.; Y. Li; and X. Zheng. “Approaching mean-variance efficiency for large portfolios.” *Review of Financial Studies*, 32 (2019), 2890–2919.
- Barroso, P., and K. Saxena. “Lest we forget: Using out-of-sample forecast errors in portfolio optimization.” *Review of Financial Studies*, 35 (2021), 1222–1278.

- Boyle, P.; L. Garlappi; R. Uppal; and T. Wang. “Keynes meets Markowitz: The trade-off between familiarity and diversification.” *Management Science*, 58 (2012), 253–272.
- Calvet, L.; J. Campbell; and P. Sodini. “Down or out: Assessing the welfare costs of household investment mistakes.” *Journal of Political Economy*, 115 (2007), 707–747.
- Campbell, J. “Household finance.” *Journal of Finance*, 61 (2006), 1553–1604.
- Campbell, J.; A. Lo; and A. Mackinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ (1997).
- Chamberlain, G. “A characterization of the distributions that imply mean-variance utility functions.” *Journal of Economic Theory*, 29 (1983), 185–201.
- Chung, M.; Y. Lee; J. H. Kim; W. C. Kim; and F. Fabozzi. “The effects of errors in means, variances, and correlations on the mean-variance framework.” *Quantitative Finance*, 22 (2022), 1893–1903.
- Clements, A.; A. Scott; and A. Silvennoinen. “On the benefits of equicorrelation for portfolio allocation.” *Journal of Forecasting*, 34 (2015), 507–522.
- Da, R.; S. Nagel; and D. Xiu. “The statistical limit of arbitrage.” *NBER working paper*, 33070 (2024).
- DeMiguel, V.; L. Garlappi; F. Nogales; and R. Uppal. “A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms.” *Management Science*, 55 (2009a), 798–812.
- DeMiguel, V.; L. Garlappi; and R. Uppal. “Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?” *Review of Financial Studies*, 22 (2009b), 1915–1953.
- DeMiguel, V.; A. Martín-Utrera; and F. Nogales. “Parameter uncertainty in multiperiod portfolio optimization with transaction costs.” *Journal of Financial and Quantitative Analysis*, 50 (2015), 1443–1471.
- Du, J.-H.; Y. Guo; and X. Wang. “High-dimensional portfolio selection with cardinality constraints.” *Journal of the American Statistical Association*, 118 (2023), 779–791.
- El Karoui, N. “High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation.” *Annals of Statistics*, 38 (2010), 3487–3566.
- El Karoui, N. “On the realized risk of high-dimensional Markowitz portfolios.” *SIAM Journal on Financial Mathematics*, 4 (2013), 737–783.
- Elton, E., and M. Gruber. “Estimating the dependence structure of share prices—Implications for portfolio selection.” *Journal of Finance*, 28 (1973), 1203–1232.
- Engle, R.; R. Ferstenberg; and J. Russell. “Measuring and modeling execution cost and risk.” *Journal of Portfolio Management*, 38 (2012), 14–28.

- Engle, R., and B. Kelly. “Dynamic equicorrelation.” *Journal of Business & Economic Statistics*, 30 (2012), 212–228.
- Fan, J.; J. Zhang; and K. Yu. “Vast portfolio selection with gross-exposure constraints.” *Journal of the American Statistical Association*, 107 (2012), 592–606.
- Frazzini, A.; R. Israel; and T. J. Moskowitz. “Trading costs.” *SSRN working paper*, (2018).
- Gandy, A., and L. Veraart. “The effect of estimation in high-dimensional portfolios.” *Mathematical Finance*, 23 (2013), 531–559.
- Gao, J., and D. Li. “Optimal cardinality constrained portfolio selection.” *Operations Research*, 61 (2013), 745–761.
- Goetzmann, W., and A. Kumar. “Equity portfolio diversification.” *Review of Finance*, 12 (2008), 433–446.
- Kan, R., and N. Lassance. “Optimal portfolio choice with fat tails and parameter uncertainty.” *Journal of Financial and Quantitative Analysis*, forthcoming (2025).
- Kan, R., and X. Wang. “Optimal portfolio choice with unknown benchmark efficiency.” *Management Science*, 70 (2023), 6117–6138.
- Kan, R.; X. Wang; and G. Zhou. “Optimal portfolio choice with estimation risk: No risk-free asset case.” *Management Science*, 68 (2021), 2047–2068.
- Kan, R., and G. Zhou. “Optimal portfolio choice with parameter uncertainty.” *Journal of Financial and Quantitative Analysis*, 42 (2007), 621–656.
- Koijen, R., and M. Yogo. “A demand system approach to asset pricing.” *Journal of Political Economy*, 127 (2019), 1475–1515.
- Lassance, N., and A. Martín-Utrera. “Sentiment-based portfolios.” *SSRN working paper*, (2023).
- Lassance, N.; A. Martín-Utrera; and M. Simaan. “The risk of expected utility under parameter uncertainty.” *Management Science*, 70 (2024a), 7644–7663.
- Lassance, N.; R. Vanderveken; and F. Vrans. “On the combination of naive and mean-variance portfolio strategies.” *Journal of Business & Economic Statistics*, 42 (2024b), 875–889.
- Ledoit, O., and M. Wolf. “Honey, I shrunk the sample covariance matrix.” *Journal of Portfolio Management*, 30 (2003), 110–119.
- Ledoit, O., and M. Wolf. “A well-conditioned estimator for large-dimensional covariance matrices.” *Journal of Multivariate Analysis*, 88 (2004), 365–411.
- Ledoit, O., and M. Wolf. “Robust performance hypothesis testing with the Sharpe ratio.” *Journal of Empirical Finance*, 15 (2008), 850–859.
- Ledoit, O., and M. Wolf. “Analytical nonlinear shrinkage of large-dimensional covariance matrices.” *Annals of Statistics*, 48 (2020), 3043–3065.

- Merton, R. “On estimating the expected return on the market: An exploratory investigation.” *Journal of Financial Economics*, 8 (1980), 323–361.
- Moreira, A., and T. Muir. “Volatility-managed portfolios.” *Journal of Finance*, 72 (2017), 1611–1643.
- Novy-Marx, R., and M. Velikov. “A taxonomy of anomalies and their trading costs.” *Review of Financial Studies*, 29 (2015), 104–147.
- Politis, D., and J. Romano. “The stationary bootstrap.” *Journal of the American Statistical Association*, 89 (1994), 1303–1313.
- Schuhmacher, F.; H. Kohrs; and B. Auer. “Justifying mean-variance portfolio selection when asset returns are skewed.” *Management Science*, 67 (2021), 7812–7824.
- Tu, J., and G. Zhou. “Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies.” *Journal of Financial Economics*, 99 (2011), 204–215.
- Yen, Y.-M. “Sparse weighted-norm minimum variance portfolios.” *Review of Finance*, 20 (2016), 1259–1287.
- Yuan, M., and G. Zhou. “Why naïve 1/N diversification is not so naïve, and how to beat it?” *Journal of Financial and Quantitative Analysis*, 59 (2024), 3601–3632.

Online Appendix to

Optimal Portfolio Size
under Parameter Uncertainty

Nathan Lassance Rodolphe Vanderveken Frédéric Vrins

This Online Appendix contains eight sections. In Section OA.1, we study the optimal portfolio size under a single-factor model assumption. In Section OA.2, we analyze the relation between the correlation of assets and the maximum utility of the MV portfolio. In Section OA.3, we study the impact of the sequence of assets on the EU and the optimal portfolio size. In Section OA.4, we detail how we implement the factor-plus-alpha portfolio strategy. In Section OA.5, we explain how we estimate the EWRF portfolio strategy. In Sections OA.6 and OA.7, we report and discuss additional simulation and empirical results. In Section OA.8, we provide the proofs of theoretical results contained in the Appendix and Online Appendix.

OA.1 Single-factor model assumption

In the main text, we find the optimal portfolio size under Assumption 1, i.e., asset returns are equicorrelated. In this section, we show that we obtain a very similar portfolio size under another assumption for the covariance matrix, which is that asset returns follow a single-factor model. This is a commonly used assumption in asset pricing and portfolio selection to obtain parsimonious representations of expected returns and covariance matrices; see, e.g., MacKinlay and Pástor (2000), Tu and Zhou (2011), and Kan et al. (2021).

Assumption OA.1 *The asset returns follow a single-factor model,*

$$\mathbf{r} = \boldsymbol{\alpha}_N + \boldsymbol{\beta}_N f + \boldsymbol{\varepsilon}, \quad (\text{OA1})$$

where $f \in \mathbb{R}$ is the factor with zero mean and variance σ_f^2 , $\boldsymbol{\beta}_N \in \mathbb{R}^N$ is the vector of factor loadings, $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}_N$, $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma_\varepsilon^2 \mathbf{I}_N$, and f is uncorrelated with $\boldsymbol{\varepsilon}$.

Under Assumption OA.1, the covariance matrix of asset returns is of the form

$$\Sigma_N = \sigma_f^2 \beta_N \beta_N' + \sigma_\varepsilon^2 \mathbf{I}_N. \quad (\text{OA2})$$

This parametrization of the covariance matrix involves $N + 2$ parameters: the N factor loadings β_N , σ_f^2 , and σ_ε^2 . This is similar to the covariance matrix under Assumption 1, which involves $N + 1$ parameters: N asset variances and one correlation coefficient. However, the equicorrelation and single-factor assumptions restrict the covariance matrix differently. Assumption 1 fixes the correlation coefficients to a single ρ , while allowing individual variances to vary freely. In contrast, Assumption OA.1 entangles variances and correlations, the variance of asset i being $\sigma_f^2 \beta_i^2 + \sigma_\varepsilon^2$, and the correlation between assets i and j being $\beta_i \beta_j \sigma_f^2$.

Given these differences, if both assumptions result in similar optimal portfolio sizes, this would support the robustness of our approach. Therefore, we proceed by comparing the optimal portfolio sizes under the two assumptions, both the oracle and estimated ones. We focus on the optimal portfolio size for the two-fund rule for conciseness; similar results can be obtained for the SMV portfolio and the three-fund rules. To find the optimal N for the two-fund rule, we derive in the next proposition the maximum squared Sharpe ratio and the EU of the optimal two-fund rule under the single-factor covariance matrix in (OA2).

Proposition OA.1 *Under Assumption OA.1, the maximum squared Sharpe ratio is*

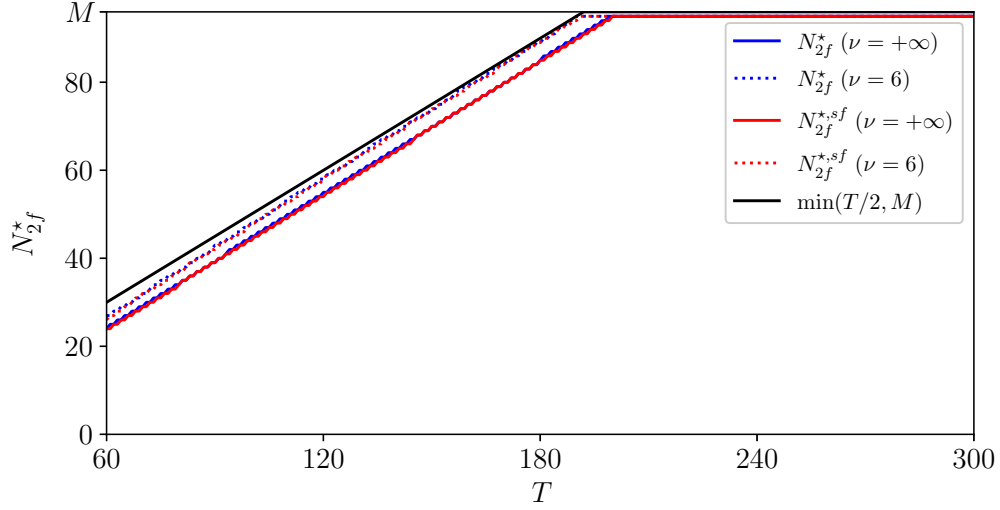
$$\theta_N^2 = \frac{N}{\sigma_\varepsilon^2} \Delta_N(\bar{\ell}_N) \quad \text{with} \quad \Delta_N(\bar{\ell}_N) = \bar{\ell}_N^\mu - \frac{N \sigma_f^2}{\sigma_\varepsilon^2 + N \sigma_f^2 \bar{\ell}_N^\beta} (\bar{\ell}_N^{\mu, \beta})^2, \quad (\text{OA3})$$

where $\bar{\ell}_N = (\bar{\ell}_N^\mu, \bar{\ell}_N^\beta, \bar{\ell}_N^{\mu, \beta})$, $\bar{\ell}_N^\mu = \frac{1}{N} \sum_{i=1}^N \mu_i^2$, $\bar{\ell}_N^\beta = \frac{1}{N} \sum_{i=1}^N \beta_i^2$, and $\bar{\ell}_N^{\mu, \beta} = \frac{1}{N} \sum_{i=1}^N \mu_i \beta_i$.

Moreover, the EU of the optimal two-fund rule $\hat{\mathbf{w}}(\alpha^*)$ is given by

$$EU(\hat{\mathbf{w}}(\alpha^*)) = \frac{N}{2\gamma c_N \sigma_\varepsilon^2} \times \frac{\kappa_{N,1}^2 \Delta_N(\bar{\ell}_N)^2}{\kappa_{N,2} \Delta_N(\bar{\ell}_N) + \kappa_{N,3} \sigma_\varepsilon^2 / T} =: EU_{2f}^{sf}(N, T, \sigma_f^2, \sigma_\varepsilon^2, \tau, \bar{\ell}_N). \quad (\text{OA4})$$

Figure OA.1: Optimal portfolio size for the two-fund rule under the single-factor model and equicorrelation assumptions



Notes. This figure depicts the optimal portfolio size for the optimal two-fund rule as a function of the sample size T under two different assumptions. First, in blue, the equicorrelation assumption, i.e., Assumption 1, which yields N_{2f}^* in (18). Second, in red, the single-factor model assumption, i.e., Assumption OA.1, which yields $N_{2f}^{*,sf}$ in (OA5). We assume that asset returns are i.i.d. multivariate t -distributed with $\nu = \infty$ (solid lines) and $\nu = 6$ (dashed lines) degrees of freedom. We calibrate the parameters on which N_{2f}^* and $N_{2f}^{*,sf}$ depend to a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, which yields $\bar{\ell}_M = (6.07 \times 10^{-5}, 1.229, 0.0082)$, $\sigma_\varepsilon^2 = 0.0014$, $\bar{\theta}_M = (0.125, 0.0169)$, $\rho = 0.74$. The single factor is the Fama-French market excess return, which has a variance $\sigma_f^2 = 0.0020$. The risk-aversion coefficient is $\gamma = 1$.

As in the main text, we use the result in Proposition OA.1 to find the optimal N for the two-fund rule as

$$N_{2f}^{*,sf} = \underset{N \in \{1, \dots, \min(M, T-5)\}}{\operatorname{argmax}} EU_{2f}^{sf}(N, T, \sigma_f^2, \sigma_\varepsilon^2, \tau, \bar{\ell}_M). \quad (\text{OA5})$$

To determine $N_{2f}^{*,sf}$, we must compute $\bar{\ell}_M$, σ_f^2 , and σ_ε^2 . To do so, we use the Fama-French market excess return as the factor f and we regress the M asset returns on f via OLS to obtain the β_i 's.

For each asset, we obtain a residual variance $\sigma_{\varepsilon_i}^2$, and we set $\sigma_\varepsilon^2 = \frac{1}{M} \sum_{i=1}^M \sigma_{\varepsilon_i}^2$.

We now compare the oracle $N_{2f}^{*,sf}$ under the single factor model to N_{2f}^* under equicorrelation. We use the same setup as in Figure 2, which yields $\bar{\ell}_M = (6.07 \times 10^{-5}, 1.229, 0.0082)$, $\sigma_\varepsilon^2 = 0.0014$, $\bar{\theta}_M = (0.125, 0.0169)$, and $\rho = 0.74$. Moreover, the variance of the market factor over the period July 1963 to August 2023 is $\sigma_f^2 = 0.0020$. Figure OA.1 depicts $N_{2f}^{*,sf}$ and N_{2f}^* as a

function of the sample size $T \in [60, 300]$ for degrees of freedom $\nu = 6$ or ∞ . We find that the equicorrelation and single-factor assumptions yield an essentially equivalent optimal N . This is because, under both assumptions, the maximum squared Sharpe ratio θ_N^2 is equal to N multiplied by a factor that has little sensitivity to N .^{OA1}

Finally, we compare the estimated optimal N under the two assumptions in empirical data. In each rolling window of size $T = 120$ months, we compute the estimate of $N_{2f}^{*,sf}$ in (OA5), $\hat{N}_{2f}^{*,sf}$, obtained by estimating $(\sigma_f^2, \sigma_\varepsilon^2, \bar{\ell}_M)$ as explained above and the parameters $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3})$ as explained in Section A.2, and compare it to $\hat{N}_{2f}^*$ in the main text.

We depict boxplots of the estimated optimal portfolio sizes under the single-factor model and equicorrelation assumptions in Figure OA.2. As in Figure OA.1, we find that the two assumptions deliver similar results, which confirms the robustness of the optimal N found under equicorrelation. Also, we observe that the single-factor estimator, $\hat{N}_{2f}^{*,sf}$, has a lower variance than $\hat{N}_{2f}^*$, because the parameters $(\sigma_f^2, \sigma_\varepsilon^2, \bar{\ell}_M)$ have less variability than $(\rho, \bar{\theta}_M)$.^{OA2}

OA.2 Equicorrelation and the maximum utility

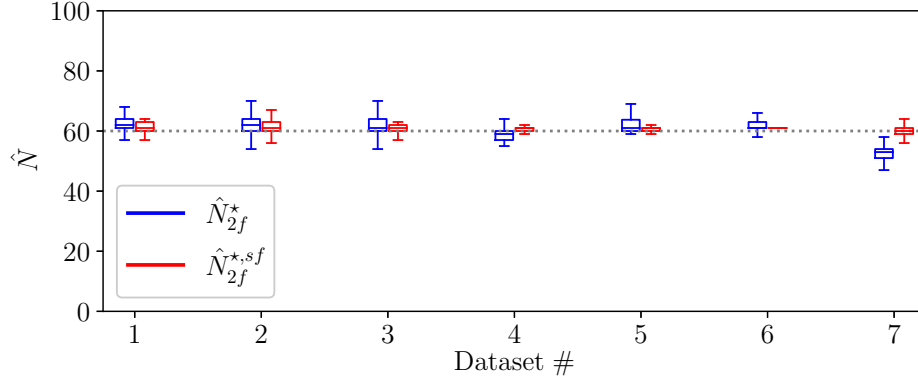
In the next proposition, we demonstrate that, under the Assumption 1, i.e., equicorrelation, $U(\mathbf{w}^*)$ is a convex function of ρ , and we derive the minimizer.

Proposition OA.2 *Under Assumption 1, the maximum utility $U(\mathbf{w}^*) = \theta_N^2/(2\gamma)$, with θ_N^2 given*

^{OA1}Engle and Kelly (2012, p. 213) also note that equicorrelation and single-factor covariance matrices are consistent: “In a one-factor world, [...] if the cross-sectional dispersion of β_j is small and idiosyncrasies have similar variance over each period, then the system is well described by Dynamic Equicorrelation.”

^{OA2}In unreported results, we find that the net out-of-sample utility delivered by the size-optimized two-fund rule is similar using either $\hat{N}_{2f}^{*,sf}$ or $\hat{N}_{2f}^*$ for all selection rules in Table 3.

Figure OA.2: Estimated optimal portfolio size in empirical data



Notes. This figure depicts boxplots of the estimated optimal portfolio size for the two-fund rule under two different assumptions: the equicorrelation assumption ($\hat{N}_{2f}^*$) and the single factor assumption ($\hat{N}_{2f}^{*,sf}$). These are the estimated counterparts of N_{2f}^* in (18) and $N_{2f}^{*,sf}$ in (OA5), respectively. The boxplots are obtained by implementing the estimation methodology in Section A.2 of the Appendix for each dataset listed in Table 1. We use a sample size $T = 120$ and set $\gamma = 1$.

by (4), is a convex function of $\rho \in (-\frac{1}{N-1}, 1)$ and is minimized by^{OA3}

$$\rho_{\min} = \frac{\frac{N}{\sqrt{N-1}} \sqrt{\bar{\theta}_{N,1}^2 (\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2) - \bar{\theta}_{N,2}}}{N(\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2) - \bar{\theta}_{N,2}}. \quad (\text{OA6})$$

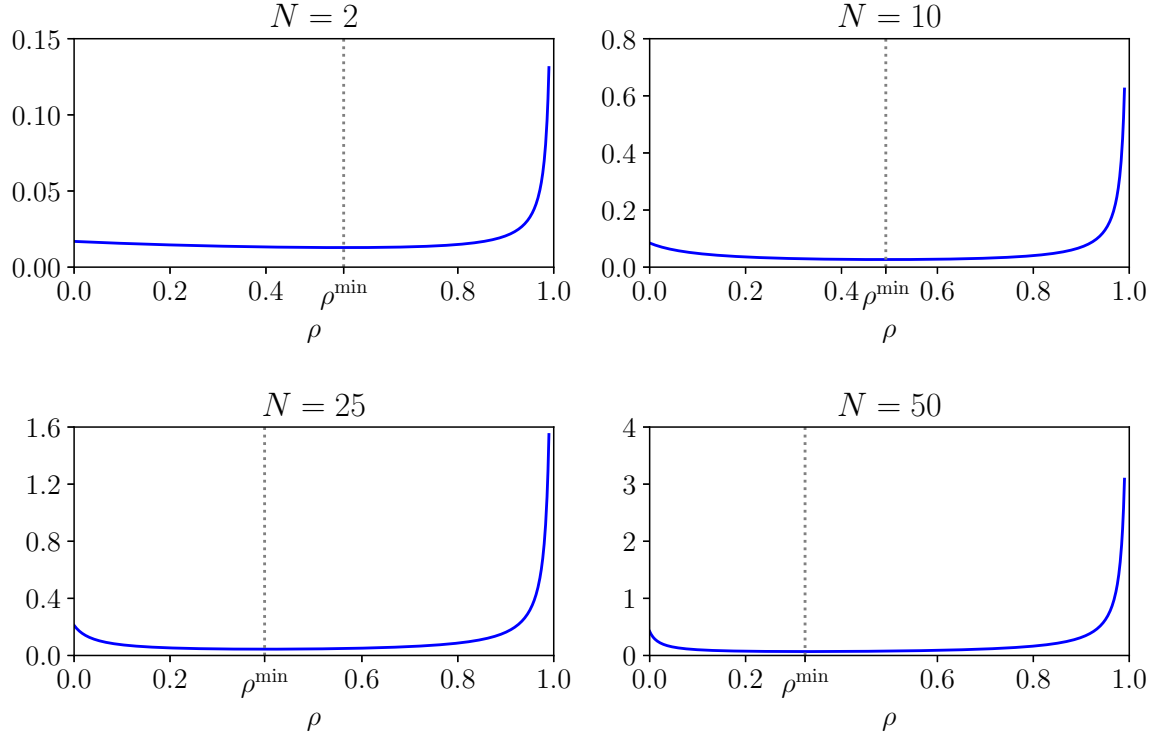
Figure OA.3 illustrates Proposition OA.2. We calibrate $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ to the 96S-BM dataset, which yields $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2}) = (0.125, 0.0169)$, and we depict $U(\mathbf{w}^*)$ as a function of ρ for $N \in \{2, 10, 25, 50\}$ and $\gamma = 1$. The figure shows that $U(\mathbf{w}^*)$ decreases with ρ for $\rho < \rho_{\min}$ and increases with ρ for $\rho > \rho_{\min}$. This result extends Gandy and Veraart (2013, p.536), who also find that the maximum utility “[...] is initially decreasing in ρ and increasing afterwards.”, and which we extend to the case where asset returns are non-homogeneous.

