OPTIMAL PORTFOLIO SIZE UNDER PARAMETER UNCERTAINTY

Rodolphe Vanderveken, Nathan Lassance, Frédéric Vrins

LIDAM Discussion Paper LFIN
2024 / 04

UCLouvain

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

LFIN

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

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

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

Optimal Portfolio Size under Parameter Uncertainty*

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

July 5, 2024

Abstract

Estimation risk in portfolio selection can be mitigated with sparse approaches such as lasso that penalizes for the norm of the portfolio weights and excludes assets from the investment universe. The latter are revealed *a posteriori*, by identifying which assets receive an optimal weight of zero. We show instead that in the presence of parameter uncertainty, it is desirable to remove assets *before* computing the portfolio weights. In particular, we show that the optimal portfolio size strikes a tradeoff between accessing additional investment opportunities and limiting estimation risk. Our approach disentangles the determination of the optimal portfolio size from the asset selection rule, making it more easily implementable and robust to estimation risk than alternative sparse methods. Empirically, our restricted portfolios substantially outperform their counterparts applied to all available assets. Our methodology renders portfolio theory valuable even when the full dataset dimension is comparable to the sample size.

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

JEL Classification: G11, G12.

*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`. We thank Richard B. Evans, Mohamad Hassan Abou Daya, Aleksy Leeuwenkamp, Marco Saerens, Majeed Simaan, and participants at the 17th International Risk Management Conference (IRMC), 6th Quantitative Finance and Financial Econometrics International Conference (QFFE), 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.

1 Introduction

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

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

Sparse portfolio selection aims at alleviating parameter uncertainty precisely by reducing the portfolio size. This is typically achieved by imposing portfolio norm penalties, constraints based on lasso (DeMiguel et al., 2009a; Fan et al., 2012; Yen, 2016; Ao et al., 2019) or cardinality constraints (Gao and Li, 2013; Du et al., 2023). However, such methods suffer from five main drawbacks. First, the optimization programs must be solved in dimension M, hence, face large estimation risk, even if the output portfolio is ultimately of lower dimension. Second, they can be computationally intensive to implement especially in high dimension. For

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

instance, cardinality constraints render the portfolio problem NP-hard. Third, they entangle the selection of the assets with the computation of portfolio weights in one single step. Therefore, sparse methods do not allow for different rebalancing frequencies for portfolio weights and the asset selection, or for flexibility in the asset selection. Fourth, they typically feature extra parameters such as a penalty strength often estimated by cross-validation, which is time consuming and adds an extra layer of estimation risk. Lastly, the resulting portfolio size N is implicit from the portfolio weights and does not have a clear meaning.

We propose a sequential *three-stage process* which addresses addresses the main drawbacks of sparse portfolio selection listed above. First, given a portfolio strategy, compute an *optimal* portfolio size $N \leq M$. Second, select the N assets to invest in among the M available ones. Third, compute the portfolio weights in dimension N instead of M. Interestingly, Ao et al. (2019) also propose such a sequential process for reasons of computational efficiency, in which they select a chosen number of N assets to maximize the Sharpe ratio and then implement their MAXSER strategy on this subset of assets. Specifically, among the S&P500 constituents, they *arbitrarily* select $N = 50$ of them. In contrast, we show how to select an *optimal* N for several portfolio rules based on estimation risk considerations.

Two key questions remain to implement our three-stage process: What is the optimal portfolio size N? How should we then select the N assets? To find the optimal N, we consider a classical setting in which investors are expected utility maximizers with mean-variance (MV) preferences, there are no investment constraints, and returns are serially independent and identically elliptically distributed. In this setting, parameter uncertainty stems from the unknown mean and covariance matrix of returns. We consider several MV portfolio rules and find, for each of them, the optimal portfolio size N that strikes a tradeoff between two opposite goals: increasing investment opportunities and reducing estimation risk. This N is *optimal* in the sense that it aims to maximize 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). To then identify which N assets to select, we propose a set of 10 simple and sensible asset selection rules. We observe empirically that determining the optimal N using our theory and then selecting the N assets using these selection rules significantly outperforms investing in all M assets in most configurations (about 90%).

Our approach has five main benefits compared to the aforementioned sparse methods. First, because we restrict portfolio size *before* computing the portfolio weights, these weights depend on a limited number of parameters, and thus, face less estimation risk. Second, our approach is less computationally expensive because our optimal N can be computed very efficiently and the portfolio weights are then computed on a universe of smaller dimension. Third, our optimal N depends on the data, the portfolio strategy, and the objective function, but is agnostic with respect to the selection rule determining which N assets to invest in, i.e., the investor has flexibility regarding asset selection. This allows different rebalancing frequencies for portfolio weights and the asset selection, which is valuable in practice.[2] It also means that our optimal N could be used as threshold in cardinality-constrained portfolios, for instance. Fourth, our approach does not require the specification of any extra parameters such as a penalty strength. Lastly, the optimal N has a clear meaning from trading off between diversification benefits and estimation risk when maximizing expected utility.

Our theoretical analysis starts with the sample mean-variance (SMV) portfolio, which is optimal *in sample* but not out of sample. Assuming that asset returns are equicorrelated,[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 the

[2]We rebalance the weights every month but the assets every year, which reduces asset turnover and beats selecting assets every month. Still, we show in Section D.7 of the supplementary material that our optimal N outperforms investing in all M assets also when the selected assets change every month. One cannot do this with norm or cardinality constraints because the selection of assets and their weights is simultaneous.

[3]Assuming equal correlations allows us to express the EU of our different MV portfolio rules as an explicit function of N. Moreover, it makes our optimal portfolio size N less sensitive to estimation errors by replacing many correlation parameters by a single parameter, which is useful because estimation errors in correlations largely impact portfolio performance (Ledoit and Wolf, 2003; Chung et al., 2022). Still, our analysis shows that our method for determining N performs well also when returns are not equicorrelated.

fat tails of elliptical returns.[4] We show that because the SMV portfolio is highly sensitive to estimation errors, its optimal portfolio size N is typically very small. As a remedy for the large estimation risk faced by the SMV portfolio, Kan and Zhou (2007) introduce a *two-fund rule* that optimally scales down the SMV portfolio by combining it with the risk-free asset so as to maximize the EU; a portfolio rule that Kan and Lassance (2023) extend to the case with elliptically distributed returns. Building on this research, we also derive the optimal N for the two-fund rule. Remarkably, for typical levels of rather high correlations observed in equity data, we can show that this optimal N is slightly below *half the sample size*. Given commonly used sample sizes (e.g., 120 months), this result means that it is optimal to hold a limited number of assets in the portfolio to optimize out-of-sample performance.

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

We test our theory first via a simulation experiment in which we evaluate the performance of our optimal N when (i) the optimal N is subject to estimation errors (bona fide estimator), (ii) asset returns are not equicorrelated, and (iii) the decision of which N assets to select out of the M available ones is done randomly. The main conclusion is that the two- and

[4] Most of the literature on portfolio choice with estimation risk assumes normally distributed returns. Our incorporation of elliptical fat tails builds on the recent work by Kan and Lassance (2023).

three-fund rules implemented with our optimal portfolio size N deliver an EU close to the maximum and substantially larger than that obtained when investing either in all M assets or in too few assets. This shows that our method delivers satisfying performance even when relaxing the assumptions underlying our theory, highlighting its practical relevance.

Finally, we turn to empirical data. We consider six datasets of characteristic and industry-sorted test assets, with M around 100. Using a rolling window of 120 months, we compare the out-of-sample utility net of costs of different strategies, including our optimally restricted two-fund and three-fund rules. For the latter, we propose 10 simple and sensible *selection rules* to decide which N assets to keep in the portfolio. We find large benefits from restricting the portfolio size N with our theory, as the two-fund and three-fund rules implemented with the 10 asset selection rules nearly consistently outperform the same portfolio rules applied to all M assets. The improvement is particularly large and statistically significant when we construct the portfolios with a shrinkage covariance matrix. Moreover, for some of the asset selection rules, the optimal two-fund and three-fund rules manage to consistently outperform EW and SGMV portfolios (with or without a risk-free asset), which is a notoriously difficult task when the dataset dimension M is comparable to the sample size, as well as benchmark MV portfolios that reduce the number of assets using weight thresholds or norm constraints.

The 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.

2 Optimal portfolio size without parameter uncertainty

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

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

$$\boldsymbol{w}^\star = \arg\max_{\boldsymbol{w}\in\mathbb{R}^N} U(\boldsymbol{w}) = \arg\max_{\boldsymbol{w}\in\mathbb{R}^N} \boldsymbol{w}'\boldsymbol{\mu}_N - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}_N\boldsymbol{w}, \tag{1}$$

where $U(\boldsymbol{w})$ is the MV utility of portfolio $\boldsymbol{w}$. Solving (1) yields the MV portfolio and the maximum achievable utility,

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

where θ_N is the maximum Sharpe ratio. Clearly, $U(\boldsymbol{w}^\star)$ is non-decreasing with N because, at worst, one can set the weight of an additional asset to zero and keep the same utility. Thus, without parameter uncertainty, it is optimal to consider the whole investment universe and set $N = M$. To relate $U(\boldsymbol{w}^\star)$ and N more explicitly, we assume the following about $\boldsymbol{\Sigma}_M$. We denote $\boldsymbol{I}_M$ the $(M \times M)$ identity matrix, and $\mathbf{1}_M$ the $(M \times 1)$ vector of ones.