OA.3 Impact of the sequence of assets on expected utility

In this section, we study the impact of the sequence in which the assets are picked on $\bar{\theta}_N$, which is defined in Proposition 1, and on the EU of the two-fund rule, $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ in (16). Specif-

^{OA3}It holds that $\lim_{N \rightarrow \bar{\theta}_{N,2}/(\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2)} \rho_{\min} = 1 - \bar{\theta}_{N,2}/(2\bar{\theta}_{N,1}^2)$, see the proof of Proposition OA.2.

Figure OA.3: Impact of correlation on maximum utility



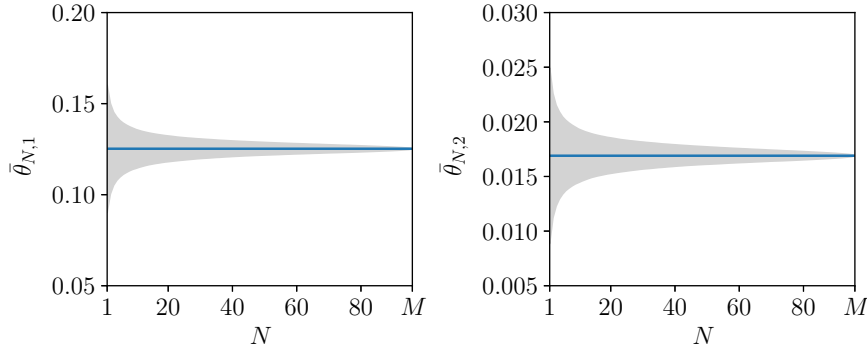
Notes. This figure depicts the maximum utility $U(\mathbf{w}^*)$ in Proposition 1 under the assumption that all assets have equal correlations ρ . We calibrate $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ to a dataset of 96 portfolios sorted on size and book-to-market spanning July 1963 to August 2023 to the 96S-BM dataset, which yields $(\bar{\theta}_{N,1}, \bar{\theta}_{N,2}) = (0.125, 0.0169)$. We depict $U(\mathbf{w}^*)$ as a function of ρ for $N \in \{2, 10, 25, 50\}$ and a risk-aversion coefficient $\gamma = 1$. The minimizer of $U(\mathbf{w}^*)$ is $\rho_{\min}$ given by (OA6).

ically, we study how close $\bar{\theta}_N$ is to $\bar{\theta}_M$ and how close our EU approximation $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$ in (18) is to $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ as a function of N .

To do so, we consider the following experiment where $\bar{\theta}_N$ is calibrated to the 96S-BM dataset. First, we randomly select $N = 1$ asset out of the M available ones and evaluate $\bar{\theta}_N$ and $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ assuming i.i.d. multivariate t -distributed returns, i.e., $\tau \sim (\nu - 2)/\chi_\nu^2$. Then, we randomly select a new asset on top of the first one and compute $\bar{\theta}_N$ and the EU on the $N = 2$ assets. We proceed until $N = M$, which allows us to depict $\bar{\theta}_N$ and the EU as a function of N for a specific sequence of assets. We repeat this exercise 10,000 times. To compute the EU, we take $T = 120$, $\rho \in \{0.2, 0.5, 0.8\}$ and $\nu \in \{6, \infty\}$.

We depict the results from this experiment in two figures. First, in Figure OA.4, we depict

Figure OA.4: Impact of the sequence of assets on $\bar{\theta}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$



Notes. This figure depicts $\bar{\theta}_{N,1}$ (left panel) and $\bar{\theta}_{N,2}$ (right panel), defined in Proposition 1, as a function of the portfolio size N . The solid blue line depicts $\bar{\theta}_{M,1}$ and $\bar{\theta}_{M,2}$, and the shaded gray area depicts the one-sigma interval of the values of $\bar{\theta}_{N,1}$ and $\bar{\theta}_{N,2}$ obtained under 10,000 random sequences of assets. We calibrate this experiment to a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023.

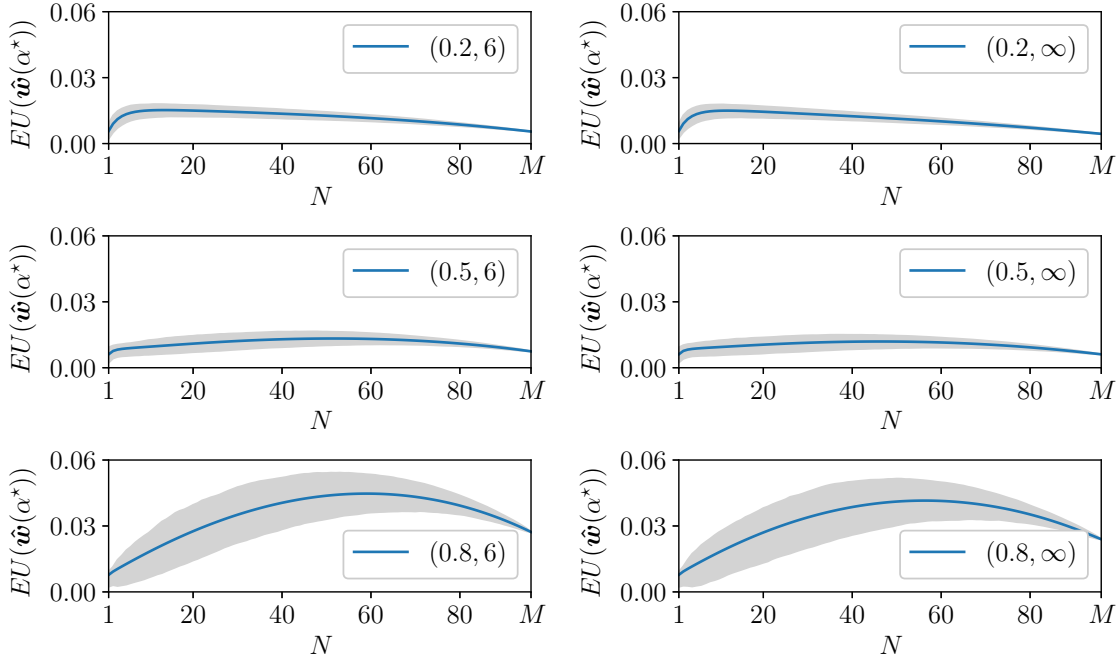
$\bar{\theta}_M$ in solid blue and a one-sigma interval of the 10,000 values of $\bar{\theta}_N$ in shaded gray, as a function of N . We observe that the interval quickly shrinks with N . Second, in Figure OA.5, we depict $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$ in solid blue and a one-sigma interval of the 10,000 values of $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ in shaded gray, as a function of N . We can observe that the gray area closely follows our curve for determining the optimal N_{2f}^* , $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$.

OA.4 Implementation of factor-plus-alpha strategy

In this section, we explain how we implement the factor-plus-alpha (F+A) strategy considered in the simulation and empirical analysis. We construct the F+A strategy using PCA and the empirical Bayes method of Da et al. (2024). Below, we try to stay close to the notation in Da et al. (2024) while ensuring consistency with our own notation.

Assume the vector of N asset excess returns $\mathbf{r}$, with mean $\boldsymbol{\mu}_N$ and covariance matrix $\boldsymbol{\Sigma}_N$,

Figure OA.5: Impact of the sequence of assets on the EU of the optimal two-fund rule



Notes. This figure depicts our approximation of the EU of the optimal two-fund rule, $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_M)$ in (18), as a function of the portfolio size N (solid blue line). We also depict in shaded gray the one-sigma interval of the 10,000 values of $EU_{2f}(N, T, \rho, \tau, \bar{\theta}_N)$ obtained with random sequences of assets. We calibrate $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3})$ to the multivariate t -distribution with $\nu \in \{6, \infty\}$ degrees of freedom, and consider $\rho \in \{0.2, 0.5, 0.8\}$. Each panel corresponds to a choice of (ρ, ν) . The risk-aversion coefficient is $\gamma = 1$. We calibrate this experiment to a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023.

follows a K -factor model of the form

$$\mathbf{r} = \underbrace{\boldsymbol{\beta}(\boldsymbol{\pi} + \mathbf{v})}_{\text{factor}} + \underbrace{\boldsymbol{\alpha}}_{\text{alpha}} + \mathbf{u}, \quad (\text{OA7})$$

where $\mathbf{u}$ and $\mathbf{v}$ have zero mean and covariance matrices $\boldsymbol{\Sigma}_u$ (of size $N \times N$) and $\boldsymbol{\Sigma}_v$ (of size $K \times K$), respectively, with $\boldsymbol{\Sigma}_u$ being diagonal. The factor-optimal and alpha-optimal portfolios are

$$\mathbf{w}_\beta = \frac{1}{\gamma} (\boldsymbol{\beta} \boldsymbol{\Sigma}_v \boldsymbol{\beta}')^{-1} \boldsymbol{\beta} \boldsymbol{\pi} \quad \text{and} \quad \mathbf{w}_\alpha = \frac{1}{\gamma} \boldsymbol{\Sigma}_u^{-1} \boldsymbol{\alpha}. \quad (\text{OA8})$$

The F+A strategy then combines $\mathbf{w}_\beta$ and $\mathbf{w}_\alpha$ to maximize the expected utility, which gives

$$\mathbf{w}_{\beta+\alpha} = \phi_\beta \mathbf{w}_\beta + \phi_\alpha \mathbf{w}_\alpha \quad \text{with} \quad \phi_x = \frac{1}{\gamma} \frac{\mathbf{w}_x' \boldsymbol{\mu}_N}{\mathbf{w}_x' \boldsymbol{\Sigma}_N \mathbf{w}_x}, \quad (\text{OA9})$$

where we assume that $\mathbf{u}$ and $\mathbf{v}$ are uncorrelated, which is the case under PCA below. Note that if we do not restrict $\boldsymbol{\Sigma}_u$ to be diagonal, we have $\mathbf{w}_{\beta+\alpha} = \mathbf{w}^* = \frac{1}{\gamma} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N$. Thus, the diagonal

constraint on Σ_u , as well as the Bayes estimation procedure for estimating $\mathbf{w}_\alpha$ explained below, help alleviate estimation risk in $\mathbf{w}^*$.

For the choice of factors $\mathbf{v}$, we make the standard choice of $\mathbf{v}$ being the first K principal components (PCs), with $K \in [1, N - 1]$. Let $\mathbf{V}$ be the $N \times K$ matrix whose columns are the first K eigenvectors of Σ_N and let $\mathbf{\Lambda}$ be the $K \times K$ diagonal matrix of the corresponding eigenvalues in decreasing order. Let also $\bar{\mathbf{V}}$ and $\bar{\mathbf{\Lambda}}$ be the corresponding matrices for the remaining PCs from $K + 1$ to N . Then, the correspondence with the above notation is $\beta = \mathbf{V}$, $\mathbf{v} = \mathbf{V}'(\mathbf{r} - \mu_N)$, $\pi = \mathbf{V}'\mu_N$, $\Sigma_v = \mathbf{\Lambda}$, $\alpha = \bar{\mathbf{V}}\bar{\mathbf{V}}'\mu_N$, and $\Sigma_u = \text{diag}(\bar{\mathbf{V}}\bar{\mathbf{\Lambda}}\bar{\mathbf{V}}')$. Therefore, the factor-optimal portfolio is $\mathbf{w}_\beta = \frac{1}{\gamma}\mathbf{V}\mathbf{\Lambda}^{-1}\pi$ and the alpha-optimal portfolio is $\mathbf{w}_\alpha = \frac{1}{\gamma}\text{diag}(\bar{\mathbf{V}}\bar{\mathbf{\Lambda}}\bar{\mathbf{V}}')^{-1}\alpha$, which combined give the F+A strategy $\mathbf{w}_{\alpha+\beta} = \phi_\beta\mathbf{w}_\beta + \phi_\alpha\mathbf{w}_\alpha$.

Regarding estimation, we first need to estimate the number of factors K . We follow the widely used method by Bai and Ng (2002), which is consistent as N and T go to infinity together. Specifically, we estimate K as

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

where $\epsilon_{it} = (\bar{\mathbf{V}}\bar{\mathbf{V}}'\mathbf{R}')_{it}$ with $\mathbf{R}$ the $T \times N$ matrix of return samples and $\bar{\mathbf{V}}$ obtained from the sample covariance matrix $\hat{\Sigma}_N$. Then, we estimate $\mathbf{w}_\beta$, ϕ_β , and ϕ_α using the sample estimators $\hat{\mu}_N$ and $\hat{\Sigma}_N$ because these are low-dimensional quantities; $\mathbf{w}_\beta$ is low-dimensional because $\hat{K}$ is typically small. However, $\mathbf{w}_\alpha$ requires substantial shrinkage because it is high-dimensional. To estimate $\mathbf{w}_\alpha$, we use the empirical Bayes method in Algorithm 1 of Da et al. (2024):

- (a) Let $\hat{s}_i = \hat{\alpha}_i / (\hat{\Sigma}_u)_{ii}^{1/2}$, where $\hat{\alpha}_i$ and $\hat{\Sigma}_u$ are obtained from $\hat{\mu}_N$ and $\hat{\Sigma}_N$.
- (b) Estimate the marginal density of the $\hat{s}_i$'s using a Gaussian kernel $\phi_{1/T}(x)$ and a bandwidth

$k_N \propto 1/\ln(N)$, i.e.,

$$\hat{p}(x) = \frac{1}{Nk_N} \sum_{i=1}^N \phi_{1/T} \left(\frac{\hat{s}_i - x}{k_N} \right), \quad \phi_{1/T} \left(\frac{\hat{s}_i - x}{k_N} \right) = \sqrt{\frac{T}{2\pi}} \exp \left(-\frac{T}{2k_N^2} (\hat{s}_i - x)^2 \right). \quad (\text{OA10})$$

(c) Define $\check{\psi}(x) = x + \frac{1+k_N^2}{T} \frac{d \ln \hat{p}(x)}{dx}$, where

$$\frac{d \ln \hat{p}(x)}{dx} = \frac{T}{Nk_N^3 \hat{p}(x)} \sum_{i=1}^N (\hat{s}_i - x) \phi_{1/T} \left(\frac{\hat{s}_i - x}{k_N} \right). \quad (\text{OA11})$$

Let $M_{\hat{\beta}} = I_N - \hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}' = I_N - \hat{V}\hat{V}'$, where $\hat{V}$ is the matrix of the first K eigenvectors of $\hat{\Sigma}_N$. Then, w_α is estimated as $\hat{w}_\alpha = \frac{1}{\gamma} M_{\hat{\beta}} \hat{\Sigma}_u^{-1/2} \hat{\psi}$, where $\hat{\Sigma}_u^{-1/2}$ is the sample estimate of $\Sigma_u^{-1/2} = \text{diag}(\bar{V} \bar{\Lambda}^{-1/2} \bar{V}')$ and $\hat{\psi}$ is obtained using an isotonic regression of $\check{\psi}(\hat{s})$:

$$\hat{\psi} = \underset{x \in \mathbb{R}^N}{\text{argmin}} \left\| x - \check{\psi}(\hat{s}) \right\|_2 \quad \text{subject to} \quad x_i \leq x_j \text{ if } \hat{s}_i \leq \hat{s}_j \quad \forall i, j. \quad (\text{OA12})$$

Like Da et al. (2024), we consider values of the bandwidth $k_N \in \{0.25, 0.5, \dots, 4\}/\ln(N)$, we split the sample into a training set composed of the first 80% of the data and a validation set composed of the remaining 20%, and we choose k_N for which $\hat{w}_{\beta+\alpha}$ achieves the maximum utility on the validation set. In the empirical analysis, we also implement the F+A strategy using the shrinkage covariance matrix of Ledoit and Wolf (2004) instead of the sample covariance matrix $\hat{\Sigma}_N$.

OA.5 Estimation of EWRF portfolio

In this section, we explain how we estimate the EWRF portfolio, a benchmark we consider in the empirical analysis that combines the EW portfolio with the risk-free asset. It is given by $w_{ewrf} = (\lambda_{ew,N}/\gamma)w_{ew}$, where $\lambda_{ew,N}$ is defined in (A11). Its sample estimator, obtained by plugging $\hat{\lambda}_{ew,N}$ in (A28), is $\hat{w}_{ewrf} = (\hat{\lambda}_{ew,N}/\gamma)w_{ew}$. Because of estimation errors in $\hat{\lambda}_{ew,N}$, it is possible to improve upon the EU of $\hat{w}_{ewrf}$. In particular, it is useful to combine the EWRF portfolio with

the risk-free asset a second time to account for estimation errors in $\hat{\lambda}_{ew,N}$. Such a “double shrinkage” approach has been considered in the literature; see, e.g., Kan and Wang (2023) who shrink the sample mean-variance portfolio toward different portfolios and, after estimating the shrinkage intensities, shrink the combination portfolio a second time toward the risk-free asset to account for estimation errors in the shrinkage intensities. In the next proposition, we derive the EU-optimal combination of $\hat{\mathbf{w}}_{ewrf}$ with the risk-free asset.

Proposition OA.3 *Let $T > 5$ and the asset returns be i.i.d. multivariate normally distributed. Then, the EU-optimal combination of $\hat{\mathbf{w}}_{ewrf} = (\hat{\lambda}_{ew,N}/\gamma)\mathbf{w}_{ew}$ with the risk-free asset is $\eta^*\hat{\mathbf{w}}_{ewrf}$ with*

$$\eta^* = \frac{T-5}{T-3} \left(\frac{\theta_{ew,N}^2}{\theta_{ew,N}^2 + \frac{1}{T}} \right) \in [0, 1]. \quad (\text{OA13})$$

To obtain the final estimator of the EWRF portfolio, we estimate η^* by estimating $\theta_{ew,N}^2$ in (A11) with a newly derived adjusted estimator given by^{OA4}

$$\hat{\theta}_{ew,N,a}^2 = \frac{(T-3)\hat{\theta}_{ew,N}^2 - 1}{T} + \frac{2(\hat{\theta}_{ew,N}^2)^{\frac{1}{2}}(1 + \hat{\theta}_{ew,N}^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\theta}_{ew,N}^2/(1+\hat{\theta}_{ew,N}^2)}\left(\frac{1}{2}, \frac{T-1}{2}\right)}, \quad (\text{OA14})$$

where $B_x(a, b)$ is the incomplete beta function defined in Section A.2.2 and $\hat{\theta}_{ew,N}^2$ is the sample estimator of $\theta_{ew,N}^2$ obtained with $\hat{\boldsymbol{\mu}}_N$ and $\hat{\boldsymbol{\Sigma}}_N$ in (5). Therefore, the final EWRF portfolio is

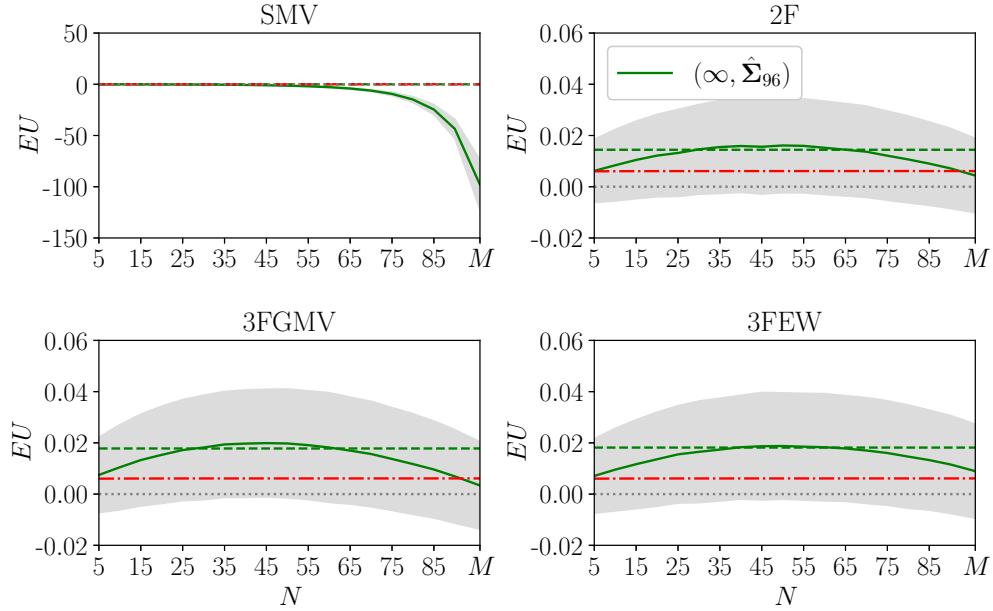
$$\hat{\eta}^* \frac{\hat{\lambda}_{ew,N}}{\gamma} \mathbf{w}_{ew} \quad \text{with} \quad \hat{\eta}^* = \frac{T-5}{T-3} \left(\frac{\hat{\theta}_{ew,N,a}^2}{\hat{\theta}_{ew,N,a}^2 + \frac{1}{T}} \right). \quad (\text{OA15})$$

OA.6 Additional simulation results

We now report simulation results that complement those in Section 4. Section OA.6.1 considers the true correlation matrix with $\nu = \infty$, Section OA.6.2 the equicorrelation matrix, and Section OA.6.3 the block-correlation matrix, respectively.

^{OA4}The adjusted estimator $\hat{\theta}_{ew,N,a}^2$ in (OA14) is obtained by noting that $\hat{\theta}_{ew,N}^2 \sim \chi_1^2(T\theta_{ew,N}^2)/\chi_{T-1}^2$ and applying Kan and Zhou (2007, Equation (A.9)).

Figure OA.6: Expected out-of-sample utility in simulated data ($\nu = \infty, \Sigma_M = \hat{\Sigma}_{96}$)



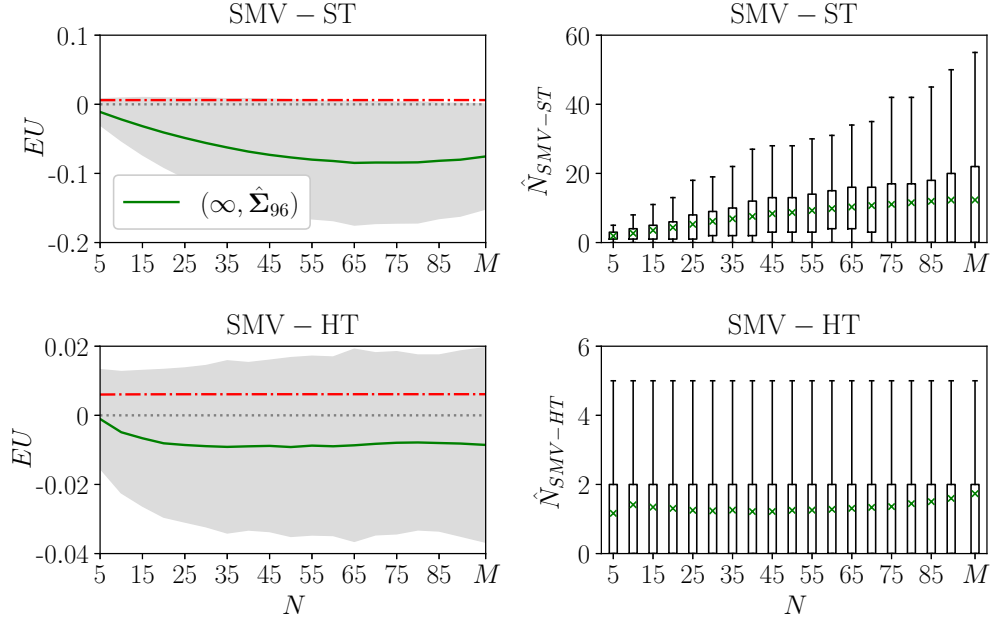
Notes. This figure depicts the expected out-of-sample utility (EU) of the sample mean-variance portfolio (SMV, top left panel), two-fund rule (2F, top right panel), GMV-three-fund rule (3FGMV, bottom left), and EW-three-fund rule (3FEW, bottom right) as a function of the portfolio size N . We simulate 10,000 samples of t -distributed returns with $\nu = \infty$ degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}$. In each panel, the solid blue line depicts the EU in (21), the shaded gray area depicts the one-sigma interval around the EU across all simulations, and the dashed horizontal blue line depicts the EU obtained when using the estimated optimal N for each strategy, i.e., $\hat{N}_{smv}^*$ for the SMV portfolio, $\hat{N}_{2f}^*$ for 2F, $\hat{N}_{3f,g}^*$ for 3FGMV, and $\hat{N}_{3f,ew}^*$ for 3FEW. The dash-dotted red line depicts the EU of the equally weighted portfolio. The dotted horizontal gray line depicts the zero EU level. The risk-aversion coefficient is $\gamma = 1$.

OA.6.1 True correlation matrix with $\nu = \infty$

In Figure OA.6, we replicate Figure 5 in the main text for a number of degrees of freedom $\nu = \infty$ (i.e., normality) instead of $\nu = 6$. We find that the two figures are very similar, and thus, our estimated optimal N delivers an EU close to the maximum for different ν .

In Figure OA.7, we replicate Figure 6 in the main text for $\nu = \infty$. Again, we find that the two figures are very similar, and thus, the conclusions are robust to the choice of ν .

Figure OA.7: Expected out-of-sample utility and estimated portfolio size of soft and hard-thresholding of the sample mean-variance portfolio in simulated data ($\nu = \infty$, $\Sigma_M = \hat{\Sigma}_{96}$)



Notes. This figure depicts the expected out-of-sample utility (EU) and the estimated optimal portfolio size of the soft and hard-thresholding versions of the sample mean-variance portfolio, SMV-ST and SMV-HT, which are described in Section 4.3. We depict these as a function of the number of assets N on which SMV-ST and SMV-HT are estimated. We simulate 10,000 samples of t -distributed returns with $\nu = 6$ degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}$. In the left panels, the solid green line depicts the EU of SMV-ST or SMV-HT in (21), the shaded gray area depicts the one-sigma interval around the EU across all simulations, the dash-dotted red line depicts the EU of the equally weighted portfolio, and the dotted horizontal gray line depicts the zero EU level. The risk-aversion coefficient is $\gamma = 1$. In the right panels, the green crosses depict the average estimated optimal portfolio size.

OA.6.2 Equicorrelation matrix

In Figure OA.8, we replicate Figure 5 in the main text using a population covariance matrix satisfying equicorrelation, i.e., $\Sigma_M = \hat{\Sigma}_{96}(\bar{\rho})$. Given that Assumption 1 is satisfied, we know from Sections 3 and A.1 the oracle optimal N for the SMV, 2F, 3FGMV, and 3FEW portfolios obtained under the true parameters $(\hat{\mu}_{96}, \hat{\Sigma}_{96}(\bar{\rho}), \nu)$. Therefore, in Figure OA.8, we depict N^* with a vertical dotted line and its EU with an horizontal dotted line; it is not visible for SMV because we start at $N = 5$ and N_{smv}^* is smaller than five. Panels A and B of Figure OA.8 consider degrees of freedom $\nu = 6$ and $\nu = \infty$, respectively. Our main observation from Figure OA.8 is that, unlike

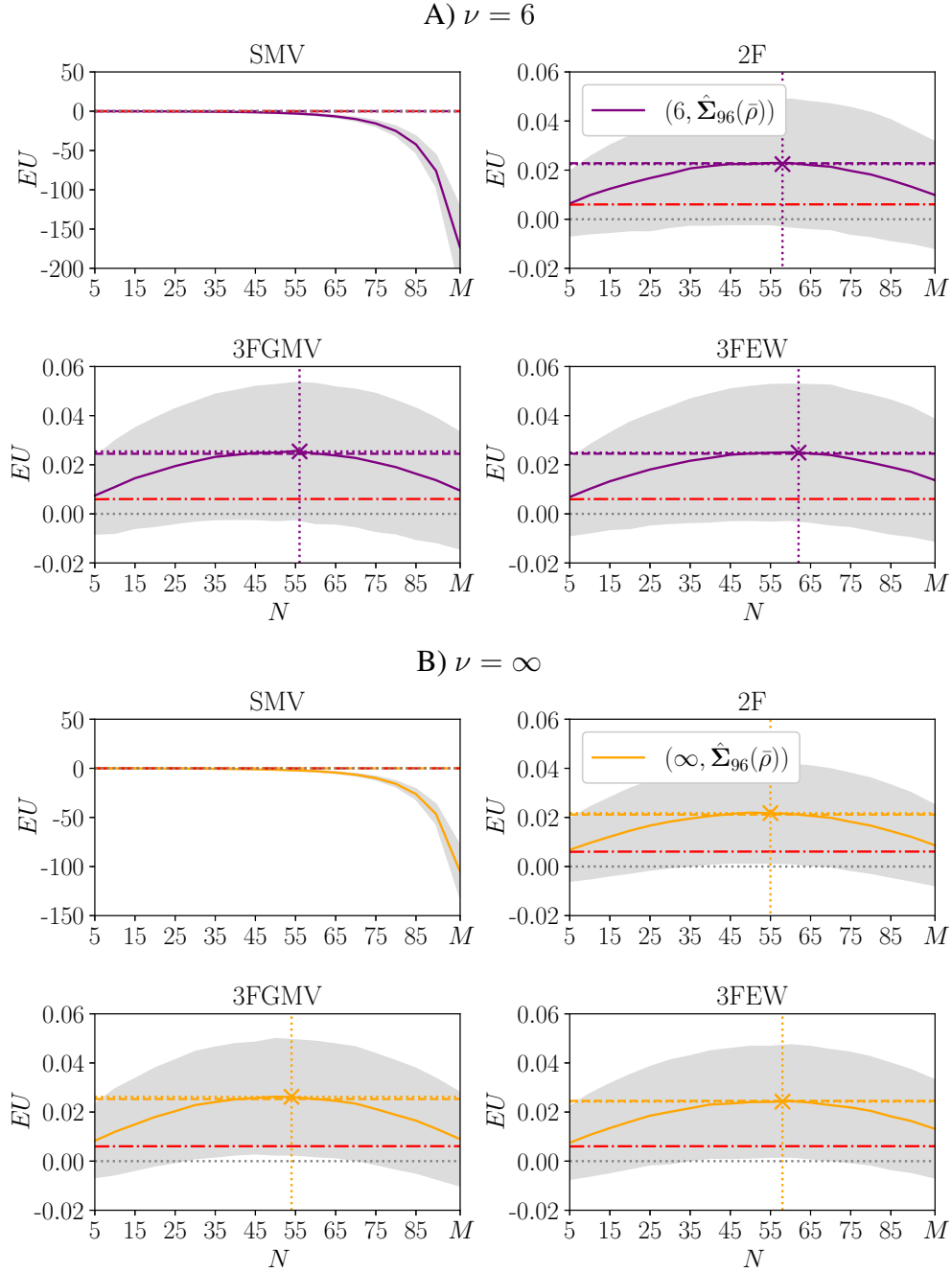
in Figures 5 and OA.6 where we run the same analysis with $\Sigma_M = \hat{\Sigma}_{96}$ violating Assumption 1, there is essentially no gap between the EU delivered by our estimated optimal N and that under the oracle optimal N .

Next, in Figure OA.9, we depict the EU of the two-fund rule under equicorrelation as a function of ρ . We set degrees of freedom $\nu = 6$ and a risk-aversion coefficient $\gamma = 1$. We depict the EU considering three different portfolio sizes: the oracle N_{2f}^* , the estimated $\hat{N}_{2f}^*$, and $N = M$. Figure 4 in the main text depicts N_{2f}^* and $\hat{N}_{2f}^*$ as a function of ρ . We observe that both N_{2f}^* and $\hat{N}_{2f}^*$ deliver a larger EU than $N = M$, as expected. Moreover, remarkably, $\hat{N}_{2f}^*$ delivers almost the same EU as N_{2f}^* for all ρ even though Figure 4 shows that $\hat{N}_{2f}^*$ has a substantial variance when ρ moves away from one. Finally, the EU of the two-fund rule is a convex function of ρ , consistent with the maximum Sharpe ratio, θ_N in (2), being also a convex function of ρ as shown in Section OA.2 of this Online Appendix.

OA.6.3 Block-correlation matrix

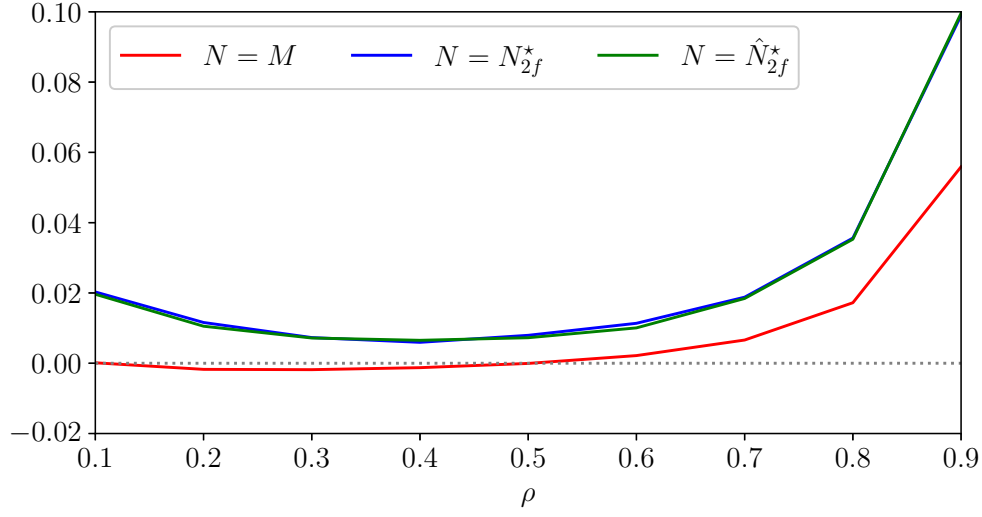
In Figure OA.10, we run the same experiment as in Sections 4.2, OA.6.1, and OA.6.2, except that we set the population covariance matrix as $\Sigma_M = \hat{\Sigma}_{96}^b(\bar{\rho}) = \hat{D}_{96} \mathbf{P}_{96}^b(\bar{\rho}) \hat{D}_{96}$, where $\mathbf{P}_{96}^b(\bar{\rho})$ is a block-correlation matrix. We consider two blocks for which the return correlation within the blocks is $\bar{\rho} = 0.74$ and that between the blocks is zero. The two blocks correspond to the first and last 48 assets in the 96S-BM dataset. We observe from Figure OA.10 that even in this setup that violates our theoretical equicorrelation assumption, our estimated optimal N still delivers an EU close to the maximum, in particular for three-fund strategies.

Figure OA.8: Expected out-of-sample utility in simulated data with equicorrelation matrix



Notes. Each panel depicts the expected out-of-sample utility (EU) of the sample mean-variance portfolio (SMV, top left panel), two-fund rule (2F, top right panel), GMV-three-fund rule (3FGMV, bottom left), the EW-three-fund rule (3FEW, bottom right) as a function of the portfolio size N . We simulate 10,000 samples of t -distributed returns with $\nu = 6$ (Panel A) and $\nu = \infty$ (Panel B) degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}(\bar{\rho})$, where $\hat{\Sigma}_{96}(\bar{\rho})$ is an equicorrelation covariance matrix and $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\Sigma}_{96}$. In each panel, the solid line depicts the EU in (21), the shaded gray area depicts the one-sigma interval around the EU across all simulations, and the dashed horizontal line depicts the EU obtained when using the estimated optimal N for each strategy, i.e., N_{smv}^* for the SMV portfolio, N_{2f}^* for 2F, N_{3fg}^* for 3FGMV, and N_{3fe}^* for 3FEW. The vertical and horizontal dotted lines depict the oracle N and its EU, respectively, with the crosses showing the intersection. The dash-dotted red line depicts the EU of the equally weighted portfolio. The dotted horizontal gray line depicts the zero EU level. The risk-aversion coefficient is $\gamma = 1$.

Figure OA.9: Impact of correlation ρ on expected out-of-sample utility of the two-fund rule in simulated data

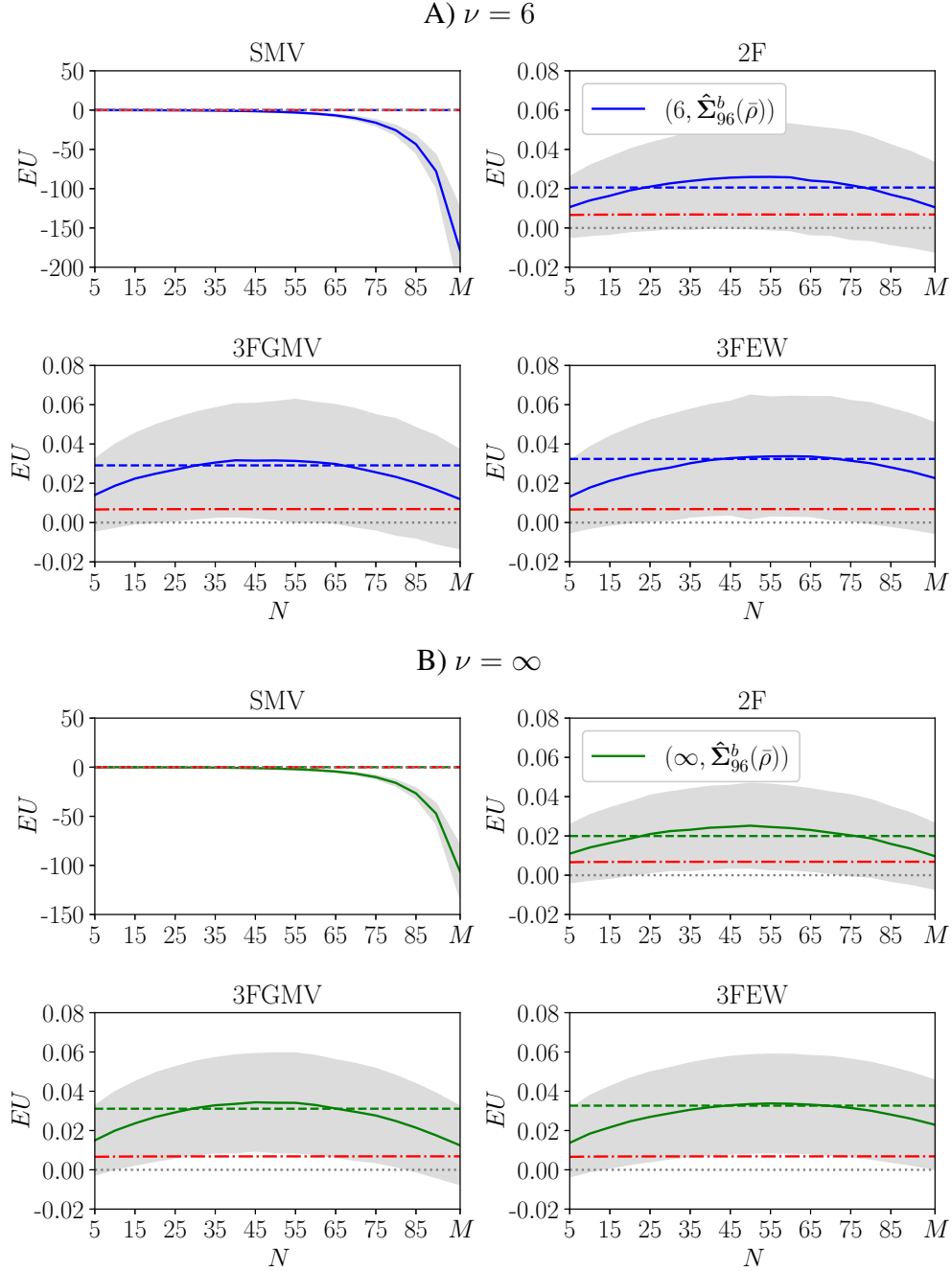


Notes. This figure depicts the expected out-of-sample utility (EU) of the two-fund rule as a function of the equicorrelation ρ for three different portfolio sizes: the oracle $N = N_{2f}^*$, the estimated $N = \hat{N}_{2f}^*$, and $N = M$. $\hat{N}_{2f}^*$ is obtained following the estimation methodology in Section A.2 of the Appendix. We simulate 10,000 samples of t -distributed returns with $\nu = 6$ degrees of freedom. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}(\rho)$, where $\hat{\Sigma}_{96}(\rho)$ is an equicorrelation covariance matrix and ρ varies between 0.1 and 0.9 with a step size of 0.1. For each ρ , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. The risk-aversion coefficient is $\gamma = 1$. The dotted horizontal gray line depicts the zero EU level.

OA.7 Additional empirical results

We now report empirical results that complement those in Section 5. Section OA.7.1 looks at the equicorrelation and elliptical assumptions, Section OA.7.2 considers a nonlinear shrinkage estimator of the covariance matrix, Section OA.7.3 applies our different asset selection rules to the SMV portfolio, Section OA.7.4 reports the out-of-sample Sharpe ratio of the considered portfolios, Section OA.7.5 evaluates the portfolio turnover and how much of it is driven by changes to the selected assets over time, Section OA.7.6 further analyzes the performance delivered by the random asset selection rule, Section OA.7.7 investigates the overlap between the selection rules implemented in the main text, and Section OA.7.8 studies the effect of the portfolio size on the out-of-sample performance of the benchmark strategies.

Figure OA.10: Expected out-of-sample utility in simulated data with block-correlation matrix



Notes. Each panel depicts the expected out-of-sample utility (EU) of the sample mean-variance portfolio (SMV, top left panel), two-fund rule (2F, top right panel), GMV-three-fund rule (3FGMV, bottom left), and EW-three-fund rule (3FEW, bottom right) as a function of the portfolio size N . We simulate 10,000 samples of t -distributed returns with $\nu = 6$ (Panel A) and $\nu = \infty$ (Panel B) degrees of freedom. For each N , we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market spanning July 1963 to August 2023, we set $\mu_M = \hat{\mu}_{96}$ and $\Sigma_M = \hat{\Sigma}_{96}^b(\bar{\rho})$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\Sigma}_{96}$ and the block-correlation matrix $\hat{\Sigma}_{96}^b(\bar{\rho})$ is described in Section OA.6.3. In each panel, the solid line depicts the EU in (21), the shaded gray area depicts the one-sigma interval around the EU across all simulations, and the dashed horizontal line depicts the EU obtained when using the estimated optimal N for each strategy, i.e., $\hat{N}_{smv}^*$ for the SMV portfolio, $\hat{N}_{2f}^*$ for 2F, $\hat{N}_{3f,g}^*$ for 3FGMV, and $\hat{N}_{3f,ew}^*$ for 3FEW. The dash-dotted red line depicts the EU of the equally weighted portfolio. The dotted horizontal gray line depicts the zero EU level. The risk-aversion coefficient is $\gamma = 1$.

OA.7.1 Equicorrelation and elliptical assumptions

We rely on two assumptions in our methodology to derive optimal portfolio combinations and portfolio sizes: Assumption 1 states that asset returns are equicorrelated and Assumption 2 states that asset returns are i.i.d. multivariate elliptically distributed. We now look at these two assumptions in the seven datasets we use in the empirical analysis of Section 5.

Regarding Assumption 1, we define a measure of distance between the correlation matrix observed in the data and the equicorrelation matrix. Specifically, we estimate ρ as the average of sample correlations obtained from the sample covariance matrix $\hat{\Sigma}_M$ as in (A32),

$$\bar{\rho} = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M \hat{\rho}_{ij}, \quad (\text{OA16})$$

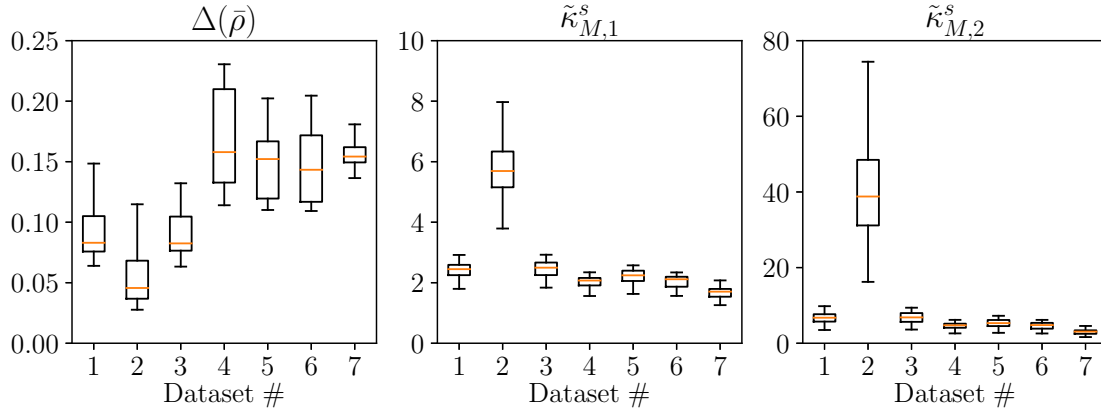
where $\hat{\rho}_{ij}$ is the sample correlation between assets i and j , and we define $\Delta(\bar{\rho})$ as the root mean squared error between the sample correlation matrix and the equicorrelation matrix obtained by setting $\rho = \bar{\rho}$, i.e.,

$$\Delta(\bar{\rho}) = \sqrt{\frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=i+1}^M (\hat{\rho}_{ij} - \bar{\rho})^2}. \quad (\text{OA17})$$

In the left panel of Figure OA.11, we depict, for each of the seven datasets, boxplots of $\Delta(\bar{\rho})$ across all estimation windows of size $T = 120$ months. For three out of the seven datasets (96S-BM, 100S-OP, 108CHA), $\Delta(\bar{\rho})$ is quite low and hovers around 5-10%. However, for the remaining datasets (107IN-CHA, 94IN-NV, 98IN-CHA-NV, STO100), $\Delta(\bar{\rho})$ is larger and around 10-20%. Thus, there are varying degrees to which the equicorrelation assumption is representative of the true correlation structure in the data, which makes the consistent outperformance delivered by our method to determine the optimal N particularly appreciable.

Regarding Assumption 2, we look at the three parameters $(\kappa_{M,1}, \kappa_{M,2}, \kappa_{M,3})$ defined in (10)–

Figure OA.11: Assumptions of equicorrelation and elliptical distribution in empirical data



Notes. This figure depicts boxplots of the estimated values of $\Delta(\bar{\rho})$ in (OA17), which measures the distance between the sample correlation matrix and the equicorrelation matrix, and $\tilde{\kappa}_{M,1}^s$ and $\tilde{\kappa}_{M,2}^s$ in (A25)–(A26), which measure the impact of elliptical fat tails on the expected out-of-sample utility, for each dataset listed in Table 1 across all rolling estimation windows of size $T = 120$ months.