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

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

where $-\frac{1}{M-1} < \rho < 1$ so that $\boldsymbol{\Sigma}_M$ is positive definite.

Under Assumption 1, the dependence structure of the returns is controlled by a single parameter ρ. It is useful under parameter uncertainty because estimation errors in correlations are known to hurt performance substantially (Ledoit and Wolf, 2003; Chung et al., 2022). It also follows that for any $N \leq M$, $\boldsymbol{\Sigma}_N$ admits the same form as in Assumption 1.

The next proposition provides the analytical expression of $U(\boldsymbol{w}^\star)$ under Assumption 1.[5]

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

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

where $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$ *is defined as*

$$\bar{\theta}_{N,k} = \frac{1}{N}\sum_{i=1}^{N} s_i^k. \tag{5}$$

Moreover, $U(\boldsymbol{w}^\star)$ *is non-decreasing in* N *and is strictly increasing from* N *to* $N+1$ *whenever* $s_{N+1} \neq \rho N \bar{\theta}_{N,1}/(1-\rho+N\rho)$.

Proposition 1 shows that the maximum utility depends on the assets' correlation structure via the correlation parameter ρ and on the assets' Sharpe ratios via $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$. Moreover, it is non-decreasing in N because including an additional asset in the portfolio cannot decrease the utility. In Figure 1, we depict $U(\boldsymbol{w}^\star)$ as a function of $N \in \{1, \ldots, M\}$ for $\rho \in \{0.2, 0.5, 0.8\}$ and assets' monthly Sharpe ratios θ_i calibrated to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.[6]

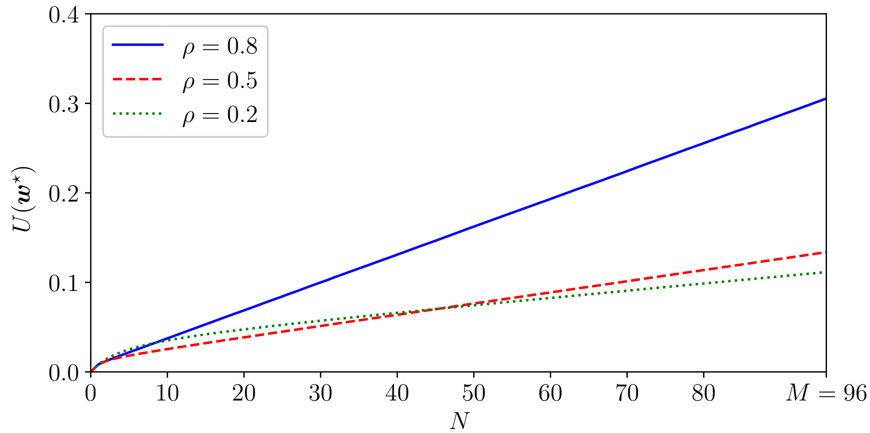
3 Optimal portfolio size under parameter uncertainty

We show in Section 2 that the utility of the MV portfolio without parameter uncertainty is non-decreasing with the portfolio size N. Intuitively, this is because excluding *a priori* some assets from the investment strategy amounts to adding a constraint to the optimization problem, which can only deteriorate the solution. In this section, we show that this no longer holds when the assets' parameters are unknown and the investor relies on sample estimates.

[5] The proofs of all results are available in Section E of the supplementary material.

[6] We remove four portfolios sorted on size and book-to-market for which there is missing data.

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



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

3.1 Sample estimates and distributional assumption

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

$$\hat{\boldsymbol{\mu}}_N = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{r}_t \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_N = \frac{1}{T}\sum_{t=1}^{T} (\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)' \tag{6}$$

be sample estimates of $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$, and we require $T > N$ so that $\hat{\boldsymbol{\Sigma}}_N$ is invertible. We study the optimal portfolio size N for *sample portfolios* implemented by relying on $\hat{\boldsymbol{\mu}}_N$ and $\hat{\boldsymbol{\Sigma}}_N$. Given a sample portfolio $\hat{\boldsymbol{w}}$, we define its optimal size $N^\star$ as the N maximizing the *expected out-of-sample utility (EU)* performance measure pioneered by Kan and Zhou (2007),

$$EU(\hat{\boldsymbol{w}}) = \mathbb{E}[U(\hat{\boldsymbol{w}})] = \mathbb{E}\left[\hat{\boldsymbol{w}}'\boldsymbol{\mu}_N - \frac{\gamma}{2}\hat{\boldsymbol{w}}'\boldsymbol{\Sigma}_N\hat{\boldsymbol{w}}\right]. \tag{7}$$

Kan and Zhou (2007) and follow-up papers like Tu and Zhou (2011), DeMiguel et al. (2015), Kan et al. (2021), Lassance et al. (2024b,a), and Yuan and Zhou (2024) all take the expectation assuming that asset returns are normally distributed, which is analytically useful but unrealistic in practice. Therefore, we follow Kan and Lassance (2023) who study the EU of various sample portfolio rules when asset returns are *elliptically* distributed. Like them, we use the representation of the elliptical distribution in El Karoui (2010, 2013).

Assumption 2 *Asset returns are elliptically distributed, namely*

$$\boldsymbol{r} \stackrel{d}{=} \boldsymbol{\mu}_M + (\tau \boldsymbol{\Sigma}_M)^{1/2} \boldsymbol{z}_M, \tag{8}$$

where $\boldsymbol{z}_M \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{I}_M)$, *$\tau$ is a positive random variable satisfying* $\mathbb{E}[\tau] = 1$, *and* $\boldsymbol{z}_M \perp \tau$.[7]

The distribution of τ in Assumption 2 captures the effect of fat tails. For example, we recover the normal distribution when $\tau = 1$ collapses to a constant, and the t-distribution when $\tau \sim (\nu - 2)/\chi^2_\nu$, where χ^2_ν denotes a chi-square distribution with $\nu > 2$ degrees of freedom. We focus on the elliptical distribution because it is consistent with MV portfolios (Chamberlain, 1983; Schuhmacher et al., 2021) and because Kan and Lassance (2023) show that it is crucial to account for fat tails in determining optimal portfolio combination rules.

3.2 Optimal portfolio size for sample portfolios

The first strategy for which we study the optimal portfolio size is the estimated counterpart of the MV portfolio in (2), i.e., the *sample mean-variance (SMV)* portfolio,

$$\hat{\boldsymbol{w}}^\star = \frac{1}{\gamma} \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N. \tag{9}$$

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

[7]Because the multivariate elliptical distribution is closed under marginalization, the joint distribution of any subset of N assets belongs to the same family. The corresponding mean and covariance parameters $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ are obtained by selecting the appropriate entries of $\boldsymbol{\mu}_M$ and $\boldsymbol{\Sigma}_M$, respectively, while the distribution of τ remains unchanged.

rule is of the form

$$\hat{\boldsymbol{w}}(\alpha) = \alpha\hat{\boldsymbol{w}}^{\star} = \frac{\alpha}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N, \tag{10}$$

where $\alpha \in \mathbb{R}$ is the *combination coefficient*. Note that the SMV portfolio in (9) corresponds to $\alpha = 1$, i.e., $\hat{\boldsymbol{w}}(1) = \hat{\boldsymbol{w}}^{\star}$.

In the next proposition, we derive the EU of the SMV portfolio and the two-fund rule when asset returns are elliptically distributed, the optimal combination coefficient $\alpha^{\star}$, and which EU it delivers. These results then allow us to determine the optimal portfolio size N. In Section A of the supplementary material, we follow the same approach for three-fund rules. Specifically, a three-fund rule based on the GMV portfolio as in, e.g., Kan and Zhou (2007), DeMiguel et al. (2015), and Yuan and Zhou (2024), which we call the *GMV-three-fund rule*, and a three-fund rule based on the EW portfolio as in, e.g., Tu and Zhou (2011), Kan and Wang (2023), and Lassance et al. (2024b), which we call the *EW-three-fund rule*.

Proposition 2 *Let $T > N + 4$, $\boldsymbol{M} = \boldsymbol{I}_T - \mathbf{1}_T\mathbf{1}_T'/T$, $\boldsymbol{Z}_N$ be a $T \times N$ matrix of independent standard normal variables, $\boldsymbol{\Lambda}$ be a diagonal matrix of T independent copies of $\tau^{1/2}$, and $\boldsymbol{\Lambda} \perp \boldsymbol{Z}_N$. Then, under Assumption 2, the EU of the two-fund rule $\hat{\boldsymbol{w}}(\alpha)$ in (10) is*

$$EU(\hat{\boldsymbol{w}}(\alpha)) = \frac{1}{2\gamma}\frac{T}{T-N-2}\left[\left(2\alpha\kappa_{N,1} - \frac{c_N\alpha^2\kappa_{N,2}T}{T-N-2}\right)\theta_N^2 - \frac{c_N\alpha^2\kappa_{N,3}N}{T-N-2}\right], \tag{11}$$

where θ_N^2 is the maximum squared Sharpe ratio given in (4), c_N is given by

$$c_N = \frac{(T-2)(T-N-2)}{(T-N-1)(T-N-4)}, \tag{12}$$

and $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$ are functions of only N, T, and the distribution of τ:

$$\kappa_{N,1} = \frac{T-N-2}{N}\mathbb{E}\left[\operatorname{tr}\left((\boldsymbol{Z}_N'\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda}\boldsymbol{Z}_N)^{-1}\right)\right], \tag{13}$$

$$\kappa_{N,2} = \frac{(T-N-2)^2}{c_N N}\mathbb{E}\left[\operatorname{tr}\left((\boldsymbol{Z}_N'\boldsymbol{\Lambda}\boldsymbol{M}\boldsymbol{\Lambda}\boldsymbol{Z}_N)^{-2}\right)\right], \tag{14}$$

$$\kappa_{N,3} = \frac{(T-N-2)^2}{c_N NT}\mathbb{E}\left[\mathbf{1}_T'\mathbf{\Lambda}\boldsymbol{Z}_N(\boldsymbol{Z}_N'\mathbf{\Lambda}\boldsymbol{M}\mathbf{\Lambda}\boldsymbol{Z}_N)^{-2}\boldsymbol{Z}_N'\mathbf{\Lambda}\mathbf{1}_T\right]. \tag{15}$$

The EU of the SMV portfolio is obtained as $EU(\hat{\boldsymbol{w}}^\star) = EU(\hat{\boldsymbol{w}}(1))$*. Moreover, the optimal combination coefficient* $\alpha^\star$ *maximizing* (11) *is*

$$\alpha^\star = \frac{T-N-2}{c_N T}\frac{\kappa_{N,1}\theta_N^2}{\kappa_{N,2}\theta_N^2 + \kappa_{N,3}\frac{N}{T}}, \tag{16}$$

and the resulting EU of the optimal two-fund rule is

$$EU(\hat{\boldsymbol{w}}(\alpha^\star)) = \frac{1}{2\gamma c_N}\frac{\kappa_{N,1}^2\theta_N^4}{\kappa_{N,2}\theta_N^2 + \kappa_{N,3}\frac{N}{T}}. \tag{17}$$

Three comments are in order. First, Proposition 2 is valid for any covariance matrix $\mathbf{\Sigma}_N$, not only those of the form $\mathbf{\Sigma}_N(\rho)$. 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 C of the supplementary material. Kan and Lassance (2023) show that they increase with estimation risk N/T and the tail heaviness. Third, because uncertainty in the parameters $\boldsymbol{\mu}_N$ and $\mathbf{\Sigma}_N$ impacts the EU, we no longer expect a larger N to always be favorable as it was the case in Section 2. Instead, when selecting the value of N maximizing (11), there is a tradeoff between improving the population performance (i.e., increasing θ_N^2) and limiting estimation risk (i.e., decreasing N/T).

Next, we proceed as in Section 2 and relate the EU with N more explicitly by assuming that the covariance matrix $\mathbf{\Sigma}_N$ complies with Assumption 1, i.e., $\mathbf{\Sigma}_N = \mathbf{\Sigma}_N(\rho)$.

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

$$EU(\hat{\boldsymbol{w}}^\star) = \frac{1}{2\gamma}\frac{NT}{T-N-2}\left[\frac{\delta_N(\bar{\boldsymbol{\theta}}_N)}{1-\rho}\left(2\kappa_{N,1} - \frac{c_N\kappa_{N,2}T}{T-N-2}\right) - \frac{c_N\kappa_{N,3}}{T-N-2}\right] \tag{18}$$

$$=: EU_{smv}(N,T,\rho,\tau,\bar{\boldsymbol{\theta}}_N),\ \textit{and} \tag{19}$$

$$EU(\hat{\boldsymbol{w}}(\alpha^\star)) = \frac{N}{2\gamma c_N(1-\rho)} \times \frac{\kappa_{N,1}^2\delta_N(\bar{\boldsymbol{\theta}}_N)^2}{\kappa_{N,2}\delta_N(\bar{\boldsymbol{\theta}}_N) + \kappa_{N,3}\frac{1-\rho}{T}} =: EU_{2f}(N,T,\rho,\tau,\bar{\boldsymbol{\theta}}_N). \tag{20}$$

Corollary 1 shows that both EU_{smv} and EU_{2f} functions depend on several parameters.[8] First, the sample size T, directly but also via 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 involved in the portfolio via $\bar{\boldsymbol{\theta}}_N = (\bar{\theta}_{N,1}, \bar{\theta}_{N,2})$.

Now, to maximize $EU_{smv}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ or $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ with respect to N, one would also need to optimize the selection of assets at the same time because $\bar{\boldsymbol{\theta}}_N$ depends on N via the chosen subset of assets. This is a difficult optimization problem because the number of possible subsets of N assets with N going from one to M grows very quickly with M.[9] Moreover, this would introduce potentially severe estimation risk and time instability in the selection of assets, which is not desirable in practice. This is why, in order to disentangle the determination of the optimal N from the asset selection, we replace $\bar{\boldsymbol{\theta}}_N$ by its counterpart for the whole investment universe, i.e., $\bar{\boldsymbol{\theta}}_M$. This approximation may not be accurate when N is small relative to M.[10] In this case, however, the estimated $\bar{\boldsymbol{\theta}}_N$ will be particularly noisy, and thus estimating it from $\bar{\boldsymbol{\theta}}_M$ can help too. All in all, the way we determine the optimal portfolio size N for the SMV portfolio and the optimal two-fund rule becomes

$$N^\star_{smv} = \underset{N \in \{1, \ldots, \min(M, T-5)\}}{\text{argmax}} EU_{smv}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M), \text{ and} \tag{21}$$

$$N^\star_{2f} = \underset{N \in \{1, \ldots, \min(M, T-5)\}}{\text{argmax}} EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M), \tag{22}$$

where we search N up to $\min(M, T-5)$ because we assume $T > N + 4$ in Proposition 2.[11]

---

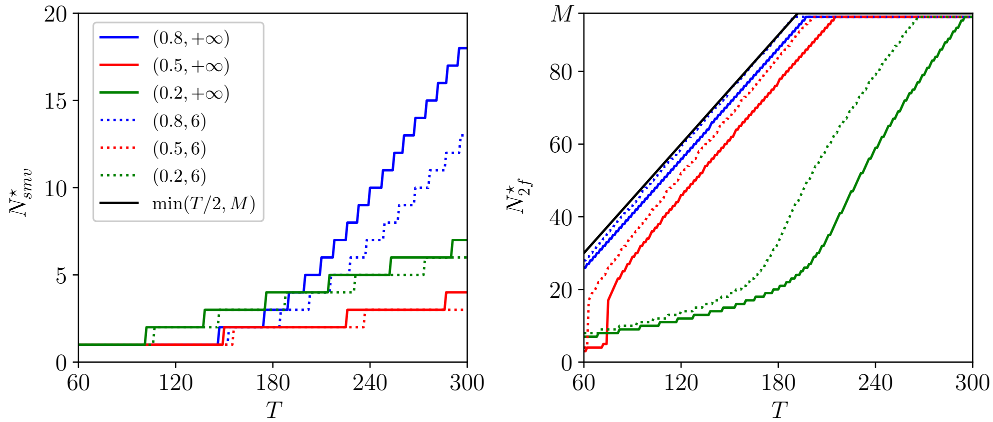
[8] The EU of the SMV portfolio and the two-fund rule is proportional to $1/\gamma$, and thus, the N optimizing the EU does not depend on γ. Therefore, we do not highlight the dependence on γ in the functions EU_{smv} and EU_{2f}. This is also true for the three-fund rules we consider in Section A of the supplementary material.

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

[10] In Section B of the supplementary material, we show that $\bar{\boldsymbol{\theta}}_N$ quickly converges to $\bar{\boldsymbol{\theta}}_M$ in practice.

[11] In Section B of the supplementary material, we compare for the optimal two-fund rule the dependence on N of the EU obtained with $\bar{\boldsymbol{\theta}}_M$ and $\bar{\boldsymbol{\theta}}_N$, i.e., $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ and $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ in (20). The function $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ depends on N via the sequence of assets considered, and thus, we generate a large number of random asset sequences, evaluate $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ for each sequence, and compare it to $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$. We find that $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ provides a good approximation and that the dispersion of $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ across the asset sequences is reasonable.

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



Notes. This figure depicts the optimal portfolio size for the SMV portfolio, $N^\star_{smv}$ in (21) (left panel), and for the optimal two-fund rule, $N^\star_{2f}$ in (22) (right panel), as a function of the sample size T. Each line represents a different choice of the correlation, $\rho = \{0.2, 0.5, 0.8\}$, and the degrees of freedom of the t-distribution, $\nu = \{6, \infty\}$. We calibrate $\bar{\boldsymbol{\theta}}_M = (0.125, 0.0169)$ to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.