(12) that control the impact of elliptical fat tails on the EU and that appear in the formulas for the optimal combination coefficients in the two-fund and three-fund rules. As explained in Section A.2.1, in the empirical analysis we estimate them following El Karoui (2010, 2013), i.e., we estimate $\kappa_{M,1}$ and $\kappa_{M,3}$ with $\tilde{\kappa}_{M,1}^s$ in (A25) and $\kappa_{M,2}$ with $\tilde{\kappa}_{M,2}^s$ in (A26). The middle and right panels of Figure OA.11 depict boxplots of $\tilde{\kappa}_{M,1}^s$ and $\tilde{\kappa}_{M,2}^s$ across estimation windows. The closer they are to one, the closer is the data to being normally distributed. We observe that $\tilde{\kappa}_{M,1}^s$ and $\tilde{\kappa}_{M,2}^s$ substantially depart from one, particularly for the 108CHA dataset, but for the other datasets too. These results indicate that it is crucial to account for the fat tails of returns in determining optimal combination rules.

OA.7.2 Nonlinear shrinkage covariance matrix

In Table OA.1, we replicate the results in Table 4 using the nonlinear shrinkage estimator of the covariance matrix of Ledoit and Wolf (2020). We observe similar results to those obtained under the linear shrinkage covariance matrix in Table 4, except for the 108CHA dataset where the nonlinear

shrinkage estimator does not behave well.

OA.7.3 Performance of size-optimized SMV portfolio

In Table OA.2, we report the annualized net out-of-sample utility of the size-optimized SMV portfolio for which N is determined using our theory in Section 3, i.e., using $\hat{N}_{smv}^*$, and the assets are selected using the different selection rules in Section 5.3. We find for all asset selection rules that reducing N substantially improves the performance of the SMV portfolio. For instance, for the 96S-BM dataset and the linear shrinkage covariance matrix, the SMV portfolio implemented on all assets delivers a net utility of -1443 , whereas BWSR, which is the *worst* selection rule for that dataset, delivers a net utility of -33.3 . However, because the SMV portfolio is too exposed to estimation risk, we still obtain negative utilities in most cases even with a reduced N . It is preferable instead to apply our method for determining the optimal N to a better reference portfolio like the two-fund and three-fund rules.

Next, we compare the portfolio size under the three different ways of optimizing the SMV portfolio size we consider in Table OA.2, across different months t . First, our estimated portfolio size $\hat{N}_{smv,t}^*$ following the methodology in Section A.2 of the Appendix. Second, the portfolio size obtained under soft-thresholding as detailed in Section 4.3. Specifically, in each month t , we solve (22) under a cross-validated δ , which yields the vector of weights $\hat{w}_{st,t}$, from which we can compute the portfolio size as

$$\hat{N}_{st,t} = \sum_{i=1}^M \mathbb{1}_{\{|\hat{w}_{st,i,t}| > 0\}}. \quad (\text{OA18})$$

Third, the portfolio size obtained under hard-thresholding as detailed in Section 4.3. Specifically, in each month t , we obtain $\hat{N}_{ht,t}$ in (23) under a cross-validated $\bar{w}$. We depict boxplots of portfolio

Table OA.1: Net out-of-sample utility with a nonlinear shrinkage covariance matrix

Port- folio strat.	Asset select. rule	Sample covariance matrix							nonlinear shrinkage covariance matrix						
		Dataset							Dataset						
		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO
2F	All	-0.31	-59.6	-15.3	99.4	-18.8	56.8	-5.47	4.64	-3E+31	1.05	44.4	2.19	16.4	0.10
2F	Rand	12.3 [•]	14.2 [•]	-7.17 [•]	145 [•]	1.46 [•]	149 [•]	-1.77 [•]	18.6 [•]	-5E+05 [•]	2.83 [•]	140 [•]	24.1 [•]	79.0 [•]	3.19 [•]
2F	MaxSR	1.38	11.6 [•]	-7.48	23.6 [○]	-10.5	-27.1 [•]	7.75 [•]	1.86 [○]	-3E+04 [•]	-5.41	28.7 [•]	14.2 [•]	10.5	12.5 [•]
2F	MinSR	25.4 [•]	73.2 [•]	7.10 [•]	146	24.9 [•]	139 [•]	-3.86 [•]	19.1 [•]	-2E+05 [•]	5.54 [•]	104 [•]	32.7 [•]	68.5 [•]	-2.49 [•]
2F	BWSR	7.39	84.4 [•]	-4.42	133	11.7 [○]	242 [•]	-29.9 [•]	23.7 [•]	-6E+05 [•]	7.02	150 [•]	39.1 [•]	102 [•]	-13.3 [•]
2F	MaxVar	25.0 [•]	124 [•]	17.9 [•]	114	21.7 [•]	83.8	-7.25 [•]	24.9 [•]	-2E+05 [•]	9.04 [•]	126 [•]	41.6 [•]	72.1 [•]	-3.85 [•]
2F	MinVar	2.47	12.5 [•]	-13.1	219 [•]	-1.88	225 [•]	-0.42 [•]	4.93	-3E+04 [•]	-3.89	150 [•]	21.2 [•]	80.9 [•]	13.0 [•]
2F	BWVar	13.0	89.3 [•]	-6.76	182 [○]	0.14	180 [•]	-0.53 [•]	22.6 [•]	-5E+05 [•]	5.16	141 [•]	30.1 [•]	93.3 [•]	5.09 [•]
2F	MinPC	27.4 [•]	112 [•]	-6.23	115	17.9 [•]	94.1	-1.01 [•]	26.3 [•]	-5E+05 [•]	-2.38	118 [•]	40.5 [•]	50.3 [•]	7.05 [•]
2F	Best θ_N^2	-1.22	14.7 [•]	-8.77	182 [○]	4.73	137 [•]	-9.71 [•]	19.1 [•]	-5E+05 [•]	-0.27	143 [•]	30.5 [•]	80.6 [•]	3.28 [•]
2F	MaxW	-127 [•]	-305 [•]	-179 [•]	-581 [•]	-129 [•]	-366 [•]	-77.2 [•]	33.4 [•]	-2E+05 [•]	9.79	205 [•]	42.3 [•]	132 [•]	-13.5 [•]
3FGMV	All	0.47	-60.5	-12.5	99.1	-19.7	55.4	-8.35	5.45	-2E+31	1.64	44.7	3.13	16.5	1.98
3FGMV	Rand	16.8 [•]	14.5 [•]	2.65 [•]	145 [•]	6.11 [•]	148 [•]	-3.74 [•]	22.0 [•]	-6E+05 [•]	10.6 [•]	141 [•]	16.3 [•]	79.8 [•]	11.5 [•]
3FGMV	MaxSR	-4.06	10.8 [•]	-1.93 [○]	15.6 [•]	-6.02	-27.4 [•]	13.5 [•]	2.99	-3E+04 [•]	3.98	30.0 [•]	24.3 [•]	11.3	23.3 [•]
3FGMV	MinSR	36.0 [•]	71.6 [•]	12.9 [•]	150	31.5 [•]	141 [•]	-2.49 [•]	25.7 [•]	-2E+05 [•]	8.83 [•]	106 [•]	37.9 [•]	71.0 [•]	4.49 [•]
3FGMV	BWSR	10.9	82.4 [•]	1.02	143	19.9 [•]	253 [•]	-7.29	30.3 [•]	-6E+05 [•]	16.5 [•]	152 [•]	39.5 [•]	104 [•]	11.7 [•]
3FGMV	MaxVar	25.1 [•]	126 [•]	24.9 [•]	117	21.7 [•]	85.0	-13.4 [•]	27.3 [•]	-2E+05 [•]	12.9 [•]	128 [•]	34.2 [•]	74.2 [•]	-0.30 [•]
3FGMV	MinVar	6.13	13.3 [•]	-14.9	222 [•]	0.35 [○]	219 [•]	9.13 [•]	7.25	-4E+04 [•]	1.63	150 [•]	25.2 [•]	80.4 [•]	21.4 [•]
3FGMV	BWVar	27.4 [•]	89.0 [•]	2.74	183 [○]	11.2 [•]	190 [•]	0.08 [•]	28.4 [•]	-5E+05 [•]	13.1 [•]	145 [•]	31.9 [•]	94.6 [•]	13.7 [•]
3FGMV	MinPC	33.3 [•]	113 [•]	6.47 [•]	115	25.9 [•]	91.3	-10.1	28.8 [•]	-6E+05 [•]	9.59 [•]	120 [•]	36.4 [•]	52.7 [•]	11.4 [•]
3FGMV	Best θ_N^2	12.3	7.39 [•]	-0.08	122	1.33 [○]	165 [•]	-4.21 [•]	22.9 [•]	-6E+05 [•]	10.2 [•]	147 [•]	27.7 [•]	81.0 [•]	11.9 [•]
3FGMV	MaxW	-136 [•]	-300 [•]	-147 [•]	-650 [•]	-110 [•]	-325 [•]	-82.9 [•]	38.3 [•]	-2E+05 [•]	20.0 [•]	210 [•]	43.7 [•]	132 [•]	12.6 [•]
3FEW	All	1.83	-62.8	-10.8	102	-16.8	60.8	14.1	6.82	-3E+31	2.95	43.2	3.06	17.2	19.1
3FEW	Rand	13.3 [•]	13.6 [•]	-0.44 [•]	146 [•]	2.76 [•]	150 [•]	15.3 [•]	18.7 [•]	-5E+05 [•]	7.61 [•]	139 [•]	17.2 [•]	78.9 [•]	19.7 [•]
3FEW	MaxSR	3.68	12.6 [•]	-0.61	23.6 [•]	-12.1	-20.4 [•]	17.6 [•]	3.60	-3E+04 [•]	2.85	28.1 [•]	6.34	11.6	17.9
3FEW	MinSR	24.8 [•]	70.8 [•]	11.3 [•]	145	24.8 [•]	146 [•]	9.96 [•]	18.6 [•]	-2E+05 [•]	7.41	102 [•]	25.9 [•]	68.0 [•]	12.1 [•]
3FEW	BWSR	8.43	84.4 [•]	3.41	135	17.9 [•]	237 [•]	12.6	24.8 [•]	-5E+05 [•]	11.4 [•]	147 [•]	34.8 [•]	101 [•]	19.6
3FEW	MaxVar	26.0 [•]	122 [•]	25.3 [•]	124	16.0 [•]	98.9	8.08 [•]	23.5 [•]	-2E+05 [•]	12.2 [•]	126 [•]	31.1 [•]	71.4 [•]	13.0 [•]
3FEW	MinVar	4.93	12.3 [•]	-5.87	220 [•]	0.40	230 [•]	19.2 [•]	7.97	-3E+04 [•]	4.06	150 [•]	15.6 [•]	82.2 [•]	23.2 [•]
3FEW	BWVar	14.2	88.6 [•]	-3.30	195 [○]	-0.15	187 [•]	18.9 [•]	21.9 [•]	-5E+05 [•]	9.71 [•]	141 [•]	21.3 [•]	93.6 [•]	21.2 [•]
3FEW	MinPC	26.0 [•]	114 [•]	2.26	120	15.9 [•]	103	19.1 [•]	23.9 [•]	-5E+05 [•]	6.26	120 [•]	31.4 [•]	52.8 [•]	24.0 [•]
3FEW	Best θ_N^2	12.5	4.58 [•]	-4.45	136	-8.83	166 [•]	15.9 [○]	19.1 [•]	-6E+05 [•]	6.75	148 [•]	21.8 [•]	78.8 [•]	20.4 [•]
3FEW	MaxW	-124 [•]	-305 [•]	-157 [•]	-650 [•]	-139 [•]	-361 [•]	-7.46 [•]	23.5 [•]	-2E+05 [•]	8.38 [○]	211 [•]	12.2 [○]	131 [•]	19.2
EW	All	8.11	2.80	8.00	3.98	7.13	5.61	13.5	8.11	2.80	8.00	3.98	7.13	5.61	13.5
EW	Rand	8.04 [•]	2.73 [•]	7.90 [•]	3.89 [•]	7.04 [•]	5.54 [•]	13.4 [•]	8.04 [•]	2.73 [•]	7.90 [•]	3.89 [•]	7.04 [•]	5.54 [•]	13.4 [•]
EW	MaxSR	8.77 [○]	4.72 [•]	8.49	5.97 [•]	7.40	6.25 [•]	11.2 [•]	8.77 [○]	4.72 [•]	8.49	5.97 [•]	7.40	6.25 [•]	11.2 [•]
EW	MinSR	7.39 [•]	0.83 [•]	7.55 [○]	2.75 [•]	6.77	5.18	14.8 [•]	7.39 [•]	0.83 [•]	7.55 [○]	2.75 [•]	6.77	5.18	14.8 [•]
EW	BWSR	7.94	2.42 [•]	7.72	2.86 [•]	7.35	5.28	14.7 [•]	7.94	2.42 [•]	7.72	2.86 [•]	7.35	5.28	14.7 [•]
EW	MaxVar	7.95	1.38 [•]	7.97	3.52	6.68	5.31	14.8 [•]	7.95	1.38 [•]	7.97	3.52	6.68	5.31	14.8 [•]
EW	MinVar	8.60	4.23 [•]	8.18	4.36	7.44	5.86	11.2 [•]	8.60	4.23 [•]	8.18	4.36	7.44	5.86	11.2 [•]
EW	BWVar	7.60 [•]	2.34 [•]	7.52 [•]	3.99	6.88	5.58	14.1 [•]	7.60 [•]	2.34 [•]	7.52 [•]	3.99	6.88	5.58	14.1 [•]
EW	MinPC	7.87	2.48 [•]	7.20 [•]	5.00 [•]	6.97	6.22 [•]	12.4 [•]	7.87	2.48 [•]	7.20 [•]	5.00 [•]	6.97	6.22 [•]	12.4 [•]
EW	Best θ_N^2	7.88 [○]	2.72	7.92	3.82	7.02	6.03 [•]	13.5	8.08	2.79	8.16	3.89	6.79 [•]	5.55	13.5
EWRF	All	0.63	-1.55	0.59	-1.84	1.00	0.59	16.4	0.63	-1.55	0.59	-1.84	1.00	0.59	16.4
SGMV		1.76	-49.7	4.25	-0.34	-2.28	-4.67	1.25	10.1	2.38	6.52	3.88	7.66	6.04	6.78
GMVRF		1.81	-26.3	-0.98	-11.4	-12.3	-5.34	-5.70	3.81	-3E+34	1.25	0.19	2.72	0.82	2.13
SMV		-4E+05	-3E+11	-1E+06	-6E+07	-8E+08	-8E+07	-9E+04	-994	-1E+44	-974	-2115	-2744	-747	-1486
SMV-ST		15.8 [•]	-108 [•]	11.7 [•]	-579 [•]	-2.91 [•]	0.88 [•]	0.77 [•]	21.6 [•]	-349 [•]	7.03 [•]	-243 [•]	-32.2 [•]	9.12 [•]	1.17 [•]
SMV-HT		-35.3 [•]	-302 [•]	-16.8 [•]	-105 [•]	-25.9 [•]	-143 [•]	-7.62 [•]	-28.0 [•]	-4E+32 [•]	-29.6 [•]	-421 [•]	-288 [•]	-84.8 [•]	-22.1 [•]
F+A		-232	-158	-368	-338	-138	-138	-110	-253	-5E+09	-394	-344	-2668	-61.0	-171

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 11 portfolio strategies in Table 2, the 11 asset selection rules in Table 3, across the seven datasets in Table 1. The table is constructed following the out-of-sample methodology described in Section 5.4. We estimate the portfolios either with the sample covariance matrix or the nonlinear shrinkage estimator of Ledoit and Wolf (2020). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p -values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under each selection rule relative to the 'All' selection rule as a benchmark, using the block bootstrap methodology described in Section 5.4. We also report p -values for SMV-HT and SMV-ST, using SMV as benchmark. The symbols ○, ●, and ● indicate that the p -value is less than 10%, 5%, and 1%, respectively.

Table OA.2: Annualized net out-of-sample utility of the size-optimized SMV portfolio

Port- folio strat.	Asset select. rule	Sample covariance matrix								nonlinear shrinkage covariance matrix							
		Dataset								Dataset							
		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO	
SMV	All	-4E+05	-3E+11	-1E+06	-6E+07	-8E+08	-8E+07	-9E+04		-1443	-107	-1399	-2360	-1457	-1143	-1844	
SMV	Rand	-5.75 [•]	-35.1 [•]	-1998 [•]	-151 [•]	-8.59 [•]	-105 [•]	-2.23 [•]		-1.26 [•]	27.9 [•]	-21.3 [•]	10.0 [•]	-4.79 [•]	-11.8 [•]	2.00 [•]	
SMV	MaxSR	-17.3 [•]	-25.3 [•]	-2011 [•]	-38.5 [•]	-27.2 [•]	-21.5 [•]	-60.9 [•]		-15.8 [•]	0.93	-34.6 [•]	-37.6 [•]	-25.7 [•]	-19.8 [•]	-42.8 [•]	
SMV	MinSR	-16.0 [•]	-2.04 [•]	-2007 [•]	-43.4 [•]	-12.8 [•]	-17.9 [•]	-36.2 [•]		-14.5 [•]	18.2	-31.7 [•]	-13.2 [•]	-13.1 [•]	-17.8 [•]	-38.2 [•]	
SMV	BWSR	-50.9 [•]	18.5 [•]	-2050 [•]	-62.2 [•]	-54.0 [•]	-54.6 [•]	-145 [•]		-33.3 [•]	95.1 [◊]	-63.4 [•]	-8.63 [•]	-44.0 [•]	-34.3 [•]	-105 [•]	
SMV	MaxVar	-5.94 [•]	-57.3 [•]	-1977 [•]	-26.4 [•]	-8.54 [•]	-18.4 [•]	-7.05 [•]		-3.90 [•]	47.1	-9.97 [•]	-23.6 [•]	-9.36 [•]	-18.6 [•]	-7.02 [•]	
SMV	MinVar	-22.1 [•]	-35.1 [•]	-1998 [•]	-52.4 [•]	0.22 [•]	14.7 [•]	-6.33 [•]		-16.8 [•]	4.38	-13.0 [•]	30.3 [•]	7.30 [•]	25.7 [•]	5.50 [•]	
SMV	BWVar	0.73 [•]	-83.2 [•]	-1986 [•]	-50.3 [•]	-14.3 [•]	-7.33 [•]	-1.63 [•]		2.36 [•]	-1.50	-8.42 [•]	-13.2 [•]	-8.47 [•]	-0.28 [•]	3.36 [•]	
SMV	MinPC	-19.3 [•]	-49.7 [•]	-2004 [•]	-12.7 [•]	-10.9 [•]	-2.23 [•]	-15.2 [•]		-16.2 [•]	-11.9	-30.8 [•]	-11.1 [•]	-10.4 [•]	-2.07 [•]	2.85 [•]	
SMV	Best θ_N^2	-0.98 [•]	34.2 [•]	-1996 [•]	22.1 [•]	-44.8 [•]	11.4 [•]	-17.9 [•]		9.97 [•]	34.7	-30.6 [•]	39.9 [•]	-16.1 [•]	3.14 [•]	-5.26 [•]	
SMV	MaxW	-15.3 [•]	-82.5 [•]	-2021 [•]	-76.2 [•]	-16.3 [•]	-78.4 [•]	-6.24 [•]		-17.5 [•]	50.9	-45.4 [•]	22.5 [•]	-0.62 [•]	-14.5 [•]	-43.4 [•]	
	SMV-ST	15.8 [•]	-108 [•]	11.7 [•]	-579 [•]	-2.91 [•]	0.88 [•]	0.77 [•]		23.9 [•]	67.8	5.13 [•]	-195 [•]	10.4 [•]	16.4 [•]	1.43 [•]	
	SMV-HT	-35.3 [•]	-302 [•]	-16.8 [•]	-105 [•]	-25.9 [•]	-143 [•]	-7.62 [•]		-29.9 [•]	-52.1	-17.8 [•]	-376 [•]	-33.6 [•]	-349 [•]	-21.1 [•]	

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the sample mean-variance portfolio across the seven datasets in Table 1 and the 11 asset selection rules in Table 3, as well as the soft and hard-thresholding versions of the SMV portfolio, SMV-ST and SMV-HT, described in Section 5.2. The table is constructed following the empirical methodology described in Section 4.3. We estimate the portfolios either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

sizes in Figure OA.12, averaged across the first six datasets in Table 1; we do not include the 100STO dataset because it consists of 100 datasets of 100 stocks, and thus, would be overweighted in the average. We observe that the portfolio sizes for the SMV portfolio are small on average compared to M , and that $\hat{N}_{st,t}$ is much more variable than $\hat{N}_{smv,t}^*$ and $\hat{N}_{ht,t}$.

OA.7.4 Out-of-sample Sharpe ratio

In Table OA.3, we report the annualized out-of-sample Sharpe ratio net of proportional transaction costs of the portfolio strategies considered in Table 4,

$$SR_k = \sqrt{12} \times \frac{\hat{\mu}_k}{\hat{\sigma}_k}, \quad (\text{OA19})$$

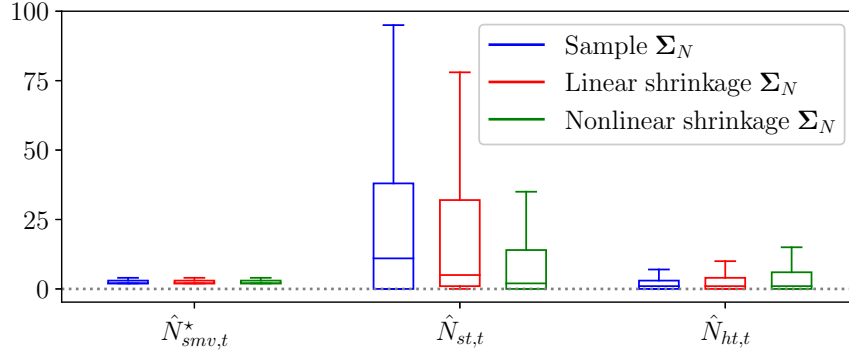
where $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the sample mean and standard deviation of the out-of-sample net portfolio returns, $r_{net,k,t}$ in (25). The main conclusions drawn from the out-of-sample utility are robust to using the out-of-sample Sharpe ratio. In particular, it is in most cases preferable to reduce the portfolio size in the 2F, 3FGMV, and 3FEW portfolio combination rules.