We now illustrate the optimal portfolio sizes $N^\star_{smv}$ and $N^\star_{2f}$. Using the full sample of the 96S-BM dataset considered previously, we compute the sample estimators $\hat{\boldsymbol{\mu}}_{96}$ and $\hat{\boldsymbol{\Sigma}}_{96}$ using (6). Then, we set the population mean $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and the population covariance matrix $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\rho) = \hat{\boldsymbol{D}}_{96}\boldsymbol{P}_{96}(\rho)\hat{\boldsymbol{D}}'_{96}$ with $\hat{\boldsymbol{D}}_{96} = \mathrm{diag}(\hat{\boldsymbol{\Sigma}}_{96})^{1/2}$, which yields $\bar{\boldsymbol{\theta}}_{96} = (0.125, 0.0169)$, and we take $\rho \in \{0.2, 0.5, 0.8\}$. We consider two elliptical distributions. First, the normal distribution, i.e., $\tau = 1$ and $\kappa_{N,1} = \kappa_{N,2} = \kappa_{N,3} = 1$. Second, the t-distribution, i.e., $\tau \sim (\nu - 2)/\chi^2_\nu$, with $\nu = 6$. Figure 2 depicts how $N^\star_{smv}$ and $N^\star_{2f}$ depend on the sample size T. We expect both to increase with T because, as T increases, the estimated portfolios become better estimates of the true optimal MV portfolio for which the optimal N is unbounded.

We make three observations from Figure 2. First, we find that $N^\star_{smv}$ is very small compared to T and M. For instance, with $T = 240$ and $\rho = 0.5$, we have $N^\star_{smv} = 3$ both for the normal and t-distributions. For a larger $\rho = 0.8$, we find a larger $N^\star_{smv}$ but it is still small. For example, when $\rho = 0.8$ and $T = 120$, 180 and 240, we have $N^\star_{smv} = 1$, 3, and 10 for the normal distribution, and $N^\star_{smv} = 1$, 2, and 7 for the t-distribution, respectively.

Second, $N^\star_{2f}$ is higher than $N^\star_{smv}$, and is below $T/2$, or equal to M if $T/2$ is enough above M. Also, $N^\star_{2f}$ gets closer to $T/2$ as ρ increases. These results can be explained based on Equations (20) and (22). Specifically, we can show that in the normal case, i.e., $\kappa_{N,1} = \kappa_{N,2} = \kappa_{N,3} = 1$, $N^\star_{2f}$ is below half of the sample size $T/2$ (or equal to M if M is sufficiently below $T/2$), and close to $T/2$ as ρ is either close to zero or one. To see this, consider the N that maximizes the first ratio of EU_{2f}, i.e., N/c_N. We have

$$\frac{N}{c_N} = \frac{N(T-N-1)(T-N-4)}{(T-2)(T-N-2)} \approx \frac{N(T-N-4)}{(T-2)}, \tag{23}$$

and the latter is maximized by

$$\underset{N\in\{1,\ldots,\min(M,T-5)\}}{\operatorname{argmax}} \frac{N(T-N-4)}{(T-2)} = \min([T/2-2], M). \tag{24}$$

Turning to the second ratio of EU_{2f} in (20), it is easy to show that if $\kappa_{N,1} = \kappa_{N,2} = \kappa_{N,3} = 1$, it is a decreasing function of N and thus is maximized by $N = 1$, which moves $N^\star_{2f}$ below (24). However, the sensitivity of this second term to N depends on how sensitive $N\rho/(1-\rho+N\rho)$ is to N, and it is less so as ρ approaches zero or one. In equity data where correlations are typically large, we thus expect $N^\star_{2f}$ to stay close to (24). Finally, in the elliptical case, $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$ are different from one and depend on N, but Figure 2 and later simulations suggest that $N^\star_{2f}$ still remains close to that under normality.

Third, fat tails positively impact $N^\star_{2f}$, while we observe the opposite for $N^\star_{smv}$. This can be explained because the two-fund rule optimally scales down the SMV portfolio according to tail heaviness via $\kappa_{N,1}$, $\kappa_{N,2}$, and $\kappa_{N,3}$. Therefore, as shown by Kan and Lassance (2023, Proposition 5), the two-fund rule often performs better when asset returns are elliptically instead of normally distributed, and thus we find a larger $N^\star_{2f}$ under elliptical returns.

4 Simulation analysis

We now run simulations to analyze two questions. First, in Section 4.1, what is the magnitude and variability of estimated optimal values of N for the SMV portfolio and the two-fund and three-fund rules. Second, in Section 4.2, whether we obtain a portfolio performance close to optimal even when we use *estimated* optimal values of N and that Assumption 1, i.e., asset returns are equicorrelated, is not fulfilled. To answer both questions, we draw monthly excess returns from a t-distribution, i.e., $\tau \sim (\nu - 2)/\chi^2_\nu$, with degrees of freedom $\nu \in \{6, \infty\}$. We calibrate the mean and covariance matrix of asset returns from the 96S-BM dataset as previously.[12] 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 $\hat{\boldsymbol{\Sigma}}_{96}$.[13] The second choice of $\boldsymbol{\Sigma}_M$ allows us to compare the estimated optimal values of N with the true optimal one that, for each portfolio strategy, is known theoretically only under Assumption 1, i.e., when $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$.

4.1 Estimated optimal portfolio size

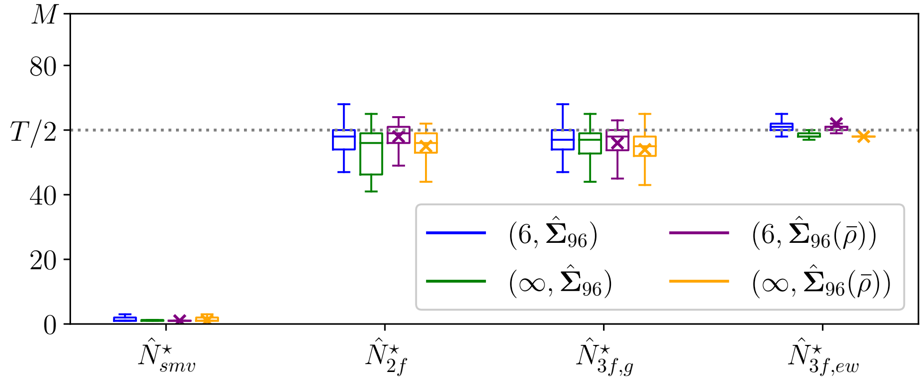
We now analyze the estimated optimal portfolio size for the SMV portfolio ($\hat{N}^\star_{smv}$), the two-fund rule ($\hat{N}^\star_{2f}$), the GMV-three-fund rule ($\hat{N}^\star_{3f,g}$), and the EW-three-fund rule ($\hat{N}^\star_{3f,ew}$). These are the estimated counterparts of $N^\star_{smv}$ in (21), $N^\star_{2f}$ in (22), $N^\star_{3f,g}$ in (SM11), and $N^\star_{3f,ew}$ in (SM22) following the methodology detailed in Section C of the supplementary material.

We simulate $K = 10,000$ times $T = 120$ returns and estimate, for each strategy, the optimal portfolio sizes in each of the K simulations. Figure 3 depicts boxplots of the estimated optimal N for the different choices of (ν,$\boldsymbol{\Sigma}_M$) across all simulations. We make a number of observations. First, in the case $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, our $\hat{N}^\star$ are close to the true $N^\star$ on average for

[12] Running our simulation analysis on the five additional datasets we consider in the empirical analysis of Section 5 yields similar insights, and thus, we focus on the 96S-BM dataset for conciseness.

[13] In Section D.1 of the supplementary material, instead of fixing $\rho = \bar{\rho}$, we study how the estimated optimal portfolio size varies with ρ.

Figure 3: Estimated optimal portfolio size in simulated data



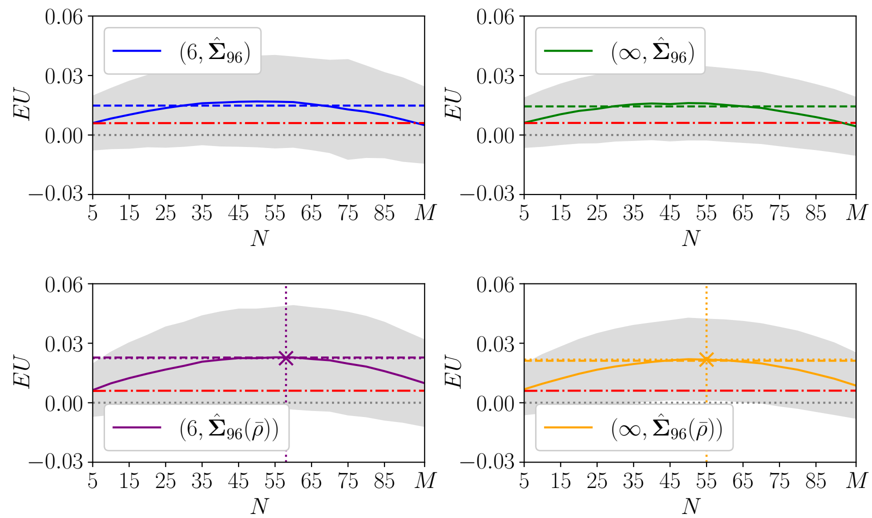
Notes. This figure depicts boxplots of the estimated optimal portfolio size N for the SMV portfolio ($\hat{N}^\star_{smv}$), two-fund rule ($\hat{N}^\star_{2f}$), GMV-three-fund rule ($\hat{N}^\star_{3f,g}$), and EW-three-fund rule ($\hat{N}^\star_{3f,ew}$). These are the estimated counterparts of $N^\star_{smv}$ in (21), $N^\star_{2f}$ in (22), $N^\star_{3f,g}$ in (SM11), and $N^\star_{3f,ew}$ in (SM22) following the estimation methodology in Section C of the supplementary material. The boxplots are obtained by simulating 10,000 times $T = 120$ t-distributed returns. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\mathbf{\Sigma}_M \in \{\hat{\mathbf{\Sigma}}_{96}, \hat{\mathbf{\Sigma}}_{96}(\bar{\rho})\}$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\mathbf{\Sigma}}_{96}$, and consider $\nu \in \{6, \infty\}$ degrees of freedom. Each boxplot corresponds to a different choice of $(\nu, \mathbf{\Sigma}_M)$. We depict with a '$\times$' marker the oracle value of the optimal N known under Assumption 1, i.e., when $\mathbf{\Sigma}_M$ is of the form $\mathbf{\Sigma}_M(\rho)$.

each strategy, highlighting the quality of the estimation procedure proposed in Section C. Second, the $\hat{N}^\star$ are similar under $\hat{\mathbf{\Sigma}}_{96}$ and $\hat{\mathbf{\Sigma}}_{96}(\bar{\rho})$, meaning that drawing returns from a distribution that does not fulfill Assumption 1 does not significantly affect the estimation of N. Third, the $\hat{N}^\star$ for the SMV portfolio is very small, in line with Figure 2. Fourth, the $\hat{N}^\star$ are similar for the two-fund and three-fund rules, meaning that they are robust to considering different portfolio combination rules, and are close to but typically below $T/2$, which is because $\bar{\rho}$ is rather large as discussed in Section 3.2. Finally, the lower the ν, and thus the fatter the tails, the higher the optimal N in the two-fund and three-fund rules, which is consistent with Figure 2.

4.2 Optimality of restricted portfolios

Section 4.1 shows the estimated optimal portfolio size $\hat{N}^\star$ for various MV combination rules. We now analyze the performance of the corresponding restricted portfolios using $N = \hat{N}^\star$. We observe that even in the unfavorable setting where (i) Assumption 1 is violated, (ii) the optimal N is estimated from the data and (iii) the N assets selected in the restricted portfolios are chosen randomly,[14] the restricted portfolio delivers an EU close to the maximum achievable one, and outperforms on average the naive equally-weighted portfolio which is not subject to estimation risk. For conciseness, we only consider the two-fund rule in this section; the results are similar for the three-fund rules in Section A of the supplementary material.

Figure 4: Expected out-of-sample utility of the two-fund rule in simulated data



Notes. This figure depicts the expected out-of-sample utility (EU) of the two-fund rule as a function of the portfolio size N. We simulate 10,000 samples of t-distributed returns with $\nu \in \{6, \infty\}$ degrees of freedom. For each N, we conduct a rolling window exercise as described in Section 4.2 and, in each rolling window, we randomly select the N assets out of the M available ones. Using a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\boldsymbol{\Sigma}}_{96}$. Each panel corresponds to a different choice of $(\nu, \boldsymbol{\Sigma}_M)$. In each panel, the solid line depicts the EU in (25), the shaded gray area depicts the one standard deviation interval around the EU across all simulations, and the dashed horizontal line depicts the EU obtained when using the estimated optimal N for the two-fund rule, i.e., the estimate of $N^{\star}_{2f}$ in (22) according to the estimation method in Section C of the supplementary material. The dash-dotted red line depicts the EU delivered by the equally weighted portfolio. In the bottom panels, the vertical and horizontal dotted lines depict the oracle N and its EU, respectively, with the '×' marker showing the intersection. The dotted horizontal gray line depicts the zero EU level. We set a risk-aversion coefficient of $\gamma = 1$.

We conduct this analysis as follows. For each choice of $(\nu, \boldsymbol{\Sigma}_M)$, where $\nu \in \{6, \infty\}$ and $\boldsymbol{\Sigma}_M \in \{\hat{\boldsymbol{\Sigma}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})\}$, we draw $K = 10{,}000$ times $L = 300$ t-distributed returns. We then use the L returns in a rolling window exercise. Specifically, for each simulation k, given the sample size $T = 120$ and a portfolio size N, we randomly select at each rolling window a

[14] In 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.

subset of N assets out of the $M = 96$ available ones, compute the estimated two-fund rule $\hat{\boldsymbol{w}}(\hat{\alpha}^{\star})$[15] over these N assets, and evaluate the out-of-sample portfolio return $r_{t,k}$ over the next month. We also implement the naive EW portfolio $\boldsymbol{w}_{ew} = \mathbf{1}_N/N$ on the same N assets for comparison purposes. We then roll the window by one month and proceed similarly until we have used the whole sample of L returns. This procedure gives us, for any N, a time series of out-of-sample portfolio returns $r_{t,k}$, where $t = 1, \dots, L-T$ and $k = 1, \dots, K$, from which we compute the EU,

$$EU = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2\right) \quad \text{with} \quad \hat{\mu}_k = \frac{1}{L-T}\sum_{t=1}^{L-T} r_{t,k}, \quad \hat{\sigma}_k^2 = \frac{1}{L-T}\sum_{t=1}^{L-T}(r_{t,k} - \hat{\mu}_k)^2. \quad (25)$$

We implement this exercise with various portfolio sizes and depict the results in Figure 4. First, for both $\hat{\boldsymbol{w}}(\hat{\alpha}^{\star})$ and $\boldsymbol{w}_{ew}$, we consider fixed values of N ranging from $N = 5$ to $N = M = 96$ with a step size of five (six for the last step), and depict the EU as a function of N with a solid line for $\hat{\boldsymbol{w}}(\hat{\alpha}^{\star})$ and with a dash-dotted red line for $\boldsymbol{w}_{ew}$. Second, for $\hat{\boldsymbol{w}}(\hat{\alpha}^{\star})$ only, we consider the estimated optimal portfolio size $N = \hat{N}^{\star}_{2f,t,k}$ which is not fixed but instead varies depending on the rolling window and the simulation, and depict its EU with a horizontal dashed line. Third, for $\hat{\boldsymbol{w}}(\hat{\alpha}^{\star})$ only and for the two bottom panels, i.e., when $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho})$, we also consider the oracle $N = N^{\star}_{2f}$ obtained under the true parameters $(\hat{\boldsymbol{\mu}}_{96}, \hat{\boldsymbol{\Sigma}}_{96}(\bar{\rho}), \nu)$. In this case, we depict the value of $N^{\star}_{2f}$ with a vertical dotted line and its EU with a horizontal dotted line, and mark the intersection with a '$\times$' marker.[16]

Several observations can be made from Figure 4. First, our estimated optimal portfolio size for the two-fund rule, $\hat{N}^{\star}_{2f}$, delivers an EU that is close to the maximum, *both* for the case in which returns are equicorrelated and in which they are not. Second, we can see that for both $\nu = 6$ and $\nu = \infty$, the EU is maximized when N is slightly below $T/2$, in line

[15] $\hat{\alpha}^{\star}$ is the estimate of the optimal combination coefficient in the two-fund rule, $\alpha^{\star}$ in (16). We detail our estimation methodology in Section C of the supplementary material.

[16] In Section D.2 of the supplementary material, we implement a similar exercise in which the correlation matrix has a block structure. Even in this setup that strongly violates Assumption 1, there are significant gains from reducing the portfolio size from M to $\hat{N}^{\star}_{2f}$.

with the theoretical results derived in Section 3.2 and the fact that the average correlation is large for the 96S-BM dataset ($\bar{\rho} = 0.74$). Third, $\hat{N}^{\star}_{2f}$ delivers an EU that is substantially larger than that when N is either too small or too close to M which highlights the value of choosing the right portfolio size and the performance risk of blindly investing in all available assets. Lastly, $\hat{N}^{\star}_{2f}$ substantially outperforms the naive EW portfolio on average, which is a challenging benchmark strategy when M is comparable to T.

5 Empirical analysis

We now test the performance of our optimally restricted MV portfolios in empirical data. 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 after obtaining the optimal portfolio size. In Section 5.4, we detail the out-of-sample methodology and performance measures. Finally, we report and discuss the results in Section 5.5.

5.1 Datasets

The cross-section of equities is composed of several thousands of stocks, which is a challenging high-dimensional environment. To explain the cross-section more parsimoniously, the standard practice in the asset pricing literature is to value-weight stocks into characteristic portfolios according to firm characteristics that have been shown to explain average returns. Another standard practice is to group stocks into a smaller number of industry portfolios.

Therefore, our empirical analysis builds on six datasets of monthly excess returns of characteristic and industry portfolios, which we list in Table 1. Given that our objective is to optimally reduce the portfolio size, the full investment universe must not be too small in the first place, and thus we consider datasets with M being around 100 assets.

The first dataset already considered previously, 96S-BM, is composed of decile portfolios sorted on size and book-to-market. The second dataset, 108CHA, is composed of the long and short legs of the 54 characteristics considered in Lassance and Martín-Utrera (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 of the long and short legs obtained from the 23 characteristics in Novy-Marx and Velikov (2015), which are available on Robert Novy-Marx's website. The fifth dataset, 107IN-CHA, is composed of 47 industry portfolios,[17] of quintile portfolios sorted on size and book-to-market and on operating profitability and investment, and of decile portfolios sorted on momentum. The sixth dataset, 98IN-CHA-NV, is composed of 47 industry portfolios, of quintile portfolios sorted on size and book-to-market ratio, of decile portfolios sorted on momentum, and of the long and short legs obtained from the eight low-turnover characteristics in Novy-Marx and Velikov (2015). The time period for each dataset is reported in Table 1.