Table OA.3: Annualized net out-of-sample Sharpe ratio (in percentage points)

Port- folio strat.	Asset select. rule	Sample covariance matrix							nonlinear shrinkage covariance matrix						
		Dataset							Dataset						
		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO
2F	All	24.7	-150	-26.8	149	-6.46	113	-7.55	89.4	214	40.1	244	72.6	209	8.12
2F	Rand	54.0●	74.5●	16.7●	180●	45.5●	175●	22.2●	69.6●	203●	19.7●	210●	58.6●	180●	28.2●
2F	MaxSR	19.4	59.5●	-1.83	97.7●	-0.85	44.7●	50.3●	19.8●	162●	-2.01●	81.5●	12.5●	55.4●	52.0●
2F	MinSR	72.3●	121●	38.3●	171	70.6●	168●	-20.5●	86.0	215	44.4	195●	77.5	180○	-23.0●
2F	BWSR	52.5●	130●	33.4●	177	65.9●	220●	17.9●	69.6○	215	32.1	223●	67.1	200	21.8●
2F	MaxVar	71.4●	163●	60.7●	157	72.7●	135	-22.1●	83.8	238●	50.7	197●	74.1	159●	-17.6●
2F	MinVar	32.0	65.4●	-13.6	211●	23.7○	212●	36.7●	38.1●	179●	-2.61●	225○	39.1●	206	49.9●
2F	BWVar	56.3●	134●	26.7●	200●	37.2●	198●	24.7●	75.8	217	33.0	214●	52.7	190	32.8●
2F	MinPC	74.2●	150●	12.3●	167	71.9●	140	25.8●	89.9	228○	1.89●	176●	79.4	137●	36.6●
2F	Best θ_N^2	39.7	75.4●	13.6●	193○	66.3●	169●	25.8●	81.7	208	9.13	228●	79.3	198●	33.6●
2F	MaxW	51.5●	3.14●	0.47●	171	29.4●	177●	12.2●	87.5	233●	28.7	238	63.2	210	30.6●
3FGMV	All	26.9	-151	-14.2	149	-3.63	112	4.79	88.5	215	57.0	246	80.5	208	39.4
3FGMV	Rand	60.8●	75.0●	35.4●	180●	52.5●	175●	33.9●	72.9●	204●	48.6●	211●	66.3●	177●	46.3●
3FGMV	MaxSR	9.59	59.1●	8.19	94.7●	15.6	46.8●	59.1●	18.2●	161●	27.7○	83.3●	35.7●	53.0●	70.0●
3FGMV	MinSR	89.9●	120●	50.8●	173	79.3●	169●	20.0●	104	215	55.8	200●	91.4	183	28.2●
3FGMV	BWSR	59.8●	128●	39.1●	181○	74.5●	225●	37.6●	77.9	215	57.3	226○	80.1	198	49.5●
3FGMV	MaxVar	71.1●	164●	70.7●	159	74.0●	136	4.75	89.8	238●	62.2	197●	76.8	156●	15.9●
3FGMV	MinVar	41.4	66.2●	-18.3	213●	32.6●	209●	54.4●	42.5●	179●	5.01●	225○	48.6●	202	68.4●
3FGMV	BWVar	74.8●	133●	37.9●	201●	53.3●	202●	38.2●	87.8	216	59.0	218●	65.2	186○	52.1●
3FGMV	MinPC	81.6●	151●	40.5●	167	80.9●	139	29.6●	88.2	229	45.7	177●	88.3	136●	45.9●
3FGMV	Best θ_N^2	53.7○	71.0●	29.6●	172	42.7●	182●	34.3●	61.1●	196●	59.5	227○	73.7	162●	47.0●
3FGMV	MaxW	47.3●	3.59●	13.5●	165	35.7●	182●	26.1●	93.8	233●	56.7	242	80.8	208	50.2●
3FEW	All	38.9	-117	12.1	151	15.9	116	58.6	40.2	5.52	32.7	135	31.8	66.6	63.9
3FEW	Rand	59.6●	74.6●	35.9●	180●	52.7●	176●	59.8●	58.6●	168●	38.8●	203●	52.5●	161●	64.3●
3FEW	MaxSR	41.8	62.3●	33.9●	98.6●	27.0	51.2●	71.2●	40.2	85.9●	36.9	67.3●	38.3○	49.6○	70.6●
3FEW	MinSR	70.5●	119●	49.2●	170	70.9●	171●	47.1●	61.4●	175●	38.4	187●	57.3●	167●	50.2●
3FEW	BWSR	58.4○	130●	45.7●	178	73.7●	218●	57.4	64.4●	190●	44.5○	214●	66.8●	181●	64.5
3FEW	MaxVar	73.3●	161●	71.6●	162	69.8●	144	45.8●	69.1●	212●	50.7●	191●	60.5●	148●	52.4●
3FEW	MinVar	45.6	66.5●	23.1	211●	44.2●	215●	72.5●	45.7	118●	34.7	217●	44.9●	172●	75.0●
3FEW	BWVar	60.8○	133●	37.3●	204●	43.2○	201●	63.8●	62.0●	191●	44.6○	210●	44.8○	175●	66.4●
3FEW	MinPC	75.3●	152●	38.2●	170	72.4●	146	68.2●	66.8●	198●	36.6	176○	65.7●	120●	72.2●
3FEW	Best θ_N^2	56.9○	69.4●	29.6○	178	37.5○	182●	60.8○	50.8●	157●	45.5●	218●	51.7●	143●	65.3●
3FEW	MaxW	50.6	2.91●	10.4	168	24.9	175●	48.9●	70.7●	154●	39.8	237●	34.1	205●	69.6●
EW	All	53.0	24.1	52.7	31.9	50.4	41.5	81.4	53.0	24.1	52.7	31.9	50.4	41.5	81.4
EW	Rand	52.6●	23.7●	52.1●	31.4●	49.8●	41.1●	80.3●	52.6●	23.7●	52.1●	31.4●	49.8●	41.1●	80.3●
EW	MaxSR	57.6●	35.0●	55.6○	44.6●	52.9	45.6●	76.8●	57.6●	35.0●	55.6○	44.6●	52.9	45.6●	76.8●
EW	MinSR	48.1●	14.4●	49.7●	24.5●	47.2○	38.5○	79.5●	48.1●	14.4●	49.7●	24.5●	47.2○	38.5○	79.5●
EW	BWSR	52.0	22.2●	50.9○	25.1●	51.4	39.4○	85.8●	52.0	22.2●	50.9○	25.1●	51.4	39.4○	85.8●
EW	MaxVar	49.5●	17.2●	49.9○	28.3●	45.0●	38.1●	76.2●	49.5●	17.2●	49.9○	28.3●	45.0●	38.1●	76.2●
EW	MinVar	58.9●	32.9●	56.7●	35.4○	55.6●	45.0○	80.3	58.9●	32.9●	56.7●	35.4○	55.6●	45.0○	80.3
EW	BWVar	50.0●	21.8●	49.7●	32.1	49.0	41.4	83.4●	50.0●	21.8●	49.7●	32.1	49.0	41.4	83.4●
EW	MinPC	52.4	22.5●	48.9●	38.9●	50.0	45.5●	82.3○	52.4	22.5●	48.9●	38.9●	50.0	45.5●	82.3○
EW	Best θ_N^2	51.4●	23.7	52.2	31.1	49.7	44.1●	80.9	53.9	23.8	51.6	31.7	49.2	43.5●	80.9
EWRF		23.4	1.77	23.1	1.81	24.4	17.1	58.1	23.4	1.77	23.1	1.81	24.4	17.1	58.1
SGMV		19.3	-193	29.6	7.82	6.33	-6.63	17.1	82.0	31.1	68.0	42.4	77.5	59.6	61.0
GMVRF		23.8	-177	9.52	-27.6	-10.4	0.81	8.19	65.6	-1.26	53.0	5.35	50.9	14.0	41.6
SMV		-50.2	-29.7	-14.7	-91.5	-16.5	-76.9	-25.8	91.2	240	40.0	243	72.1	206	10.9
SMV-ST		77.5●	167●	64.0●	163●	78.9●	177●	20.7●	77.0	175●	43.8	157●	65.5	147●	19.6●
SMV-HT		36.0●	-97.3●	29.6●	24.7●	20.7	46.4●	25.6●	21.2●	103●	18.8	128●	15.6●	108●	24.4●
F+A		98.9	189	43.9	157	70.5	164	63.8	99.6	189	44.6	147	71.9	151	64.2

Notes. This table reports the annualized net out-of-sample Sharpe ratio, in percentage points, for the 11 portfolio strategies in Table 2, the 11 asset selection rules in Table 3, across the seven datasets in Table 1. The table is constructed following the out-of-sample methodology described in Section 5.4. We estimate the portfolios either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p -values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under each selection rule relative to the 'All' selection rule as a benchmark, using the block bootstrap methodology described in Section 5.4. We also report p -values for SMV-HT and SMV-ST, using SMV as benchmark. The symbols ○, ●, and ● indicate that the p -value is less than 10%, 5%, and 1%, respectively.

Figure OA.12: Estimated portfolio sizes for the SMV portfolio



Notes. This figure depicts boxplots of the portfolio sizes obtained in empirical data, across different months t , for the SMV portfolio with three different strategies as detailed in Section OA.7.3. First, our estimated portfolio size for the SMV portfolio, $\hat{N}_{smv,t}^*$. Second, the portfolio size obtained under soft-thresholding, $\hat{N}_{st,t}$ in (OA18). Third, the portfolio size obtained under hard-thresholding, $\hat{N}_{ht,t}$. We consider the sample covariance matrix (5) in blue and the linear and nonlinear shrinkage estimators of Ledoit and Wolf (2004, 2017) in red and green, respectively. The boxplots are obtained by averaging across the first six datasets listed in Table 1.

OA.7.5 Portfolio turnover with asset selection rules

In the empirical analysis of Section 5, we consider different asset selection rules to select in which N assets to invest when reducing the portfolio size from M to N . The portfolio turnover is impacted by the changes to the investment universe over time implied by these selection rules. The larger the number of changes in selected assets in a given period, the lower the *stability* of the selection rule, and the higher the portfolio turnover.

The impact of a selection rule on portfolio turnover also depends the *selection frequency*, which is the frequency at which we decide to change the optimal N and the selected assets. In the empirical analysis of Section 5, we choose to rebalance the portfolio weights every month but to change the optimal N and the selected assets once a year. The rationale is to avoid excessive changes to the investment universe and make our method less costly and more practically implementable. However, we could choose a different selection frequency, and it is important to test the robustness of our results to this choice.

Therefore, we have two objectives in this section. First, we investigate the stability of the different asset selection rules we consider and how they impact the portfolio turnover. Second, we evaluate the robustness of our results to using a selection frequency of once every month, six months, and 24 months instead of 12 months.

We begin by defining a measure of selection rule stability. To do so, note that except for Rand and Best θ_N^2 , each selection rule listed in Table 3, which is applied to the 2F, 3FGMV, 3FEW, and EW portfolio strategies, is implemented in a similar way following three steps. First, for each selection rule k , we compute at time t the $M \times 1$ vector of parameters on which the selection rule is based, which we denote as $\delta_{k,t} = (\delta_{1,k,t}, \delta_{2,k,t}, \dots, \delta_{M,k,t})$. For instance, if we consider a selection rule based on the in-sample Sharpe ratio (MaxSR, MinSR, or BWSR), then $\delta_{k,t} = \hat{\theta}_t$ where $\hat{\theta}_t = (\hat{\theta}_{1,t}, \hat{\theta}_{2,t}, \dots, \hat{\theta}_{M,t})$ is the $M \times 1$ vector of in-sample Sharpe ratios.

Second, we rank assets from 1 to M depending on their parameter values $\delta_{k,t}$, where 1 represents the highest value and M the lowest. Specifically, we define the $M \times 1$ vector of ranks $\phi_{k,t} = (\phi_{1,k,t}, \phi_{2,k,t}, \dots, \phi_{M,k,t})$, where $\phi_{i,k,t}$ denotes the rank of asset i based on $\delta_{k,t}$,

$$\phi_{i,k,t} = \sum_{j=1}^M \mathbb{1}_{\{\delta_{i,k,t} \leq \delta_{j,k,t}\}}. \quad (\text{OA20})$$

Third, we select assets depending on their rank $\phi_{i,k,t}$ and the type of selection rule. For that purpose, we denote $y_{i,k,t}$ the binary variable indicating whether, for selection rule k , asset i is selected at time t : $y_{i,k,t} = 1$ if asset i is selected, and 0 otherwise. We consider four types of asset selection rules in Table 3. For the Rand rule, $y_{i,k,t} = 0$ or 1 at random, subject to $\sum_{i=1}^M y_{i,k,t} = N$. For the MaxSR, MaxVar, and MaxW rules that select the N assets with the highest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} \leq N\}}$. For the MinSR, MinVar and MinPC rules that select the N assets with the lowest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} > M-N\}}$. Finally, for the BWSR and BWVar rules

that select the $\lceil N/2 \rceil$ assets with the highest $\delta_{i,k,t}$ values and the $\lfloor N/2 \rfloor$ assets with the lowest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} \leq \lceil N/2 \rceil \text{ or } \phi_{i,k,t} > M - \lfloor N/2 \rfloor\}}$.^{OA5}

Given this three-step process with which the asset selection rules are implemented, we want to evaluate the impact on portfolio turnover of changes in the set of selected assets. We first introduce *selection turnover*, which measures, for a portfolio k , the number of changes in selected assets from month $t - 1$ to t :

$$\text{selectionturnover}_{k,t} = \sum_{i=1}^M |y_{i,k,t} - y_{i,k,t-1}|, \quad t = T + 1, \dots, T_{tot}. \quad (\text{OA21})$$

The less stable the asset selection rule and the higher the selection frequency, the higher the selection turnover, and the more we expect an increase in portfolio turnover due to changes in selected assets. To precisely measure the impact of selection turnover on the overall portfolio turnover, we define $\text{turnover}_{k,t}^{\text{selection}}$ as the turnover of portfolio k coming from those assets that were selected at time t but not at time $t - 1$, and vice-versa:

$$\text{turnover}_{k,t}^{\text{selection}} = \sum_{i: y_{i,k,t} \neq y_{i,k,t-1}} |w_{i,k,t} - w_{i,k,(t-1)+}|, \quad t = T + 1, \dots, T_{tot}, \quad (\text{OA22})$$

where $w_{i,k,t}$ is the weight of asset i at month t and $w_{i,k,(t-1)+}$ is the prior-time weight before rebalancing at month t . By construction, $\text{turnover}_{k,t}^{\text{selection}} \leq \text{turnover}_{k,t}$, where $\text{turnover}_{k,t}$ in (26) is the total turnover computed on all assets. Finally, we also define the proportion of the total turnover that originates from changes in the selected assets:^{OA6}

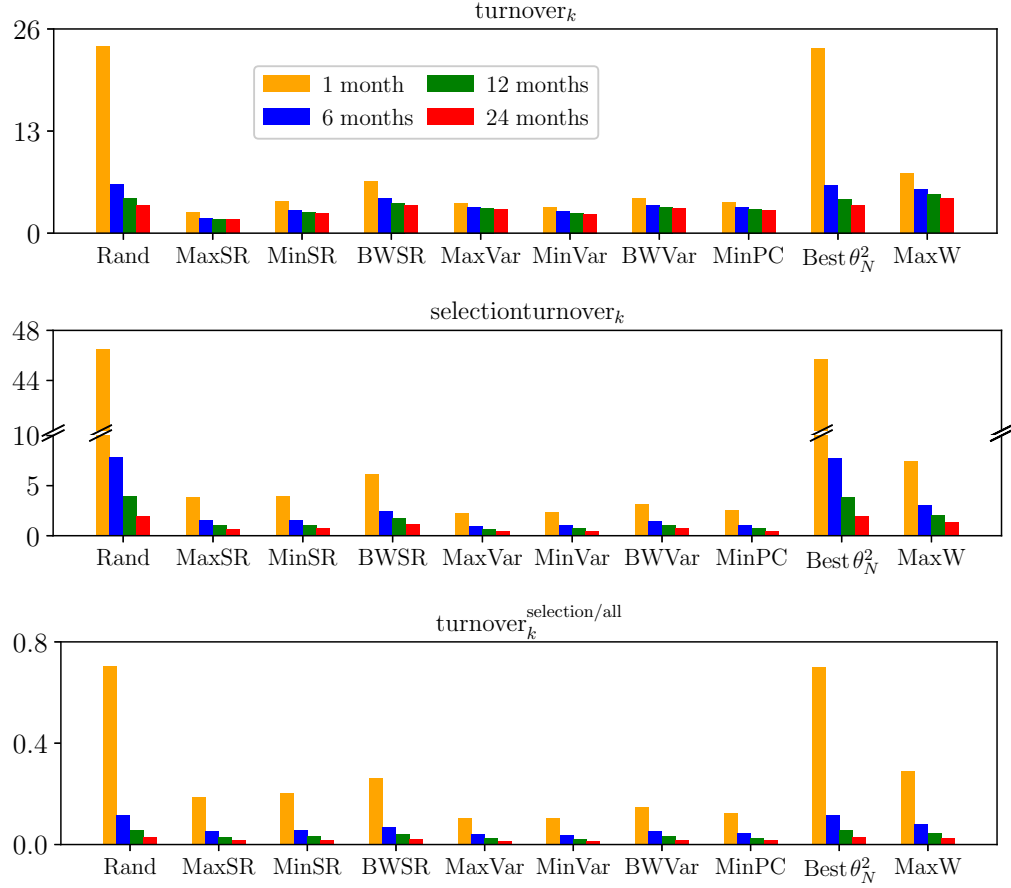
$$\text{turnover}_{k,t}^{\text{selection/all}} = \frac{\text{turnover}_{k,t}^{\text{selection}}}{\text{turnover}_{k,t}} \leq 1, \quad (\text{OA23})$$

which we expect to increase with $\text{selectionturnover}_{k,t}$.

^{OA5}The Best θ_N^2 selection rule is an exception as it is not based on a ranking of assets $\phi_{k,t}$. Instead, in that case, $y_{i,k,t} = 1$ if at time t , asset i is included in the randomly drawn selection of size N which corresponds to the 95% quantile of estimated values of θ_N^2 , as detailed in Section 5.3. Otherwise, $y_{i,k,t} = 0$.

^{OA6}The measures $\text{selectionturnover}_{k,t}$ and $\text{turnover}_{k,t}^{\text{selection}}$ are non-zero only when the investment universe changes from $t - 1$ to t , which happens according to the chosen selection frequency (1, 6, 12, or 24 months).

Figure OA.13: Portfolio turnover with asset selection rules and different selection frequencies



Notes. This figure depicts different turnover metrics for the two-fund portfolio rule across the ten asset selection rules in Table 3 (we exclude the ‘All’ selection rule). The top panel depicts the average of the monthly portfolio turnover, defined in (26). The middle panel depicts the average of the monthly selection turnover, defined in (OA21), i.e., the average of the number of changes in selected assets each month. The bottom panel depicts the average of the proportion of the total turnover coming from changes in selected assets, defined in (OA22). These metrics are averaged across the first six datasets in Table 1. We report them for four different selection frequencies: every 1, 6, 12, and 24 months. The selection frequency is the frequency with which we change the optimal portfolio size N and the selected assets. We estimate the two-fund rule with the linear shrinkage covariance matrix of Ledoit and Wolf (2004) and a risk-aversion coefficient of $\gamma = 1$. The figure is constructed following the out-of-sample methodology described in Section 5.4. The middle panel includes an axis break for better visibility.

In Figure OA.13, we depict the portfolio turnover in (26), the selection turnover in (OA21), and the proportion of portfolio turnover explained by changes in the selected assets in (OA23). We average these measures across all months t and across the first six datasets described in Table 1. We do not include the 100STO dataset because it consists of 100 datasets of 100 stocks, and thus, would be overweighted in the average. We report results for each of the four considered selection frequencies, i.e., every 1, 6, 12, and 24 months. We only consider the 2F portfolio combination rule for conciseness; the results for the 3FGMV and 3FEW portfolios are similar. Moreover, we only report the results for the case where we use the linear shrinkage covariance matrix of Ledoit and Wolf (2004) because, as discussed in Section 5.5, it leads to a better out-of-sample performance, and also a lower portfolio turnover.

We make several observations from Figure OA.13. First, we look at the stability of the different asset selection rules. The middle panel of Figure OA.13 shows that the Rand and Best θ_N^2 selection rules have the largest selection turnover, especially for a monthly selection frequency. This is expected as these two rules are based on random selections. Then, the selection rule MaxW has the second largest selection turnover because it ranks assets according to portfolio weights and these weights can vary substantially from one period to another. Then, selection rules that rank assets based on the marginal Sharpe ratio are also quite unstable because they depend on estimates of expected returns that are notoriously noisy.^{OA7} In contrast, selection rules built on the variance and PCA are remarkably stable, with the set of selected assets changing by only one or two assets on average each month when we use a monthly selection frequency. Finally, we also observe that

^{OA7}The number of changes in selected assets can be substantial. For instance, on average across the first six datasets, the selection turnover of BWSR in Figure OA.13 is equal to 5.2 when using a monthly selection frequency. This means that, on average, 5.2 assets enter or exit the set of selected assets from one month to another. Given that the number of selected assets N dictated by our theory is close to $T/2 = 60$, this represents a 8.67% change of the investment universe each month on average.

the best-worst asset selection rules, BWSR and BWVar, are more unstable because they select the assets with the most extreme rankings, which are also subject to more estimation errors.

Second, the bottom panel of Figure OA.13 shows that a higher selection turnover is indeed linked with a higher proportion of portfolio turnover originating from assets that are selected in the portfolio at time t but not at time $t - 1$, and vice-versa.

Third, the bottom panel of Figure OA.13 shows that, except for Rand and Best θ_N^2 , the proportion of the total portfolio turnover explained by assets that enter and exit the set of selected assets each month is around 10% to 30% on average when we use a monthly selection frequency but, as expected, decreases substantially as we decrease the selection frequency. This proportion of 10% to 30% with a monthly selection frequency is not negligible, particularly given that the selection turnover in the middle panel of Figure OA.13 is generally well below 10% of the total number of selected assets. Still, it means that the majority of the portfolio turnover comes from changes in weights allocated to assets that remain selected from one period to another, particularly as we decrease the selection frequency.

We now turn to our second objective, which is to evaluate the robustness of our out-of-sample performance results to the choice of the selection frequency. Figure OA.13 confirms that decreasing the selection frequency, and thus changing the investment universe less frequently, significantly reduces selection turnover and thus the overall portfolio turnover. However, this does not automatically mean that decreasing the selection frequency improves the net out-of-sample performance because the monthly frequency could deliver better gross performance. To test the impact of the selection frequency on the net out-of-sample portfolio performance, we report in Table OA.4 the annualized net out-of-sample utility of the 2F portfolio implemented with the 10 asset selection rules in Table 3 and a selection frequency of 1, 6, 12, and 24 months. The unreported results for

the 3FGMV and 3FEW portfolio combination rules are similar. We make three main observations from Table OA.4.

First, in the vast majority of cases, using a lower selection frequency than monthly is beneficial, and the performance gain can be substantial. For instance, consider the MinSR rule for the 96S-BM dataset and the sample covariance matrix. In this case, a monthly selection frequency delivers an annualized net utility of 19.3 whereas the 6, 12, and 24-month frequencies deliver a net utility of 25.5, 25.4 and 25.4, respectively. A similar observation holds for the linear shrinkage covariance matrix, but with lower differences in performance. This is because portfolios implemented with the linear shrinkage covariance matrix yield a lower turnover than those under the sample covariance matrix, and thus, benefit relatively less from an extra reduction in turnover.

Second, the 6, 12, and 24-month selection frequencies deliver a similar net out-of-sample performance overall, which means that the results under a 12-month frequency in the main text are robust to other possible choices.

Third, asset selection rules that outperform investing in all assets do so under all considered selection frequencies, including monthly. This shows that our optimal diversification method is valuable regardless of the frequency at which the selection rules are implemented.