In Section D.3 of the supplementary material, we look at how Assumption 1 (equicorrelated returns) and Assumption 2 (elliptical returns) transpire in the six datasets. First, we find that the equicorrelation assumption is reasonable for the 96S-BM, 100S-OP, and 108CHA datasets, with a root mean squared error (RMSE) between the observed and equicorrelation matrices of around 5-10%, but less so for the three other datasets, with a RMSE around 10-20%. Thus, we can assess the performance of our optimal portfolio size across different correlation structures. Second, we find that the parameters $(\kappa_{M,1}, \kappa_{M,2}, \kappa_{M,3})$ defined in (13)–(15), which control the impact of elliptical fat tails of asset returns on the EU, substantially depart from one, i.e., from normality. Thus, it is crucial to account for fat tails when implementing combination rules given their impact on optimal combination coefficients.

[17]We consider 47 industry portfolios instead of 48 because of missing data in the healthcare portfolio.

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

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, the dataset name, the dataset description, the size of the dataset M, and the time period. The industry portfolios and the portfolios sorted on size, book-to-market, operating profitability, investment, and momentum are downloaded from Kenneth French's website. The characteristic portfolios in Novy-Marx and Velikov (2015) are downloaded from Robert Novy-Marx's website. In the construction of these datasets, we have removed assets with missing data over the time period considered. The 108CHA dataset is the same as that used by Lassance and Martín-Utrera (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.

5.2 Portfolio strategies

We evaluate the out-of-sample performance of the 10 portfolio strategies in Table 2. The first three are the optimal combination rules in Section 3.2 and Section A of the supplementary material: the two-fund rule (2F), the GMV-three-fund rule (3FGMV), and the EW-three-fund-rule (3FEW). We apply these strategies to all M assets or to a subset of N assets, with N chosen optimally as $\hat{N}^{\star}_{2f}$ in (22) for 2F, $\hat{N}^{\star}_{3f,g}$ in (SM11) for 3FGMV, and $\hat{N}^{\star}_{3f,ew}$ in (SM22) for 3FEW, which we estimate following Section C of the supplementary material.

The next two portfolio strategies are individual portfolios: the EW portfolio in (SM12) and the SGMV portfolio in (SM1). We then consider two portfolio strategies that combine

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

Name	Description
2F	Two-fund rule that combines the sample mean-variance portfolio and the risk-free asset. See Equation (10) and Proposition 2.
3FGMV	Three-fund rule that combines the sample mean-variance portfolio, the sample global minimum-variance portfolio, and the risk-free asset. See Equation (SM3) and Proposition SM.1.
3FEW	Three-fund rule that combines the sample mean-variance portfolio, the equally weighted portfolio, and the risk-free asset. See Equation (SM14) and Proposition SM.3.
EW	Equally weighted portfolio. See Equation (SM12).
EWRF	Combination of EW and risk-free asset. See Equation (26).
SGMV	Sample global minimum-variance portfolio. See Equation (SM1).
GMVRF	Combination of SGMV and the risk-free asset. See Equation (27).
SMV	Sample mean-variance portfolio. See Equation (9).
SMV-L_1	SMV portfolio subject to a L_1-norm constraint with the threshold selected via cross-validation. See Equation (28).
SMV-TW	SMV portfolio with a threshold on weights selected via cross-validation.

Notes. This table lists the portfolio strategies we consider in the empirical analysis, which we detail in Section 5.2. We estimate the covariance matrix in these portfolio strategies either with the sample estimator in (6), the linear shrinkage estimator of Ledoit and Wolf (2004), or the nonlinear shrinkage estimator of Ledoit and Wolf (2020). The results for the latter are available in Section D.4 of the supplementary material. We apply the 2F, 3FGMV, 3FEW, and EW portfolio strategies on a subset of all M assets, and the asset selection rules we implement for that purpose are described in Section 5.3.

EW and SGMV with the risk-free asset to maximize the EU, EWRF and GMVRF, which might be preferred to the 2F, 3FGMV, and 3FEW combination rules because they disregard the SMV portfolio that faces high estimation risk. We implement those as

$$\hat{w}_{ewrf} = \frac{\hat{\lambda}_{ew,N}}{\gamma}\frac{\mathbf{1}_N}{N} \qquad (26) \qquad \text{and} \qquad \hat{w}_{gmvrf} = \frac{\hat{\mu}_{g,N}}{c_N\gamma}\hat{\mathbf{\Sigma}}^{-1}\mathbf{1}, \qquad (27)$$

where $\hat{\lambda}_{ew,N}$ and $\hat{\mu}_{g,N}$ are defined in (SM30)–(SM31).

The last portfolio strategy we consider is the SMV portfolio, which we implement in three different ways. In the first case, we compute the SMV portfolio on all M assets using Equation (9). In the second and third case, we reduce the portfolio size using two methods in which either the estimation of weights and the choice of N are achieved in one step, or N is chosen after having first computed the SMV portfolio weights on all M assets.[18]

[18]In Section D.5 of the supplementary material, we also report the performance of the restricted SMV portfolio whose portfolio size is determined using our theory in Section D.5, i.e., using $\hat{N}^{\star}_{smv}$, and the asset

The first SMV portfolio with a reduced size is SMV-L_1, which solves the mean-variance portfolio problem subject to an L_1-norm constraint,

$$\hat{\boldsymbol{w}}_{L_1} = \underset{\boldsymbol{w}\in\mathbb{R}^M}{\operatorname{argmax}} \ \boldsymbol{w}'\hat{\boldsymbol{\mu}}_M - \frac{\gamma}{2}\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}_M\boldsymbol{w} \quad \text{subject to} \quad ||\boldsymbol{w}||_1 = \sum_{i=1}^{M}|w_i| \leq \delta. \tag{28}$$

The L_1-norm performs asset selection and limits the amount of short-selling, which helps improve out-of-sample performance (DeMiguel et al., 2009a). We calibrate the threshold δ using cross-validation with out-of-sample utility as a decision criterion. Specifically, we search the optimal δ in an interval of 1,000 equally spaced values ranging from zero to $\delta_{\max}$, defined as the norm of the unconstrained solution to (28), i.e., $\delta_{\max} = ||\frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_M^{-1}\hat{\boldsymbol{\mu}}_M||_1$.

The second SMV portfolio with a reduced size is SMV-TW, which operates a trimming on the weights. The idea is that after having computed the SMV portfolio on all M assets, investors may want to select assets whose absolute portfolio weights are above a given threshold $\bar{w}$, in which case the value of N is determined implicitly by $\bar{w}$ as

$$N_{\bar{w}} = \sum_{i=1}^{M} \mathbb{1}_{\{|\hat{w}_i^\star| \geq \bar{w}\}} \quad \text{where} \quad \hat{\boldsymbol{w}}^\star = \frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_M^{-1}\hat{\boldsymbol{\mu}}_M. \tag{29}$$

Once the $N_{\bar{w}}$ assets are selected, we recompute the SMV portfolio on these $N_{\bar{w}}$ assets, which delivers better performance than keeping the original weights that are estimated under a high level of estimation risk. We select the threshold $\bar{w}$ via cross-validation with out-of-sample utility as a decision criterion. Specifically, we search the optimal $\bar{w}$ in an interval of 10 equally spaced values ranging from zero (all assets selected) to $\bar{w}_{\max}$ (no asset selected), where $\bar{w}_{\max} = \max|\frac{1}{\gamma}\hat{\boldsymbol{\Sigma}}_M^{-1}\hat{\boldsymbol{\mu}}_M|$ is the largest absolute weight of the unrestricted SMV portfolio.[19]

Finally, we consider three different estimates of the covariance matrix. First, the sample

selection rules in Section 5.3. We find that for all selection rules, the restricted SMV portfolio outperforms the unrestricted SMV on all M assets. However, we don't report these results in the main text because, even with a reduced size, the SMV portfolio largely underperforms the 2F, 3FGMV, and 3FEW portfolios.

[19]Dividing the interval in more than 10 values leads to similar results because in most cases, given that the SMV portfolio weights are subject to a high level of estimation risk, the cross-validation selects $\bar{w}$ close to $\bar{w}_{\max}$ such that we keep a very small number of assets.

estimator in (6) for consistency with our theory. Second, the linear shrinkage estimator of Ledoit and Wolf (2004). Third, the nonlinear shrinkage estimator of Ledoit and Wolf (2020), for which we report the results in Section D.4 of the supplementary material for conciseness.

5.3 Asset selection rules

Given a portfolio size N for the 2F, 3FGMV, and 3FEW portfolios, we must determine a *selection rule* to decide which N assets to select among all M assets. Our purpose is not to find which selection rule is most effective. Instead, our main objective is to demonstrate the added value of optimally reducing the portfolio size, and that even simple but sensible selection rules can deliver substantial gains over choosing all M assets.