OA.7.6 Random asset selection rule with different frequencies

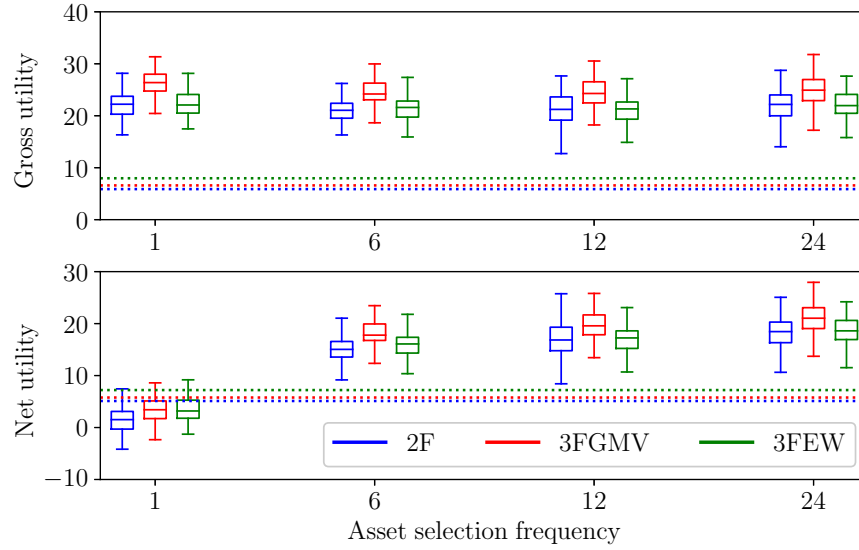
One of the asset selection rules we consider in the empirical analysis, as well as in the simulations, is the random asset selection rule. In this section, we further analyze the performance delivered by this selection rule when applied to our two-fund and three-fund rules. Specifically, in Figure OA.14, we report boxplots of the annualized gross and net out-of-sample utility delivered by the

Table OA.4: Net out-of-sample utility in empirical data with different selection frequencies

Port- folio strat.	Asset select. rule	Sample covariance matrix							nonlinear shrinkage covariance matrix						
		Dataset							Dataset						
		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO
All	n.a.	−0.31	−59.6	−15.3	99.4	−18.8	56.8	−5.47	5.11	1.20	1.09	41.6	1.09	17.7	0.16
Rand	1 month	−44.5 [●]	−533 [●]	−48.8 [●]	−74.2 [●]	−82.6 [●]	−97.8 [●]	−6.71 [●]	1.33 [●]	27.5 [●]	−8.39 [●]	87.1 [●]	−5.02 [●]	40.9 [●]	2.43 [●]
Rand	6 months	7.00 [●]	−48.8 [●]	−8.19 [●]	122 [●]	−8.27 [●]	119 [●]	−1.71 [●]	15.0 [●]	50.1 [●]	2.54 [●]	128 [●]	11.0 [●]	75.6 [●]	4.48 [●]
Rand	12 months	12.3 [●]	14.2 [●]	−7.17 [●]	145 [●]	1.46 [●]	149 [●]	−1.77 [●]	17.0 [●]	52.4 [●]	1.60	132 [●]	12.8 [●]	79.5 [●]	3.48 [●]
Rand	24 months	16.2 [●]	48.3 [●]	−0.71 [●]	156 [●]	5.96 [●]	158 [●]	−3.39 [●]	18.4 [●]	53.1 [●]	4.07 [●]	133 [●]	13.1 [●]	80.9 [●]	1.28 [●]
MaxSR	1 month	−5.80	−30.0 [○]	−7.72	2.25 [●]	−21.7	−35.2 [●]	8.51 [●]	−1.98 [●]	19.8 [●]	−1.94	21.4 [●]	0.66	7.66 [●]	13.1 [●]
MaxSR	6 months	0.77	3.62 [●]	−5.15	31.7	−9.86	−31.1 [●]	8.67 [●]	0.66 [●]	22.3 [●]	−2.71	26.5 [●]	3.72	10.3 [○]	13.0 [●]
MaxSR	12 months	1.38	11.6 [●]	−7.48	23.6 [○]	−10.5	−27.1 [●]	7.75 [●]	1.93 [○]	23.1 [●]	−6.10	23.1 [●]	0.69	10.8 [○]	12.2 [●]
MaxSR	24 months	2.02	16.2 [●]	−4.14	34.8 [○]	−13.3	−18.8 [●]	2.97 [●]	2.56 [○]	22.8 [●]	−1.54	22.4 [●]	−0.05	8.59 [●]	6.79 [●]
MinSR	1 month	19.3 [●]	25.9 [●]	6.84 [●]	123	19.2 [●]	112 [○]	−1.56 [●]	16.0 [●]	50.7 [●]	5.21 [○]	97.9 [●]	12.0 [●]	67.3 [●]	−0.61 [●]
MinSR	6 months	25.5 [●]	70.1 [●]	8.49 [●]	145	26.8 [●]	130 [●]	−1.22 [●]	17.4 [●]	53.4 [●]	6.54 [●]	98.3 [●]	14.8 [●]	70.3 [●]	−0.53 [●]
MinSR	12 months	25.4 [●]	73.2 [●]	7.10 [●]	146	24.9 [●]	139 [●]	−3.86 [●]	18.0 [●]	53.2 [●]	5.21 [●]	101 [●]	16.0 [●]	72.1 [●]	−2.49 [●]
MinSR	24 months	25.4 [●]	85.3 [●]	9.09 [●]	167	19.3 [●]	150 [●]	−4.58	19.6 [●]	53.9 [●]	6.28 [●]	107 [●]	15.2 [●]	74.7 [●]	−2.77 [●]
BWSR	1 month	−2.88	14.3 [●]	−14.5	161	−8.83	168 [●]	−28.7 [●]	16.5 [○]	61.8 [●]	3.02	140 [●]	18.8 [●]	88.8 [●]	−8.27 [●]
BWSR	6 months	0.55	66.1 [●]	−2.22	154	−4.83	232 [●]	−28.2 [●]	18.1 [●]	63.1 [●]	8.21	143 [●]	20.1 [●]	95.8 [●]	−8.21 [●]
BWSR	12 months	7.39	84.4 [●]	−4.42	133	11.7 [○]	242 [●]	−29.9 [●]	19.2 [●]	63.8 [●]	4.74	143 [●]	18.8 [●]	99.5 [●]	−11.8 [●]
BWSR	24 months	16.0	80.7 [●]	−0.23	96.1	10.8 [○]	219 [●]	−20.1 [●]	21.7 [●]	64.0 [●]	8.18	133 [●]	13.0	103 [●]	−9.57 [●]
MaxVar	1 month	27.7 [●]	112 [●]	10.0 [●]	112	20.9 [●]	76.9	−5.45	26.4 [●]	69.1 [●]	9.46 [●]	127 [●]	20.5 [●]	77.6 [●]	−1.16 [●]
MaxVar	6 months	26.5 [●]	127 [●]	17.0 [●]	118	21.9 [●]	73.3	−5.30	24.1 [●]	70.0 [●]	10.6 [●]	125 [●]	22.6 [●]	78.0 [●]	−1.91 [●]
MaxVar	12 months	25.0 [●]	124 [●]	17.9 [●]	114	21.7 [●]	83.8	−7.25 [●]	23.2 [●]	70.5 [●]	9.99 [●]	122 [●]	21.1 [●]	77.7 [●]	−3.62 [●]
MaxVar	24 months	29.8 [●]	129 [●]	18.6 [●]	113	19.5 [●]	91.9	−5.73	25.1 [●]	69.8 [●]	12.0 [●]	120 [●]	18.0 [●]	78.5 [●]	−3.73 [●]
MinVar	1 month	1.97	−18.5 [●]	−14.5	215 [●]	−0.88 [○]	199 [●]	3.57 [●]	5.17	32.6 [●]	−2.38	136 [●]	7.27 [●]	80.1 [●]	14.6 [●]
MinVar	6 months	5.32	7.83 [●]	−10.5	213 [●]	0.57 [○]	222 [●]	4.60 [●]	6.30	32.7 [●]	−1.07	141 [●]	6.66 [●]	78.8 [●]	15.7 [●]
MinVar	12 months	2.47	12.5 [●]	−13.1	219 [●]	−1.88	225 [●]	−0.42 [●]	6.08	33.1 [●]	−5.25 [○]	140 [●]	5.40 [○]	81.4 [●]	12.4 [●]
MinVar	24 months	2.04	20.0 [●]	−12.6	227 [●]	−3.33	226 [●]	−0.11 [●]	5.68	33.8 [●]	−4.89 [○]	140 [●]	4.11	81.8 [●]	10.0 [●]
BWVar	1 month	19.4 [●]	75.0 [●]	−0.95 [○]	189 [●]	−4.76	151 [●]	−0.55 [●]	21.2 [●]	62.9 [●]	9.34 [○]	127 [●]	11.1 [●]	85.7 [●]	6.24 [●]
BWVar	6 months	17.5 [○]	90.4 [●]	3.84 [●]	206 [●]	−0.03	168 [●]	3.00 [●]	21.2 [●]	63.6 [●]	7.90	132 [●]	12.3 [●]	88.9 [●]	7.59 [●]
BWVar	12 months	13.0	89.3 [●]	−6.76	182 [○]	0.14	180 [●]	−0.53 [●]	19.8 [●]	63.4 [●]	5.39	133 [●]	10.6 [●]	89.9 [●]	5.38 [●]
BWVar	24 months	20.8 [●]	101 [●]	1.00 [○]	190 [○]	−0.89	190 [●]	−1.83 [●]	21.7 [●]	62.8 [●]	7.48	133 [●]	9.12 [○]	89.0 [●]	2.74 [●]
MinPC	1 month	25.7 [●]	92.8 [●]	−8.43	86.8	17.0 [●]	88.9	−3.18	24.2 [●]	66.2 [●]	−1.14	108 [●]	22.1 [●]	49.2 [●]	6.10 [●]
MinPC	6 months	30.9 [●]	105 [●]	−7.68	89.2	27.8 [●]	98.3	−1.16 [●]	25.4 [●]	67.8 [●]	−2.59	105 [●]	22.9 [●]	52.6 [●]	6.73 [●]
MinPC	12 months	27.4 [●]	112 [●]	−6.23	115	17.9 [●]	94.1	−1.01 [●]	24.9 [●]	67.7 [●]	−2.55	109 [●]	22.8 [●]	50.5 [●]	6.68 [●]
MinPC	24 months	20.8 [○]	106 [●]	−4.26	101	22.3 [●]	106 [○]	0.42 [●]	22.8 [●]	67.9 [●]	2.32	107 [●]	24.4 [●]	50.7 [●]	5.63 [●]
Best θ_N^2	1 month	−43.1 [●]	−525 [●]	−44.7 [●]	−138 [●]	−81.0 [●]	−88.0 [●]	−17.3 [●]	4.20	28.2 [●]	−7.66 [●]	83.8 [●]	−2.72	40.7 [●]	3.70 [●]
Best θ_N^2	6 months	4.06	−47.3	−3.26	159	−1.21	102	−12.9 [●]	19.6 [●]	52.5 [●]	5.82	120 [●]	16.6 [●]	72.9 [●]	4.04 [●]
Best θ_N^2	12 months	−1.22	14.7 [●]	−8.77	182 [○]	4.73	137 [●]	−9.71 [●]	24.4 [●]	59.2 [●]	−4.03	149 [●]	23.0 [●]	95.7 [●]	3.33 [●]
Best θ_N^2	24 months	14.6	50.1 [●]	−2.44	206 [●]	−2.04	143 [●]	−7.61	21.7 [●]	56.0 [●]	4.13	134 [●]	20.7 [●]	87.4 [●]	2.04
MaxW	1 month	−307 [●]	−798 [●]	−350 [●]	−740 [●]	−300 [●]	−543 [●]	−173 [●]	30.3 [●]	79.2 [●]	9.19	199 [●]	17.9 [○]	130 [●]	−12.9 [●]
MaxW	6 months	−171 [●]	−421 [●]	−234 [●]	−749 [●]	−151 [●]	−456 [●]	−111 [●]	29.2 [●]	79.4 [●]	3.28	203 [●]	19.4 [○]	131 [●]	−10.7 [●]
MaxW	12 months	−127 [●]	−305 [●]	−179 [●]	−581 [●]	−129 [●]	−366 [●]	−77.2 [●]	32.3 [●]	79.0 [●]	2.19	197 [●]	19.8 [○]	129 [●]	−13.5 [●]
MaxW	24 months	−97.6 [●]	−74.8	−115 [●]	−416 [●]	−56.3 [○]	−244 [●]	−39.6 [●]	32.2 [●]	77.4 [●]	3.59	195 [●]	16.0	126 [●]	−13.3 [●]

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 2F portfolio implemented with the 10 asset selection rules in Table 3 (we exclude the ‘All’ selection rule), across the seven datasets in Table 1. For each selection rule, we report results for four different selection frequencies: every 1, 6, 12, and 24 months. The selection frequency is the frequency with which we change the optimal portfolio size N and the selected assets. The table is constructed following the empirical methodology described in Section 5.4. We estimate the portfolios either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). We rebalance the portfolio weights every month. The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p -values for the statistical test of the difference between the utility of the 2F portfolio under the ‘All’ selection rule versus the 10 other selection rules, for all selection frequencies, using the block bootstrap methodology described in Section 5.4. The symbols ○, ●, and ● indicate that the p -value is less than 10%, 5%, and 1%, respectively.

Figure OA.14: Performance of the random asset selection rule for the 96S-BM dataset



Notes. This figure depicts boxplots of the annualized gross (top panel) and net (bottom panel) out-of-sample utility delivered by the two-fund and three-fund rules considered in the empirical analysis, implemented with the random asset selection rule on the 96S-BM dataset. Specifically, we depict boxplots of utilities for the 2F (blue), 3FGMV (red), and 3FEW (green) portfolio strategies, obtained across 100 repetitions of the out-of-sample analysis. We depict the utilities for different asset selection frequencies as in Section OA.7.5 of this Online Appendix, i.e., we change the selected assets in our rolling window method every 1, 6, 12, or 24 months. The horizontal dotted lines depict the utility of each portfolio strategy implemented on all M assets. We use a risk-aversion coefficient $\gamma = 1$ and estimate the portfolios using the linear shrinkage covariance matrix.

2F, 3FGMV, and 3FEW portfolio strategies under the estimated optimal portfolio size N , where the N assets are selected randomly. The boxplots are obtained across 100 repetitions of the out-of-sample analysis. We focus on the 96S-BM dataset and we consider different asset selection frequencies as in Section OA.7.5 of this Online Appendix, i.e., we change the selected assets in our rolling window method every 1, 6, 12, or 24 months.

We make three main observations from Figure OA.14. First, consistent with the simulation results in Section 4, selecting the N assets randomly, with N determined using our theory, systematically outperforms selecting all M assets in terms of gross utility. This is because this figure focuses on portfolio strategies subject to estimation risk. Second, the random asset selection rule also delivers a larger utility than choosing all assets when accounting for transaction costs when using an asset selection frequency of 6, 12 or 24 months. In contrast, changing the set of selected

assets every month increases turnover and transaction costs too significantly and underperforms investing in all M assets. Third, although the random asset selection rule outperforms selecting all assets, its performance is highly variable across different random selections and, as discussed in Section 5.5, it underperforms the majority of the more sensible selection rules in Table 3.

OA.7.7 Overlap between selection rules

We show in Section 5 of the main text that reducing the portfolio size is beneficial in most cases for the different selection rules listed in Table 3. These selection rules are different by construction to test whether our results hold for different types of selection rule. Thus, we raise the following question: do these rules indeed select dissimilar assets in practice?

To answer this question, we must measure the *overlap* between two selection rules, i.e., the number of assets jointly selected by both selection rules. Following Section OA.7.5, we denote by $\mathbf{y}_{k,t} = (y_{1,k,t}, y_{2,k,t}, \dots, y_{M,k,t})$ the $(M \times 1)$ selection vector for any selection rule k : $y_{i,k,t} = 1$ if asset i is selected in rule k at time t , and 0 otherwise. The overlap between rules k and l is then given by $\mathbf{y}'_{k,t} \mathbf{y}_{l,t} = \sum_{i=1}^M y_{i,k,t} y_{i,l,t}$.

We then compare the overlap between the selection rules listed in Table 3 to the overlap between two independent random selection rules. If the average overlap between our selection rules is close to or lower than that of a pair of two independent random selections, then it means that the selected assets are indeed dissimilar.

Consider two selection rules k and l that randomly select N_t assets out of M at time t . In the next proposition, we derive the expected overlap if k and l are independent.

Proposition OA.4 *Consider two independent selection rules that randomly select N_t assets out of*

M available ones. Then, the expected overlap between the two rules is

$$\bar{N}_{t,M} = M \frac{\binom{M-1}{N_t-1}}{\binom{M}{N_t}}. \quad (\text{OA24})$$

For instance, if two independent selection rules randomly select $N_t = 50$ assets out of $M = 100$ available ones, we expect them to jointly select $\bar{N}_{t,M} = 25$ assets.

Now, to identify how a pair of selection rules (k, l) selects assets relative to two independent random selection rules, we use the expected overlap in (OA24) to build a measure of the *excess overlap* of any (k, l) pair relative to that of a pair of independent random selection rules. We denote this measure $\text{eo}_{k,l,t}$ and define it as

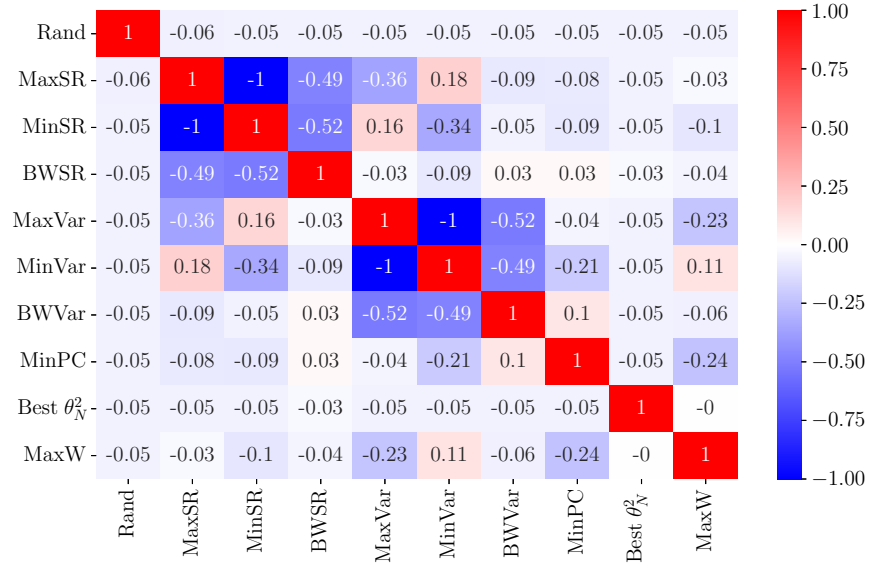
$$\text{eo}_{k,l,t} = (\mathbf{y}'_{k,t} \mathbf{y}_{l,t} - \bar{N}_{t,M}) \times \left(\frac{\mathbb{I}_{\{\mathbf{y}'_{k,t} \mathbf{y}_{l,t} - \bar{N}_{t,M} \geq 0\}}}{N_t - \bar{N}_{t,M}} - \frac{\mathbb{I}_{\{\mathbf{y}'_{k,t} \mathbf{y}_{l,t} - \bar{N}_{t,M} < 0\}}}{\max(0, 2N_t - M) - \bar{N}_{t,M}} \right). \quad (\text{OA25})$$

Several comments are in order. First, if $\mathbf{y}'_{k,t} \mathbf{y}_{l,t} = \bar{N}_{t,M}$, then $\text{eo}_{k,l,t} = 0$. This means that there is no excess overlap if the overlap between k and l is equal to the expected overlap between two independent random selection rules. Second, the excess overlap is positive (resp. negative) if there is more (resp. less) overlap between k and l compared to independent random rules, i.e., $\text{eo}_{k,l,t} > 0$ if $\mathbf{y}'_{k,t} \mathbf{y}_{l,t} > \bar{N}_{t,M}$ and vice-versa. Third, the measure of excess overlap is bounded between -1 and 1 . A value of 1 indicates the *maximum* overlap and is obtained if and only if k and l are *identical*, i.e., $\mathbf{y}_{k,t} = \mathbf{y}_{l,t}$ such that $\mathbf{y}'_{k,t} \mathbf{y}_{l,t} = N_t$. A value of -1 indicates the *minimum* overlap, i.e., $\mathbf{y}'_{k,t} \mathbf{y}_{l,t} = \max(0, 2N_t - M)$.^{OA8}

In Figure OA.15, we depict a heatmap of the excess overlap in (OA25) for all pairs of selection rules in Table 3. We average it across all months t and the first six datasets listed in Table 1; we do not include the 100STO dataset because it consists of 100 datasets of 100 stocks, and thus, would

^{OA8}If $M \geq 2N_t$, then it is possible for two rules to have zero overlap as each rule can select N_t different assets. However, if $M < 2N_t$, then the minimum possible overlap will be strictly positive and equal to $2N_t - M$ assets. For instance, if $M = 90$ and $N_t = 50$, two rules must have an overlap of at least $2N_t - M = 10$ assets.

Figure OA.15: Average excess overlap between selection rules



Notes. This figure depicts the average excess overlap, defined in (OA25), for each pair of selection rule listed in Table 3 (we exclude the ‘All’ rule). The excess overlap is averaged over the first six datasets in Table 1 and all estimation windows. We report the results for the optimal portfolio size of the two-fund rule and the sample covariance matrix.

be overweighted in the average. We only consider the optimal portfolio size of the two-fund rule for conciseness. Moreover, selection rules are based on sample estimates, and we report results under the sample covariance matrix.

Two main observations can be made from Figure OA.15. First, the results match our expectations. As expected, opposite rules (MaxSR and MinSR, MaxVar and MinVar) have an excess overlap of -1 . Also, the sign of the excess overlap is coherent. For instance, the average excess overlap is positive between MaxVar and MinSR (0.16) because both rules tend to select assets with high marginal variance, while it is negative between MaxVar and MaxSR (-0.36) because they tend to select different assets. The excess overlap between the Rand rule and the other rules is also close to zero. Second, we observe that for essentially all pairs of selection rules we consider, there is either around the same or less overlap than in the random selection case. This answers our initial question, as it shows that the selection rules in Table 3 do select dissimilar assets.

OA.7.8 Effect of portfolio size on benchmark portfolio strategies

In this section, we report the empirical performance of the seven benchmark portfolio strategies considered in the bottom panel of Table 4 in the paper, i.e., EWRF, SGMV, GMVRF, SMV, SMV-ST, SMV-HT, and F+A, for different N . In particular, we consider $N = 20, 40, 60, 80$, and M . To select the N assets when $N < M$, we consider three asset selection rules: Rand, MaxSR, and MinVar. The results for the EWRF, SGMV, and GMVRF benchmark portfolios are available in Table OA.5, for SMV, SMV-ST, and SMV-HT in Table OA.6, and for F+A in Table OA.7.