For comparison purposes, we will also apply some of the proposed selection rules to the EW portfolio to evaluate their intrinsic value without them being affected by estimation errors in optimized portfolio weights as it is the case for the 2F, 3FGMV, and 3FEW portfolios. In that case, we will consider a subset of assets of size $N = T/2$, where we use a sample size $T = 120$ months as explained in Section 5.4, because the optimal N is generally close to $T/2$ for typical levels of equity correlations as explained in our theory and our simulations.

The asset selection rules we consider are listed in Table 3. We propose 11 rules that we select according to two criteria. First, they are *simple* to understand and implement. Second, they are *sensible*, i.e., they all have an economic intuition. The first selection rule consists in choosing all assets and setting $N = M$. The 10 remaining selection rules select a subset of $N \leq M$ assets. Among those, the first six are based solely on marginal information about the assets, whereas the last four rules also consider information about their dependence. Moreover, the first nine rules select the assets before computing the portfolio weights (ex-ante approach), whereas the last rule selects the subset of assets after computing the weights on all assets (ex-post approach). We now describe these 10 asset selection rules.[20]

[20]Our three-stage approach, as described in Section 1, is flexible in that it allows any type of selection

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

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

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

Maximum Sharpe ratio (MaxSR). We select the N assets with the maximum in-sample Sharpe ratios over the estimation window, because the EU of the SMV portfolio in (18) and of the two-fund rule in (16) both depend on the assets' Sharpe ratios. This selection rule is expected to work particularly well when there is momentum in the assets.[21] However, as a downside, it ranks the assets according to average returns that are notoriously noisy (Merton, 1980). We can anticipate two other issues with this rule. First, the selected assets have the best Sharpe ratios, but estimation risk may lead us to assign a wrong sign to portfolio weights and mistakenly short some good assets. Second, we do not include the assets with the worst Sharpe ratios although shorting them might deliver performance gains.

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

rule, and we propose a specific set as an illustration. One may also consider more sophisticated rules based on asset clustering, time-series models, or machine learning.

[21] We use datasets of characteristic and industry portfolios, and portfolios exhibit more momentum (i.e., positive return autocorrelation) than individual stocks (Campbell et al., 1997, p.74).

to the assets with the best Sharpe ratios to form long positions, and ending up with mistaken long positions on assets that have bad Sharpe ratios as a result of estimation noise.

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

The MaxSR, MinSR, and BWSR selection rules rank assets based on their in-sample Sharpe ratios.[22] Given that average returns are notoriously noisy, these rankings are bound to be unstable.[23] 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 rules rank assets based on the in-sample variance instead of the Sharpe ratio. The resulting rankings of assets are expected to be more stable over time than those under the Sharpe ratio because the variance is quite persistent over time. The MaxVar rule selects assets with the maximum variances, which are bound to also have larger expected returns if a positive risk-return tradeoff stands in the data. However, the low-risk anomaly implies that assets and portfolios with lower variances often have larger expected returns (Ang et al., 2006; Moreira and Muir, 2017), an anomaly that the MinVar rule can exploit.

The selection rules considered up to now exploit marginal information about the assets. We now propose four other selection rules that account for the dependence between assets.

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

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

Maximum and minimum correlation with the first principal component (MaxPC, MinPC). We select those assets whose returns have the maximum or the minimum correlations with the first principal component (PC). On the one hand, those assets with minimum correlations with the first PC are those least explained by the remaining assets, and thus, that offer higher potential for diversification gains. On the other hand, those assets with maximum correlations may lead to less extreme weights than those optimized on more dissimilar assets.

Best portfolio Sharpe ratio (Best θ_N^2). This selection rule follows Ao et al. (2019) and consists in choosing the assets so that they deliver a high maximum squared Sharpe ratio, θ_N^2 in (2). We implement this rule in three steps. First, we form 1,000 random selections of N assets, where N is our optimal portfolio size. Second, for each selection, we estimate θ_N^2 as detailed in Section C of the supplementary material. Third, we pick the selection which corresponds to the 95% quantile of all estimated values of θ_N^2, to avoid outliers.

The selection rules above are *ex-ante* rules because the selection of the N assets is done prior to and independently of the portfolio strategy. However, the in-sample portfolio weights on all M assets can be a valuable information for deciding which N assets to select. Therefore, the last selection rule we consider in Table 3 is an *ex-post* selection rule in which we select the assets based on the portfolio strategy that will then be applied to the selected assets.

Maximum absolute weights (MaxW). This selection rule works in three steps. First, compute $\hat{\boldsymbol{w}} \in \mathbb{R}^M$, the portfolio weights on all M assets using a given strategy (2F, 3FGMV, or 3FEW). Second, select the N assets with the maximum absolute weights $|\hat{w}_i|$. Third, recompute the portfolio on the N selected assets using the same strategy as before to obtain $\hat{\boldsymbol{w}} \in \mathbb{R}^N$.[24] MaxW selects the assets whose weights are large, and thus, are identified as relevant by the portfolio. We can anticipate the issue for MaxW that the selection of assets inherits large estimation errors because it is based on portfolio weights optimized under a

[24]We find that dropping this third step and keeping the initial portfolio weights computed on the M assets, with or without rescaling, delivers a poor performance because these weights are computed under a high level of estimation risk, i.e., a high value of M/T.

high level of estimation risk, i.e., a high value of M/T.

In Section D.8, we study how different is the selection of assets with these different rules. Specifically, we introduce a measure that determines how much more or less overlap there is between two asset selection rules relative to a pair of independent random selection rules. We find that for essentially all pairs of selection rules we consider, there is either around the same or less overlap than the random selection case. That is, our selection rules do select dissimilar assets, which allows us to test our portfolio strategies on different investment universes.

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.[25] We repeat this process iteratively, resulting in $\tau - T$ out-of-sample gross returns $r_{gross,k,t}$, where τ 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

$$r_{net,k,t} = (1 + r_{gross,k,t})\,(1 - \kappa \times \text{turnover}_{k,t-1}) - 1 \quad \text{if} \quad t = T+2, \ldots, \tau, \tag{30}$$

where κ is the proportional transaction cost parameter and

$$\text{turnover}_{k,t} = \sum_{i=1}^{N} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, \tau, \tag{31}$$

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 average bid-ask spreads reported by Engle et al. (2012) and Frazzini et al. (2018). Finally, we compare the portfolio

[25] Given that our theoretical results require $T > M + 4$, and that the largest $M = 108$, we need $T > 112$. If we choose a sample size that is too large, such as $T = 240$, then the optimal N for the 2F, 3FGMV, and 3FEW portfolios, which is close to $T/2$ as shown in Section 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.

strategies in terms of *annualized out-of-sample utility net of transaction costs*,

$$U_k = 12 \times \left(\hat{\mu}_k - \frac{\gamma}{2} \hat{\sigma}_k^2 \right), \tag{32}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the sample mean and variance of $r_{net,k,t}$. We set a risk-aversion coefficient of $\gamma = 1$; the intuition behind our results remains very similar for other values of γ.[26] In Section D.6 of the supplementary material, we also consider the out-of-sample Sharpe ratio.

For the 2F, 3FGMV, and 3FEW portfolios, the selected assets change as we re-estimate the optimal N and apply an asset selection rule. Because our methodology disentangles the determination of the optimal N from the choice of the N assets, we can also disentangle the rebalancing of portfolio weights from changes in the N assets. Therefore, to make our method less costly and more practically implementable, we rebalance the portfolio weights every month but change the optimal N and select assets once a year. In Section D.7 of the supplementary material, we also consider changing the selected assets every one, six and 24 months, and analyze the impact on performance. We find that setting a lower frequency than monthly substantially reduces turnover and improves the net-of-cost performance. Moreover, frequencies of 6, 12, or 24 months yield similar results overall.

Finally, we compute two-sided p-values for the statistical test of the difference between the net utility of the 2F, 3FGMV, and 3FEW portfolios computed on all assets ('All' selection rule) versus under each of the 10 other selection rules in Table 3. The p-values for SMV-TW and SMV-L_1 are computed by taking SMV as a benchmark. To compute the p-values, we generate 10,000 bootstrap samples using the stationary block bootstrap approach of Politis and Romano (1994) with an average block size of five, and then use the methodology of Ledoit and Wolf (2008, Remark 3.2) to produce the resulting p-values. We use the symbols ○, ◑, and ● to indicate that the p-value is less than 10%, 5%, and 1%, respectively.

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

5.5 Discussion of results

In Table 4, we report the annualized net out-of-sample utility, in percentage points, of the 10 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.[27]

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 the theory for finding optimal portfolio strategies considers the sample covariance matrix for tractability, it is beneficial for investors to use the shrinkage covariance matrix. Overall, the ranking of portfolio strategies and selection rules is similar for both covariance matrix estimators, hence the discussion of results below essentially applies to either case.

First, we address our main objective which is to assess the value of optimally reducing the portfolio size. For all three portfolio strategies, i.e., 2F, 3FGMV, and 3FEW, we find that optimally reducing the portfolio size delivers substantial performance gains. Specifically, the asset selection rule 'All', which sets $N = M$, is nearly systematically outperformed by the remaining selection rules that select a subset of assets whose size is determined using our theory.[28] The differences are often statistically significant, particularly when using a shrinkage covariance matrix. Moreover, the differences are systematically positive for the MinSR, BWSR, MaxVar, BWVar, MinPC, and MaxPC asset selection rules, and are often substantial. For instance, under the shrinkage covariance matrix, and on average across datasets, the 2F strategy applied to all M assets delivers an annualized net utility of 11.30 percentage points, whereas it delivers 58.17 percentage points with the BWSR selection rule. We depict this main finding in Figure 5 which shows, across the six datasets and for the

[27] The MaxW asset selection rule is not applicable to the EW portfolio since all weights are equal.