We first discuss the effect of the portfolio size N on the EWRF, SGMV, and GMRF portfolios in Table OA.5. For the EWRF portfolio, which does not depend on the covariance matrix estimator, its performance slightly improves on average with N under the Rand asset selection rule, which is because EWRF is not much affected by estimation risk and thus can profit from more investment opportunities. However, when choosing the best assets first with the MinVar selection rule, the best performance is achieved with a small $N = 20$ or 40 among the values of N considered. This finding does not apply to the MaxSR selection rule, which does not generalize as well out of sample because it selects assets based on their noisy in-sample Sharpe ratio. Turning to the SGMV and GMVRF portfolios, we find that when they are implemented with the sample covariance matrix, reducing portfolio size is highly valuable and they overall reach their best performance for $N = 20$ or $N = 40$ under each of the Rand, MaxSR, and MinVar selection rules. However, when the SGMV and GMVRF portfolios are implemented with the shrinkage covariance matrix, they can profit from a larger N . For example, compared to choosing $N < M$ with the Rand or MaxSR selection rule, the SGMV portfolio performs best under $N = M$ in three out of seven datasets. However, when choosing the assets with minimum in-sample variances with the MinVar selection

Table OA.5: Effect of portfolio size on the annualized net out-of-sample utility of the EWRF, SGMV, and GMVRF benchmark portfolios (in percentage points)

		Sample covariance matrix							Linear shrinkage covariance matrix						
Asset select. rule	N	Dataset							Dataset						
		96	108	100	94	107	98IN-	100	96	108	100	94	107	98IN-	100
		S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	STO	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	STO
Panel (a): EWRF															
All	M	0.63	-1.55	0.59	-1.84	1.00	0.59	16.4	0.63	-1.55	0.59	-1.84	1.00	0.59	16.4
Rand	20	0.47	-1.63	0.32	-1.84	0.79	0.66	15.1	0.47	-1.63	0.32	-1.84	0.79	0.66	15.1
Rand	40	0.44	-1.66	0.35	-1.94	0.82	0.48	15.7	0.44	-1.66	0.35	-1.94	0.82	0.48	15.7
Rand	60	0.48	-1.59	0.44	-1.89	0.86	0.49	15.9	0.48	-1.59	0.44	-1.89	0.86	0.49	15.9
Rand	80	0.60	-1.56	0.50	-1.84	0.95	0.49	16.1	0.60	-1.56	0.50	-1.84	0.95	0.49	16.1
MaxSR	20	-2.36	-0.62	-5.59	-4.86	-6.40	-3.36	16.4	-2.36	-0.62	-5.59	-4.86	-6.40	-3.36	16.4
MaxSR	40	0.16	-1.33	-0.68	-0.97	-2.49	-0.35	20.2	0.16	-1.33	-0.68	-0.97	-2.49	-0.35	20.2
MaxSR	60	1.30	-1.90	0.98	-0.86	-0.41	0.54	22.2	1.30	-1.90	0.98	-0.86	-0.41	0.54	22.2
MaxSR	80	1.61	-1.83	1.61	-1.35	-0.15	0.81	21.8	1.61	-1.83	1.61	-1.35	-0.15	0.81	21.8
MinVar	20	4.16	-0.26	1.49	-0.10	4.97	4.88	28.9	4.16	-0.26	1.49	-0.10	4.97	4.88	28.9
MinVar	40	3.53	-0.59	2.71	0.19	4.10	4.36	31.1	3.53	-0.59	2.71	0.19	4.10	4.36	31.1
MinVar	60	3.46	-1.12	3.22	-0.81	4.18	2.98	26.4	3.46	-1.12	3.22	-0.81	4.18	2.98	26.4
MinVar	80	2.09	-1.25	2.59	-1.61	3.06	1.90	20.9	2.09	-1.25	2.59	-1.61	3.06	1.90	20.9
Panel (b): SGMV															
All	M	1.76	-49.7	4.25	-0.34	-2.28	-4.67	1.25	10.3	2.33	7.75	4.02	8.48	6.65	6.89
Rand	20	8.69	1.75	6.22	3.11	7.01	6.13	8.33	8.61	5.31	6.55	4.74	7.24	6.60	9.19
Rand	40	9.52	-1.20	6.18	1.48	7.88	5.60	7.10	9.41	4.28	6.67	4.38	7.89	6.79	8.27
Rand	60	9.20	-5.05	6.42	0.05	8.01	4.97	6.01	9.73	3.41	7.00	3.94	7.98	6.79	7.70
Rand	80	7.18	-11.5	6.55	-0.08	7.80	2.91	4.50	10.1	2.87	7.34	4.02	8.19	6.61	7.24
MaxSR	20	6.17	-2.16	7.82	5.86	3.99	5.13	10.2	7.20	1.23	8.21	6.12	4.71	5.15	10.5
MaxSR	40	5.39	-3.52	3.66	6.44	5.24	4.54	9.01	6.42	1.64	5.73	6.26	6.10	5.62	9.92
MaxSR	60	3.66	-6.33	2.50	4.48	4.79	5.10	7.59	5.48	2.38	5.32	5.85	6.61	5.57	9.54
MaxSR	80	3.22	-13.2	4.04	1.62	4.14	4.53	5.53	7.43	2.55	6.15	4.89	6.91	5.57	8.47
MinVar	20	7.49	-0.36	7.29	5.53	7.74	9.42	7.87	7.42	3.94	7.19	6.22	7.68	7.76	8.84
MinVar	40	8.49	-3.01	4.91	1.84	7.90	7.79	7.42	8.28	2.43	5.89	5.27	7.56	7.59	9.17
MinVar	60	6.82	-7.40	1.45	-1.81	6.31	3.25	6.43	7.26	1.64	4.43	2.58	6.86	5.94	8.71
MinVar	80	2.89	-12.2	2.29	0.71	3.43	1.75	5.10	6.72	2.25	5.86	2.52	6.14	5.99	8.17
Panel (c): GMVRF															
All	M	1.81	-26.3	-0.98	-11.4	-12.3	-5.34	-5.70	4.01	-0.00	1.47	0.14	0.66	0.49	2.10
Rand	20	14.8	1.73	6.57	0.50	9.04	7.92	16.0	13.8	6.43	8.13	0.53	10.5	7.80	20.7
Rand	40	20.0	-5.51	9.75	-1.69	13.0	9.01	8.50	17.3	2.29	10.5	-1.07	12.4	8.50	18.6
Rand	60	18.6	-11.0	10.2	-3.34	12.8	8.05	1.19	14.5	0.74	9.11	-0.60	9.67	6.80	14.0
Rand	80	11.6	-16.5	8.49	-7.44	9.47	2.63	-3.86	9.01	0.25	5.79	0.17	5.94	3.12	7.80
MaxSR	20	-0.14	-1.42	8.07	9.59	-4.53	0.16	28.2	4.01	2.59	9.47	10.1	-0.29	1.36	36.7
MaxSR	40	-2.18	-5.51	-0.93	3.36	0.80	2.21	21.5	1.48	0.73	3.20	3.93	4.29	3.99	33.5
MaxSR	60	-4.08	-12.1	0.79	-2.54	2.81	-1.65	12.0	1.58	0.29	3.54	2.20	4.66	1.74	22.4
MaxSR	80	-4.59	-17.9	1.72	-0.93	-0.17	-2.73	2.67	3.25	0.26	3.02	4.62	2.97	1.43	10.1
MinVar	20	9.02	-9.29	5.14	2.37	16.8	33.4	24.6	7.46	2.21	6.26	-1.36	16.1	21.3	34.1
MinVar	40	7.43	-4.46	1.44	-5.95	12.4	24.7	19.5	7.27	1.16	4.57	-8.24	9.07	13.7	32.0
MinVar	60	7.41	-11.5	-6.92	-8.73	4.05	0.77	7.89	6.91	0.20	1.98	-0.21	4.61	2.39	19.8
MinVar	80	3.41	-17.4	-5.07	-15.2	0.85	1.31	0.98	3.86	0.07	2.53	-2.06	2.06	2.06	9.58

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the EWRF, SGMV, and GMVRF benchmark strategies, which are described in Table 2 of the paper. We consider as asset selection rules All, Rand, MaxSR, and MinVar, described in Table 3 of the paper, and portfolio sizes ranging from $N = 20$ to $N = M$, the latter being the All selection rule. We report results across the seven datasets in Table 1 and estimate the portfolios either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). The table is constructed following the methodology described in Section 5.4. The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

Table OA.6: Effect of portfolio size on the annualized net out-of-sample utility of the SMV, SMV-ST, and SMV-HT benchmark portfolios (in percentage points)

		Sample covariance matrix							Linear shrinkage covariance matrix						
Asset select. rule	N	Dataset							Dataset						
		96	108	100	94	107	98IN-	100	96	108	100	94	107	98IN-	100
		S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	STO	S-BM	CHA	S-OP	IN-NV	IN-CHA	CHA-NV	STO
Panel (a): SMV															
All	M	-4E+05	-3E+11	-1E+06	-6E+07	-8E+08	-8E+07	-9E+04	-1443	-107	-1399	-2360	-1457	-1143	-1844
Rand	20	-240	-719	-253	-575	-272	-325	-234	-86.8	82.8	-111	-38.2	-100	-23.7	-121
Rand	40	-1131	-4753	-1043	-3709	-1233	-2442	-993	-266	84.3	-295	-343	-270	-128	-366
Rand	60	-3980	-6E+04	-3419	-1E+04	-4213	-1E+04	-3304	-565	62.6	-594	-916	-523	-355	-722
Rand	80	-1E+04	-3E+06	-9166	-3E+05	-2E+04	-3E+05	-1E+04	-990	11.9	-948	-1709	-859	-714	-1207
MaxSR	20	-95.7	-286	-98.1	-256	-230	-217	-183	-33.9	40.2	-45.9	-87.2	-81.6	-95.5	-70.4
MaxSR	40	-573	-2158	-458	-1616	-822	-884	-733	-170	99.3	-157	-198	-166	-165	-178
MaxSR	60	-2006	-4E+04	-1862	-6927	-2678	-4486	-2389	-357	136	-319	-433	-342	-396	-344
MaxSR	80	-7543	-2E+06	-6950	-5E+05	-9556	-7E+04	-8904	-759	95.8	-610	-813	-622	-657	-656
MinVar	20	-216	-405	-225	-882	-269	-707	-222	-83.3	54.0	-105	-2.57	-82.7	105	-44.0
MinVar	40	-916	-2132	-733	-4335	-986	-2750	-955	-232	121	-234	-276	-162	41.9	-206
MinVar	60	-3205	-7E+04	-2440	-3E+04	-3086	-1E+04	-3200	-492	138	-444	-781	-387	-82.5	-489
MinVar	80	-7688	-2E+06	-6551	-7E+05	-1E+04	-3E+05	-1E+04	-833	117	-660	-1484	-660	-471	-952
Panel (b): SMV-ST															
All	M	15.8	-108	11.7	-579	-2.91	0.88	0.77	23.9	67.8	5.13	-195	10.4	16.4	1.43
Rand	20	-7.40	-39.5	-3.95	-87.9	-8.19	-31.5	10.7	-5.02	8.34	-2.74	-52.8	-4.74	-12.4	10.4
Rand	40	-7.19	-29.9	-9.75	-124	-16.0	-52.4	10.0	-0.64	30.4	-6.19	-72.2	-8.79	-19.2	10.0
Rand	60	-1.96	1.31	-7.17	-148	-21.1	-33.2	9.45	6.27	64.6	-3.32	-61.4	-9.08	10.1	9.43
Rand	80	5.28	-5.48	-0.41	-228	-13.4	7.76	6.81	13.6	88.0	3.26	-103	-3.43	52.6	9.24
MaxSR	20	9.61	-7.96	7.30	5.49	-1.13	-2.83	13.9	9.97	5.17	6.45	5.04	-0.40	-3.54	-129
MaxSR	40	8.72	-15.5	4.81	-32.1	5.01	-17.0	12.6	11.7	15.8	2.49	-12.7	9.84	4.35	-155
MaxSR	60	3.14	74.3	-11.1	-71.9	-5.22	-7.79	10.9	12.3	60.0	0.57	-79.3	7.65	0.18	-184
MaxSR	80	-3.62	33.1	1.86	-94.4	-12.1	-31.1	5.93	5.82	118	-0.25	-89.5	-3.89	-45.5	-210
MinVar	20	0.56	8.66	5.95	9.03	-8.59	92.3	13.3	2.36	22.3	5.50	34.1	-3.14	103	-55.4
MinVar	40	-4.36	39.5	2.52	-120	-16.6	43.7	14.3	-1.21	50.9	4.42	-0.08	-3.53	68.5	-52.0
MinVar	60	-19.2	26.7	-4.91	-159	-16.1	81.3	10.8	-5.85	95.6	-0.33	52.4	-12.1	114	-73.9
MinVar	80	-18.9	30.0	-4.49	-345	17.4	82.5	6.46	-14.2	112	-5.07	-55.2	8.89	148	-88.8
Panel (c): SMV-HT															
All	M	-35.3	-302	-16.8	-105	-25.9	-143	-7.62	-29.9	-52.1	-17.8	-376	-33.6	-349	-21.1
Rand	20	-41.0	-439	-29.3	-168	-45.1	-106	-9.61	-24.0	-20.2	-17.7	-60.1	-26.9	-35.2	-5.80
Rand	40	-47.8	-630	-41.0	-167	-49.6	-131	-10.8	-26.2	-23.4	-22.7	-72.8	-33.8	-70.8	-10.5
Rand	60	-48.8	-680	-35.2	-130	-50.4	-132	-11.2	-27.8	-24.3	-19.5	-123	-36.7	-116	-14.1
Rand	80	-41.6	-574	-31.3	-135	-55.2	-163	-10.1	-28.2	-21.2	-19.5	-193	-38.5	-177	-17.8
MaxSR	20	-16.1	-45.3	-15.9	-17.3	-15.3	-15.0	-16.8	-14.7	-5.57	-21.0	-26.7	-19.9	-6.57	-15.8
MaxSR	40	-17.1	-128	-11.2	-30.1	-10.9	-7.93	-14.8	-26.1	-13.7	-13.7	-58.1	-28.5	-7.68	-15.9
MaxSR	60	-2.09	-110	-13.7	-51.3	-28.6	-32.9	-12.7	-14.0	2.54	-9.84	-159	-19.1	-42.1	-15.4
MaxSR	80	-19.5	-161	-2.96	-194	-14.1	-35.7	-10.9	-19.1	-10.6	-5.14	-300	-16.0	-86.9	-22.9
MinVar	20	-33.6	-138	-11.4	-125	-20.8	-113	-19.0	-19.4	-9.65	-15.9	-8.69	-4.99	32.3	-16.2
MinVar	40	-36.1	-212	-15.9	-146	-14.8	-35.7	-14.4	-22.8	1.76	-22.6	-79.7	-28.9	-28.7	-18.9
MinVar	60	-21.8	-575	-10.4	-79.0	-15.6	-55.4	-14.1	-9.68	-8.84	-10.7	-100	-46.8	5.40	-15.9
MinVar	80	-14.7	-272	-7.82	-49.4	-20.0	-100	-10.3	-16.3	-5.00	-16.7	-248	-36.9	-156	-18.0

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the SMV, SMV-ST, and SMV-HT benchmark strategies, which are described in Table 2 of the paper. We consider as asset selection rules All, Rand, MaxSR, and MinVar, described in Table 3 of the paper, and portfolio sizes ranging from $N = 20$ to $N = M$, the latter being the All selection rule. We report results across the seven datasets in Table 1 and estimate the portfolios either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). The table is constructed following the methodology described in Section 5.4. The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

Table OA.7: Effect of portfolio size on the annualized net out-of-sample utility of the F+A benchmark portfolio (in percentage points)

Asset select. rule	N	Sample covariance matrix							Linear shrinkage covariance matrix						
		Dataset							Dataset						
		96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN- CHA-NV	100 STO
All	M	-232	-158	-368	-338	-138	-138	-110	-113	19.7	-227	-214	-113	-37.5	-111
Rand	20	-143	-283	-162	-170	-135	-123	-50.7	-97.4	-75.2	-109	-102	-101	-70.6	-41.6
Rand	40	-192	-368	-242	-203	-171	-146	-71.5	-134	-88.8	-167	-136	-133	-82.8	-67.6
Rand	60	-212	-337	-293	-212	-187	-139	-83.8	-142	-62.7	-200	-151	-146	-73.3	-83.6
Rand	80	-222	-303	-328	-228	-184	-117	-95.9	-138	-29.1	-216	-161	-141	-47.0	-97.3
MaxSR	20	-49.8	-62.3	-46.9	-107	-87.2	-76.1	-35.2	-36.7	-12.0	-25.4	-66.9	-59.7	-64.1	-38.8
MaxSR	40	-129	-69.4	-127	-155	-118	-112	-45.2	-74.7	-19.2	-68.2	-103	-71.4	-64.5	-44.7
MaxSR	60	-164	-57.0	-154	-235	-123	-169	-35.9	-98.0	-10.9	-92.4	-168	-81.4	-80.6	-35.8
MaxSR	80	-228	-84.6	-224	-175	-159	-233	-56.0	-161	-15.1	-137	-62.6	-101	-121	-52.7
MinVar	20	-120	-98.0	-119	-121	-111	-27.2	-73.7	-80.1	8.60	-83.0	-44.6	-71.5	92.4	-58.5
MinVar	40	-199	-99.1	-173	-108	-161	-13.4	-143	-158	5.02	-127	-12.1	-129	69.6	-127
MinVar	60	-263	-164	-250	-312	-141	-104	-196	-172	-9.45	-176	-159	-104	50.7	-182
MinVar	80	-214	-120	-310	-436	-181	-201	-210	-139	9.13	-210	-270	-126	-9.56	-200

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the F+A benchmark strategy, which is described in Table 2 of the paper. We consider as asset selection rules All, Rand, MaxSR, and MinVar, described in Table 3 of the paper, and portfolio sizes ranging from $N = 20$ to $N = M$, the latter being the All selection rule. We report results across the seven datasets in Table 1 and estimate the portfolio either with the sample covariance matrix or the linear shrinkage estimator of Ledoit and Wolf (2004). The table is constructed following the methodology described in Section 5.4. The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

rule, both the SGMV and GMVRF portfolios perform overall best for $N = 20$.

Next, we discuss the effect of the portfolio size N on the SMV, SMV-ST, SMV-HT, and F+A portfolios in Tables OA.6 and OA.7. For the SMV portfolio, increasing N has a large negative impact on the EU. This result is consistent with our theory that sets a very small optimal N for SMV; see Figure 2 in the main text. The insight is similar for the F+A strategy, consistent with the left plot of Figure 7 in our simulations. SMV-ST and SMV-HT alleviate this negative impact of N by setting a limited number of non-zero weights using an L_1 -norm constraint and a hard threshold on the weights, respectively. Looking at SMV-HT, although it improves upon SMV, increasing N still has a negative impact on its EU in general. We can explain this finding because the assets in which SMV-HT invests are determined from the weights of SMV, which face high estimation risk. However, looking at SMV-ST, increasing N is no longer systematically unfavorable. In particular, under the shrinkage covariance matrix and the Rand asset selection rule, the best N among the

values considered is often $N = 80$ or M . For the MinVar and MaxSR asset selection rules, where we pick the best assets first, the best N under the shrinkage covariance matrix is often either quite small (20 or 40) or it is $N = M$. Setting N small is favorable because we select the best assets and we have less estimation risk. Setting $N = M$ is favorable too for SMV-ST because we still end up with a limited number of non-zero weights due to the L_1 -norm constraint, so that estimation risk is not excessive even for $N = M$. Moreover, by setting $N = M$, we do not change the investment universe over time, so that the asset selection done by the L_1 norm is more stable and SMV-ST incurs less transaction costs.

OA.8 Proofs of theoretical results in Appendix

In this section, we provide the proofs for all theoretical results given in the Appendix and the Online Appendix.

OA.8.1 Proof of Proposition A.1

Let $\tilde{\mu}_1 = \mathbb{E}[\hat{\boldsymbol{\mu}}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\mu}_N]$, $\tilde{\mu}_2 = \mathbb{E}[\boldsymbol{\mu}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}_N]$, $\tilde{\sigma}_1^2 = \mathbb{E}[\hat{\boldsymbol{\mu}}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\Sigma}_N \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N]$, $\tilde{\sigma}_2^2 = \mathbb{E}[\mathbf{1}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\Sigma}_N \hat{\boldsymbol{\Sigma}}_N^{-1} \mathbf{1}_N]$, and $\tilde{\sigma}_{12} = \mathbb{E}[\hat{\boldsymbol{\mu}}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \boldsymbol{\Sigma}_N \hat{\boldsymbol{\Sigma}}_N^{-1} \mathbf{1}_N]$. Then, the EU of the GMV-three-fund rule $\hat{\mathbf{w}}(\boldsymbol{\alpha})$ in (A3) is

$$EU(\hat{\mathbf{w}}(\boldsymbol{\alpha})) = \frac{1}{2\gamma} (2\alpha_1 \tilde{\mu}_1 + 2\alpha_2 \tilde{\mu}_2 - \alpha_1^2 \tilde{\sigma}_1^2 - \alpha_2^2 \tilde{\sigma}_2^2 - 2\alpha_1 \alpha_2 \tilde{\sigma}_{12}). \quad (\text{OA26})$$

Kan and Lassance (2025, Equations (IA80)–(IA84)) show that

$$\tilde{\mu}_1 = \frac{\kappa_{N,1} \theta_N^2 T}{T - N - 2}, \quad (\text{OA27})$$

$$\tilde{\mu}_2 = \frac{\kappa_{N,1} \lambda_{g,N} T}{T - N - 2}, \quad (\text{OA28})$$

$$\tilde{\sigma}_1^2 = \frac{c_N T^2}{(T - N - 2)^2} \left(\kappa_{N,2} \theta_N^2 + \kappa_{N,3} \frac{N}{T} \right), \quad (\text{OA29})$$

$$\tilde{\sigma}_2^2 = \frac{c_N \kappa_{N,2} T^2}{\sigma_{g,N}^2 (T - N - 2)^2}, \quad (\text{OA30})$$

$$\tilde{\sigma}_{12} = \frac{c_N \kappa_{N,2} \lambda_{g,N} T^2}{(T - N - 2)^2}, \quad (\text{OA31})$$

and plugging them into (OA26) yields the desired result in Equation (A4). It is then easy to show that the $\alpha = (\alpha_1, \alpha_2)$ maximizing (A4) is equal to (A5), which also corresponds to Kan and Lassance (2025, Equations (51)–(52)). Finally, after some developments, plugging (A5) into (A4) yields Equation (A6), which concludes the proof.

OA.8.2 Proof of Proposition A.2

The squared Sharpe ratio of the GMV portfolio depends on μ_N and Σ_N as

$$\theta_{g,N}^2 = \frac{(\mathbf{1}_N' \Sigma_N^{-1} \mu_N)^2}{\mathbf{1}_N' \Sigma_N^{-1} \mathbf{1}_N}. \quad (\text{OA32})$$

Now, under Assumption 1, Σ_N^{-1} is given by (A38), and thus,

$$\mathbf{1}_N' \Sigma_N^{-1} \mu_N = \frac{N}{1 - \rho} \left(\bar{\lambda}_N - \frac{N\rho}{1 - \rho + N\rho} \bar{\theta}_{N,1} \bar{\sigma}_{N,-1} \right), \quad (\text{OA33})$$

$$\mathbf{1}_N' \Sigma_N^{-1} \mathbf{1}_N = \frac{N}{1 - \rho} \left(\bar{\sigma}_{N,-2} - \frac{N\rho}{1 - \rho + N\rho} \bar{\sigma}_{N,-1}^2 \right). \quad (\text{OA34})$$

Plugging (OA33)–(OA34) into (OA32) yields the desired result in Equation (A7), which concludes the proof.

OA.8.3 Proof of Proposition A.3

The EU of the EW-three-fund rule $\hat{w}(\beta)$ in (A12) is

$$EU(\hat{w}(\beta)) = \frac{1}{2\gamma} \left(2\beta_1 \tilde{\mu}_1 + 2\beta_2 \mu_{ew,N} - \beta_1^2 \tilde{\sigma}_1^2 - \beta_2^2 \sigma_{ew,N}^2 - 2\mathbb{E}[\hat{\mu}' \hat{\Sigma}^{-1} \Sigma w_{ew}] \right), \quad (\text{OA35})$$

where $\tilde{\mu}_1$ is given by (OA27), $\tilde{\sigma}_1^2$ by (OA29), and $\mathbb{E}[\hat{\boldsymbol{\mu}}' \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma} \mathbf{w}_{ew}] = \frac{\kappa_{N,1} \mu_{ew,N} T}{T-N-2}$, which yields the desired result in Equation (A13). It is then easy to show that the $\boldsymbol{\beta} = (\beta_1, \beta_2)$ maximizing (A13) is equal to (A14). Finally, after some developments, plugging (A14) into (A13) yields Equation (A15), which concludes the proof.

OA.8.4 Proof of Proposition A.4

We have that $\theta_{ew,N}^2 = \mu_{ew,N}^2 / \sigma_{ew,N}^2$, where $\mu_{ew,N} = \bar{\mu}_N$ and, under Assumption 1,

$$\sigma_{ew,N}^2 = \frac{1}{N^2} \mathbf{1}_N' \boldsymbol{\Sigma} \mathbf{1}_N = \frac{1}{N} \left(\bar{\sigma}_{N,2} + \frac{\rho}{N} \sum_{i=1}^N \sum_{j \neq i}^N \sigma_i \sigma_j \right). \quad (\text{OA36})$$

This yields the desired result in Equation (A16) after noticing that

$$\frac{\rho}{N} \sum_{i=1}^N \sum_{j \neq i}^N \sigma_i \sigma_j = \rho N \bar{\sigma}_{N,1}^2 - \rho \bar{\sigma}_{N,2}, \quad (\text{OA37})$$

which concludes the proof.

OA.8.5 Proof of Proposition A.5

Throughout the proof, we denote

$$f(N, \rho) := \frac{N\rho}{1 - \rho + N\rho}. \quad (\text{OA38})$$

Bias of $\hat{\delta}_N$. The expectation of $\hat{\delta}_N$ is

$$\mathbb{E}[\hat{\delta}_N] = \frac{1}{M} \sum_{i=1}^M \mathbb{E}[\hat{s}_i^2] - \frac{f(N, \rho)}{M^2} \mathbb{E} \left[\left(\sum_{i=1}^M \hat{s}_i \right)^2 \right]. \quad (\text{OA39})$$

Given that $\mathbb{E}[\hat{s}_i^2] = s_i^2 + \frac{1}{T}$, we have

$$\frac{1}{M} \sum_{i=1}^M \mathbb{E}[\hat{s}_i^2] = \bar{\theta}_{M,2} + \frac{1}{T}. \quad (\text{OA40})$$

Moreover,

$$\frac{1}{M^2} \mathbb{E} \left[\left(\sum_{i=1}^M \hat{s}_i \right)^2 \right] = \bar{\theta}_{M,1}^2 + \frac{1}{M^2} \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{s}_i \right], \quad (\text{OA41})$$

where

$$\frac{1}{M^2} \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{s}_i \right] = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \mathbb{C}\text{ov}[\hat{s}_i, \hat{s}_j] = \frac{1}{T} \left(\frac{1}{M} + \frac{M-1}{M} \rho \right). \quad (\text{OA42})$$

Plugging (OA40)–(OA42) into (OA39) yields

$$\mathbb{E}[\hat{\delta}_N] = \delta_N + \frac{1}{T} \left(1 - \frac{f(N, \rho)(1 - \rho + M\rho)}{M} \right), \quad (\text{OA43})$$

which corresponds to (A33), as desired.

Bias of $\hat{r}_{g,N}$. The expectation of $\hat{r}_{g,N}$ is

$$\mathbb{E}[\hat{r}_{g,N}] = \frac{\mathbb{E} \left[\left(\frac{1}{M} \sum_{i=1}^M \hat{\lambda}_i - \frac{f(N, \rho)}{M} \bar{\sigma}_{M,-1} \sum_{i=1}^M \hat{s}_i \right)^2 \right]}{\bar{\sigma}_{M,-2} - f(N, \rho) \bar{\sigma}_{M,-1}^2}. \quad (\text{OA44})$$

Given that $\mathbb{E}[\hat{\lambda}_i] = \lambda_i$ and $\mathbb{E}[\hat{s}_i] = s_i$, we have

$$\begin{aligned} \mathbb{E}[\hat{r}_{g,N}] &= r_{g,N} + \frac{1}{M^2} \times \\ &\frac{\mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{\lambda}_i \right] + f(N, \rho)^2 \bar{\sigma}_{M,-1}^2 \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{s}_i \right] - 2f(N, \rho) \bar{\sigma}_{M,-1} \mathbb{C}\text{ov} \left[\sum_{i=1}^M \hat{\lambda}_i, \sum_{i=1}^M \hat{s}_i \right]}{\bar{\sigma}_{M,-2} - f(N, \rho) \bar{\sigma}_{M,-1}^2}, \end{aligned} \quad (\text{OA45})$$

where $\frac{1}{M^2} \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{s}_i \right]$ is given by (OA42) and

$$\frac{1}{M^2} \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{\lambda}_i \right] = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \mathbb{C}\text{ov}[\hat{\lambda}_i, \hat{\lambda}_j] = \frac{1}{T} \left(\frac{1-\rho}{M} \bar{\sigma}_{M,-2} + \rho \bar{\sigma}_{M,-1}^2 \right), \quad (\text{OA46})$$

$$\frac{1}{M^2} \mathbb{C}\text{ov} \left[\sum_{i=1}^M \hat{\lambda}_i, \sum_{i=1}^M \hat{s}_i \right] = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \mathbb{C}\text{ov}[\hat{\lambda}_i, \hat{s}_j] = \frac{1-\rho+M\rho}{MT} \bar{\sigma}_{M,-1}. \quad (\text{OA47})$$

Plugging (OA42) and (OA46)–(OA47) into (OA45) yields

$$\mathbb{E}[\hat{r}_{g,N}] = r_{g,N} + \frac{1}{T} \frac{\frac{1-\rho}{M} \bar{\sigma}_{M,-2} + \left(\rho + \frac{f(N,\rho)^2(1-\rho+M\rho)}{M} - \frac{2f(N,\rho)(1-\rho+M\rho)}{M} \right) \bar{\sigma}_{M,-1}^2}{\bar{\sigma}_{M,-2} - f(N,\rho) \bar{\sigma}_{M,-1}^2}, \quad (\text{OA48})$$

which corresponds to (A34), as desired.

Bias of $\hat{r}_{ew,N}$. The expectation of $\hat{r}_{ew,N}$ is

$$\mathbb{E}[\hat{r}_{ew,N}] = \frac{(1-\rho)}{(1-\rho)\bar{\sigma}_{M,2} + \rho N \bar{\sigma}_{M,1}^2} \frac{1}{M^2} \mathbb{E} \left[\left(\sum_{i=1}^M \hat{\mu}_i \right)^2 \right]. \quad (\text{OA49})$$

Given that $\mathbb{E}[\hat{\mu}_i] = \mu_i$, we have

$$\frac{1}{M^2} \mathbb{E} \left[\left(\sum_{i=1}^M \hat{\mu}_i \right)^2 \right] = \bar{\mu}_M^2 + \frac{1}{M^2} \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{\mu}_i \right]. \quad (\text{OA50})$$

Moreover,

$$\frac{1}{M^2} \mathbb{V}\text{ar} \left[\sum_{i=1}^M \hat{\mu}_i \right] = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \mathbb{C}\text{ov}[\hat{\mu}_i, \hat{\mu}_j] = \frac{(1-\rho)\bar{\sigma}_{M,2} + \rho M \bar{\sigma}_{M,1}^2}{MT}. \quad (\text{OA51})$$

Plugging (OA50)–(OA51) into (OA49) yields

$$\mathbb{E}[\hat{r}_{ew,N}] = r_{ew,N} + \frac{1-\rho}{MT} \times \frac{(1-\rho)\bar{\sigma}_{M,2} + \rho M \bar{\sigma}_{M,1}^2}{(1-\rho)\bar{\sigma}_{M,2} + \rho N \bar{\sigma}_{M,1}^2}, \quad (\text{OA52})$$

which corresponds to (A35) and concludes the proof.

OA.8.6 Proof of Proposition OA.1

Under Assumption OA.1, the covariance matrix of $\mathbf{r}$ is

$$\Sigma_N = \sigma_f^2 \boldsymbol{\beta}_N \boldsymbol{\beta}_N' + \sigma_\varepsilon^2 \mathbf{I}_N, \quad (\text{OA53})$$

and its inverse is

$$\Sigma_N^{-1} = \frac{1}{\sigma_\varepsilon^2} \left[\mathbf{I}_N - \frac{\sigma_f^2}{\sigma_\varepsilon^2 + \sigma_f^2 \boldsymbol{\beta}_N' \boldsymbol{\beta}_N} \boldsymbol{\beta}_N \boldsymbol{\beta}_N' \right]. \quad (\text{OA54})$$

We have that θ_N^2 becomes

$$\theta_N^2 = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \frac{N}{\sigma_\varepsilon^2} \left[\bar{\ell}_N^\mu - \frac{N\sigma_f^2}{\sigma_\varepsilon^2 + N\sigma_f^2 \bar{\ell}_N^\beta} (\bar{\ell}_N^{\mu,\beta})^2 \right], \quad (\text{OA55})$$

which corresponds to (OA3). Plugging (OA55) in (9) and (13) yields the desired result and thus completes the proof.

OA.8.7 Proof of Proposition OA.2

We begin by proving that $U(\mathbf{w}^*)$ in Proposition 1 is a convex function by showing that its second derivative with respect to ρ is positive for all $\rho \in (-\frac{1}{N-1}, 1)$. The first derivative is

$$\frac{\partial}{\partial \rho} U(\mathbf{w}^*) = \frac{N}{2\gamma} \left[\frac{\bar{\theta}_{N,2}(1 - \rho + N\rho)^2 - N\bar{\theta}_{N,1}^2(1 - \rho^2 + N\rho^2)}{(1 - \rho)^2(1 - \rho + N\rho)^2} \right], \quad (\text{OA56})$$

and the second derivative is

$$\frac{\partial^2}{\partial \rho^2} U(\mathbf{w}^*) = \frac{N}{\gamma(1 - \rho)^3} \left[(\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2) + \frac{(N - 1)^2(1 - \rho)^3}{(1 - \rho + N\rho)^3} \bar{\theta}_{N,1}^2 \right], \quad (\text{OA57})$$

which is positive because $\bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2 \geq 0$ and $(1 - \rho + N\rho)^3 \geq 0$ for all $\rho \in (-\frac{1}{N-1}, 1)$.

Next, we prove that the value of $\rho \in (-\frac{1}{N-1}, 1)$ minimizing $U(\mathbf{w}^*)$ is given by $\rho^{\min}$ in (OA6).

There are two roots to the first derivative in (OA56) given by

$$\rho^\pm = \frac{-\bar{\theta}_{N,2} \pm \frac{N}{\sqrt{N-1}} |\bar{\theta}_{N,1}| \sqrt{\delta}}{N\delta - \bar{\theta}_{N,2}}, \quad (\text{OA58})$$

where we define $\delta = \bar{\theta}_{N,2} - \bar{\theta}_{N,1}^2 \geq 0$. We proceed by showing that $\rho^- \notin (-\frac{1}{N-1}, 1)$ whereas $\rho^+ = \rho^{\min} \in (-\frac{1}{N-1}, 1)$, and thus, is a minimizer of $U(\mathbf{w}^*)$ that is convex in that interval.

We first treat the case $\rho^- \notin (-\frac{1}{N-1}, 1)$. We have $\lim_{N \rightarrow \bar{\theta}_{N,2}/\delta} \rho^- = \infty$ and thus we can leave out the case $N = \bar{\theta}_{N,2}/\delta$. Let us consider the case $N > \bar{\theta}_{N,2}/\delta$. Then, ρ^- is negative and we need to show that $\rho^- < -\frac{1}{N-1}$. This is equivalent to

$$\frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N\sqrt{\delta}}{N-1} \left(\sqrt{N-1} |\bar{\theta}_{N,1}| - \sqrt{\delta} \right) > 0, \quad (\text{OA59})$$

which holds because, given $N > \bar{\theta}_{N,2}/\delta$, we have

$$\begin{aligned}
& \frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N\sqrt{\delta}}{N-1} \left(\sqrt{N-1}|\bar{\theta}_{N,1}| - \sqrt{\delta} \right) \\
& > \frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N\sqrt{\delta}}{N-1} \left(\sqrt{\frac{\bar{\theta}_{N,2}}{\delta}} - 1|\bar{\theta}_{N,1}| - \sqrt{\delta} \right) \\
& = \frac{N\bar{\theta}_{N,2}}{N-1} + \frac{N(2\bar{\theta}_{N,1}^2 - \bar{\theta}_{N,2})}{N-1} > 0.
\end{aligned} \tag{OA60}$$

Then, we consider the case $N < \bar{\theta}_{N,2}/\delta$. In this case, ρ^- is positive and we need to show that $\rho^- > 1$. This is equivalent to

$$N\delta + \frac{N|\bar{\theta}_{N,1}|\sqrt{\delta}}{\sqrt{N-1}} > 0, \tag{OA61}$$

which holds true.

We treat next the case $\rho^+ \in (-\frac{1}{N-1}, 1)$. When $N = \bar{\theta}_{N,2}/\delta$, $\rho^+ = 0/0$ is indeterminate and we can show using L'Hospital rule that

$$\lim_{N \rightarrow \bar{\theta}_{N,2}/\delta} = 1 - \frac{\bar{\theta}_{N,2}}{2\bar{\theta}_{N,1}^2}, \tag{OA62}$$

which is strictly smaller than one and, moreover, $1 - \bar{\theta}_{N,2}/(2\bar{\theta}_{N,1}^2) > 1 - \bar{\theta}_{N,2}/\bar{\theta}_{N,1}^2 = -1/(N-1)$

when $N = \bar{\theta}_{N,2}/\delta$. We then consider the case $N > \bar{\theta}_{N,2}/\delta$. In this case, $\rho^+ < 1$ is equivalent to

$$\frac{N\sqrt{\delta}}{\sqrt{N-1}} \left(\sqrt{(N-1)\delta} - |\bar{\theta}_{N,1}| \right) > 0, \tag{OA63}$$

which holds if $\delta > \bar{\theta}_{N,1}^2/(N-1)$ and this holds true because, given $N > \bar{\theta}_{N,2}/\delta$,

$$\frac{\bar{\theta}_{N,1}^2}{N-1} < \frac{\bar{\theta}_{N,1}^2}{\bar{\theta}_{N,2}/\delta - 1} = \delta. \tag{OA64}$$

Next, still in the case $N > \bar{\theta}_{N,2}/\delta$, $\rho^+ > -\frac{1}{N-1}$ is equivalent to

$$\frac{N}{N-1} \left(\delta + |\bar{\theta}_{N,1}|\sqrt{(N-1)\delta} - \bar{\theta}_{N,2} \right) > 0, \tag{OA65}$$

which holds true because

$$\delta + |\bar{\theta}_{N-1}| \sqrt{(N-1)\delta} - \bar{\theta}_{N,2} > \delta + |\bar{\theta}_{N-1}| \sqrt{\left(\frac{\bar{\theta}_{N,2}}{\delta} - 1\right)\delta} - \bar{\theta}_{N,2} = 0. \quad (\text{OA66})$$

We turn then to the case where $N < \bar{\theta}_{N,2}/\delta$. In this case, $\rho^+ < 1$ is equivalent to

$$\frac{N\sqrt{\delta}}{\sqrt{N-1}} \left(|\bar{\theta}_{N,1}| - \sqrt{(N-1)\delta} \right) > 0, \quad (\text{OA67})$$

which holds if $\delta < \bar{\theta}_{N,1}^2/(N-1)$ and this holds true because, given $N < \bar{\theta}_{N,2}/\delta$, inequality (OA64)

is reversed. Finally, still in the case $N < \bar{\theta}_{N,2}/\delta$, $\rho^+ > -\frac{1}{N-1}$ is equivalent to

$$\frac{N}{N-1} \left(\delta + |\bar{\theta}_{N-1}| \sqrt{(N-1)\delta} - \bar{\theta}_{N,2} \right) < 0, \quad (\text{OA68})$$

which holds true because, given $N < \bar{\theta}_{N,2}/\delta$, (OA66) is reversed. This concludes the proof.

OA.8.8 Proof of Proposition OA.3

The EU of the portfolio $\eta \hat{\mathbf{w}}_{ewrf}$ is

$$EU(\eta \hat{\mathbf{w}}_{ewrf}) = \eta \mathbb{E}[\hat{\mathbf{w}}'_{ewrf} \boldsymbol{\mu}] - \frac{\gamma}{2} \eta^2 \mathbb{E}[\hat{\mathbf{w}}'_{ewrf} \boldsymbol{\Sigma} \hat{\mathbf{w}}_{ewrf}]. \quad (\text{OA69})$$

Therefore, the optimal η is $\eta^* = (1/\gamma) \mathbb{E}[\hat{\mathbf{w}}'_{ewrf} \boldsymbol{\mu}] / \mathbb{E}[\hat{\mathbf{w}}'_{ewrf} \boldsymbol{\Sigma} \hat{\mathbf{w}}_{ewrf}]$. Now, using the fact that,

under i.i.d. multivariate normality, $\hat{\boldsymbol{\mu}}_{ew,N} \sim \mathcal{N}(\boldsymbol{\mu}_{ew,N}, \sigma_{ew,N}^2/T)$, $T\hat{\sigma}_{ew,N}^2 \sim \sigma_{ew,N}^2 \chi_{T-1}^2$, and they

are independent of each other, the first two moments of $\hat{\lambda}_{ew,N} = \frac{T-3}{T} \hat{\boldsymbol{\mu}}_{ew,N} / \hat{\sigma}_{ew,N}^2$ are

$$\mathbb{E}[\hat{\lambda}_{ew,N}] = \lambda_{ew,N} \quad \text{and} \quad \mathbb{E}[\hat{\lambda}_{ew,N}^2] = \frac{T-3}{T-5} \left(\lambda_{ew,N}^2 + \frac{1}{T\sigma_{ew,N}^2} \right), \quad (\text{OA70})$$

and thus,

$$\mathbb{E}[\hat{\mathbf{w}}'_{ewrf} \boldsymbol{\mu}] = \frac{\theta_{ew,N}^2}{\gamma} \quad \text{and} \quad \mathbb{E}[\hat{\mathbf{w}}'_{ewrf} \boldsymbol{\Sigma} \hat{\mathbf{w}}_{ewrf}] = \frac{T-3}{\gamma^2(T-5)} \left(\theta_{ew,N}^2 + \frac{1}{T} \right). \quad (\text{OA71})$$

Therefore, the optimal η^* is equal to (OA13), which concludes the proof.

OA.8.9 Proof of Proposition OA.4

Consider a pair of selection rules (k, l) that both select N_t assets out of M available ones. They randomly select assets from a Bernoulli distribution without replacement and are mutually independent. The selections are stored in the $(M \times 1)$ selection vectors $\mathbf{y}_{k,t}$ and $\mathbf{y}_{l,t}$.

The number of assets selected in rule $\mathbf{y}_{k,t}$ is $\sum_{i=1}^M y_{i,k,t} = \mathbf{y}'_{k,t} \mathbf{1}_M$ and the overlap between rules k and l is $\mathbf{y}'_{k,t} \mathbf{y}_{l,t} = \sum_{i=1}^M y_{i,k,t} y_{i,l,t}$. We are interested in the expected overlap between these two rules, denoted $\bar{N}_{t,M}$. Using the independence properties of the two rules and that $y_{i,k,t} \in \{0, 1\} \forall i, k, t$, we can write $\bar{N}_{t,M}$ as

$$\begin{aligned} \bar{N}_{t,M} &= \mathbb{E} \left[\sum_{i=1}^M y_{i,k,t} y_{i,l,t} \middle| \sum_{i=1}^M y_{i,k,t} = \sum_{i=1}^M y_{i,l,t} = N_t \right] \\ &= \sum_{i=1}^M \mathbb{P} [y_{i,k,t} = 1 \mid \mathbf{y}'_{k,t} \mathbf{1}_M = N_t] \times \mathbb{P} [y_{i,l,t} = 1 \mid \mathbf{y}'_{l,t} \mathbf{1}_M = N_t]. \end{aligned} \quad (\text{OA72})$$

Consider the probability of getting $\mathbf{y}_{k,t} = \tilde{\mathbf{y}}$ where $\tilde{\mathbf{y}}$ is a binary vector such that $\sum_{i=1}^M \tilde{y}_i = n$. Given that the sum of i.i.d. Bernoulli random variables follows a binomial distribution, we can write, using Bayes theorem,

$$\begin{aligned} \mathbb{P} [\mathbf{y}_{k,t} = \tilde{\mathbf{y}} \mid \mathbf{y}'_{k,t} \mathbf{1}_M = N_t] &= \frac{\mathbb{P} [\mathbf{y}'_{k,t} \mathbf{1}_M = N_t \mid \mathbf{y}_{k,t} = \tilde{\mathbf{y}}] \times \mathbb{P} [\mathbf{y}_{k,t} = \tilde{\mathbf{y}}]}{\mathbb{P} [\mathbf{y}'_{k,t} \mathbf{1}_M = N_t]} \\ &= \frac{\mathbb{1}_{\{n=N_t\}} \times p^n (1-p)^{M-n}}{\binom{M}{N_t} p^{N_t} (1-p)^{M-N_t}} = \frac{\mathbb{1}_{\{n=N_t\}}}{\binom{M}{N_t}}. \end{aligned} \quad (\text{OA73})$$

Marginalizing this joint distribution over the entries $j \neq i$ yields the marginal distribution of $y_{i,k,t}$ conditional upon $\sum_{i=1}^M y_{i,k,t} = N_t$,

$$\mathbb{P} [y_{i,k,t} = 1 \mid \mathbf{y}'_{k,t} \mathbf{1}_M = N_t] = \mathbb{P} \left[y_{i,k,t} = 1 \middle| \sum_{j \neq i} y_{j,k,t} = N_t - 1 \right] = \sum_{h \neq i} \frac{\mathbb{1}_{\{\sum_{j \neq h} y_{j,k,t} = N_t - 1\}}}{\binom{M}{N_t}} = \frac{\binom{M-1}{N_t-1}}{\binom{M}{N_t}}. \quad (\text{OA74})$$

Plugging (OA74) into (OA72) yields the desired result and concludes the proof.

References in Online Appendix

- Bai, J., and S. Ng. “Determining the number of factors in approximate factor models.” *Econometrica*, 70 (2002), 191–221.
- Da, R.; S. Nagel; and D. Xiu. “The statistical limit of arbitrage.” *NBER working paper*, 33070 (2024).
- El Karoui, N. “High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation.” *Annals of Statistics*, 38 (2010), 3487–3566.
- El Karoui, N. “On the realized risk of high-dimensional Markowitz portfolios.” *SIAM Journal on Financial Mathematics*, 4 (2013), 737–783.
- Engle, R., and B. Kelly. “Dynamic equicorrelation.” *Journal of Business & Economic Statistics*, 30 (2012), 212–228.
- Gandy, A., and L. Veraart. “The effect of estimation in high-dimensional portfolios.” *Mathematical Finance*, 23 (2013), 531–559.
- Kan, R., and N. Lassance. “Optimal portfolio choice with fat tails and parameter uncertainty.” *Journal of Financial and Quantitative Analysis*, forthcoming (2025).
- Kan, R., and X. Wang. “Optimal portfolio choice with unknown benchmark efficiency.” *Management Science*, 70 (2023), 6117–6138.
- Kan, R.; X. Wang; and G. Zhou. “Optimal portfolio choice with estimation risk: No risk-free asset case.” *Management Science*, 68 (2021), 2047–2068.
- Kan, R., and G. Zhou. “Optimal portfolio choice with parameter uncertainty.” *Journal of Financial and Quantitative Analysis*, 42 (2007), 621–656.
- Ledoit, O., and M. Wolf. “A well-conditioned estimator for large-dimensional covariance matrices.” *Journal of Multivariate Analysis*, 88 (2004), 365–411.
- Ledoit, O., and M. Wolf. “Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks.” *Review of Financial Studies*, 30 (2017), 4349–4388.
- Ledoit, O., and M. Wolf. “Analytical nonlinear shrinkage of large-dimensional covariance matrices.” *Annals of Statistics*, 48 (2020), 3043–3065.
- MacKinlay, A., and L. Pástor. “Asset pricing models: Implications for expected returns and portfolio selection.” *Review of Financial Studies*, 13 (2000), 883–916.
- Tu, J., and G. Zhou. “Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies.” *Journal of Financial Economics*, 99 (2011), 204–215.