[28] Out of the 180 configurations (10 selection rules $\times$ 6 datasets $\times$ 3 portfolio strategies), the restricted portfolios outperform the 'All' selection rule in 153, 161, and 165 cases under the sample, linear shrinkage, and nonlinear shrinkage covariance matrix, respectively. If we restrict to cases in which the p-value for the performance gain is below 10%, these numbers reduce to 93, 147, and 152, respectively.

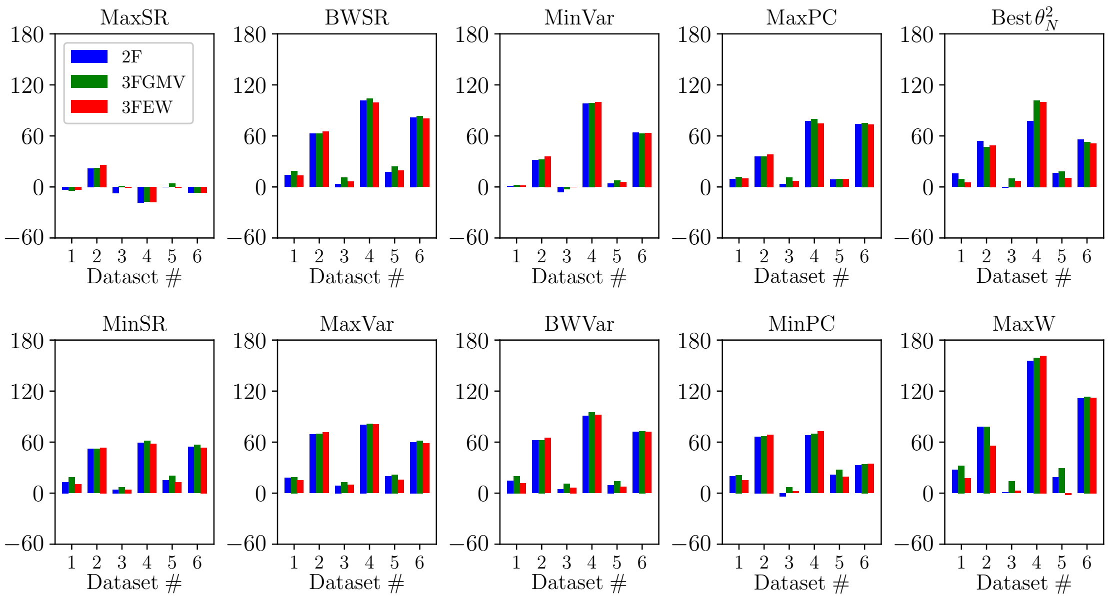
Table 4: Annualized net out-of-sample utility in empirical data (in percentage points)

		Sample covariance matrix						Linear shrinkage covariance matrix					
		Dataset						Dataset					
Port-folio strat.	Asset select. rule	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV
2F	All	−0.31	−59.6	−15.3	99.4	−18.8	56.8	5.11	1.20	1.09	41.6	1.09	17.7
2F	MaxSR	1.38	11.6●	−7.48	23.6○	−10.5	−27.1◑	1.93○	23.1●	−6.10	23.1●	0.69	10.8○
2F	MinSR	25.4●	73.2●	7.10●	146	24.9●	139◑	18.0●	53.2●	5.21○	101●	16.0●	72.1●
2F	BWSR	7.39	84.4●	−4.42	133	11.7○	242●	19.2◑	63.8●	4.74	143●	18.8◑	99.5●
2F	MaxVar	25.0◑	124●	17.9●	114	21.7●	83.8	23.2●	70.5●	9.99◑	122●	21.1●	77.7●
2F	MinVar	2.47	12.5●	−13.1	219◑	−1.88	225●	6.08	33.1●	−5.25○	140●	5.40○	81.4●
2F	BWVar	13.0	89.3●	−6.76	182○	0.14	180●	19.8●	63.4●	5.39	133●	10.6◑	89.9●
2F	MaxPC	8.80	18.3●	3.67◑	169○	9.62●	237●	14.7●	36.9●	4.67	119●	9.81●	91.7●
2F	MinPC	27.4●	112●	−6.23	115	17.9◑	94.1	24.9●	67.7●	−2.55	109●	22.8●	50.5●
2F	Best θ_N^2	3.77	14.7●	−8.77	182○	−7.87	137◑	20.7●	55.1●	0.49	119●	17.6●	73.2●
2F	MaxW	−127●	−305●	−179●	−581●	−129●	−366●	32.3●	79.0●	2.19	197●	19.8○	129●
3FGMV	All	0.47	−60.5	−12.5	99.1	−19.7	55.4	5.77	1.21	1.79	41.9	1.31	17.9
3FGMV	MaxSR	−4.06	10.8●	−1.93○	15.6◑	−6.02	−27.4◑	1.63	23.3●	3.04	24.6●	5.32	10.8○
3FGMV	MinSR	36.0●	71.6●	12.9●	150	31.5●	141◑	24.3●	53.4●	8.94◑	103●	21.5●	74.8●
3FGMV	BWSR	10.9	82.4●	1.02	143	19.9◑	253●	24.7●	64.0●	13.0◑	146●	25.2●	101●
3FGMV	MaxVar	25.1◑	126●	24.9●	117	21.7●	85.0	24.6●	70.9●	14.3●	123●	23.1●	79.6●
3FGMV	MinVar	6.13	13.3●	−14.9	222◑	0.35○	219●	7.94	33.2●	−0.72	141●	8.69◑	80.7●
3FGMV	BWVar	27.4◑	89.0●	2.74	183○	11.2◑	190●	25.8●	63.5●	12.9◑	137●	15.4●	90.8●
3FGMV	MaxPC	12.2	16.7●	10.8●	176○	7.50◑	225●	17.7●	36.9●	13.1●	122●	11.0●	93.1●
3FGMV	MinPC	33.3●	113●	6.47◑	115	25.9●	91.3	27.0●	68.2●	8.74○	112●	28.7●	52.0●
3FGMV	Best θ_N^2	12.3	7.39●	−0.08	122	1.33○	165●	15.1○	48.2●	11.9●	144●	19.6●	71.0●
3FGMV	MaxW	−136●	−300●	−147●	−650●	−110●	−325●	38.2●	79.4●	15.7	201●	30.4●	131●
3FEW	All	1.83	−62.8	−10.8	102	−16.8	60.8	7.21	−2.56	3.01	40.3	2.38	18.5
3FEW	MaxSR	3.68	12.6●	−0.61	23.6◑	−12.1	−20.4◑	4.05	23.4●	2.23	22.4◑	1.79	12.0
3FEW	MinSR	24.8●	70.8●	11.3●	145	24.8●	146◑	17.6◑	51.0●	7.15	98.2●	15.3◑	72.0●
3FEW	BWSR	8.43	84.4●	3.41	135	17.9◑	237●	20.6●	62.6●	9.48○	140●	22.0●	98.8●
3FEW	MaxVar	26.0◑	122●	25.3●	124	16.0●	98.9	22.2●	68.7●	12.8●	121●	18.0●	77.0●
3FEW	MinVar	4.93	12.3●	−5.87	220◑	0.40	230●	8.94	33.5●	2.79	140●	8.08○	82.2●
3FEW	BWVar	14.2	88.6●	−3.30	195○	−0.15	187●	19.0●	62.4●	9.50○	132●	9.65○	90.4●
3FEW	MaxPC	9.42	18.1●	11.9●	162	5.87◑	247●	17.0●	35.7●	9.98◑	115●	11.5●	92.0●
3FEW	MinPC	26.0◑	114●	2.26	120	15.9◑	103	22.1●	66.3●	5.27	113●	21.4●	52.8●
3FEW	Best θ_N^2	12.5	4.58●	−4.45	136	−8.83	166●	12.5	46.2●	10.1◑	140●	13.2●	69.5●
3FEW	MaxW	−124●	−305●	−157●	−650●	−139●	−361●	24.8●	52.9●	5.80	202●	0.52	131●
EW	All	8.11	2.80	8.00	3.98	7.13	5.61	8.11	2.80	8.00	3.98	7.13	5.61
EW	MaxSR	8.77	4.72●	8.49	5.97◑	7.40	6.25◑	8.77	4.72●	8.49	5.97◑	7.40	6.25
EW	MinSR	7.39●	0.83●	7.55●	2.75	6.77●	5.18◑	7.39◑	0.83●	7.55	2.75●	6.77◑	5.18●
EW	BWSR	7.94	2.42●	7.72	2.86	7.35◑	5.28●	7.94●	2.42●	7.72○	2.86●	7.35●	5.28●
EW	MaxVar	7.95◑	1.38●	7.97●	3.52	6.68●	5.31	7.95●	1.38●	7.97●	3.52●	6.68●	5.31●
EW	MinVar	8.60	4.23●	8.18	4.36◑	7.44	5.86●	8.60	4.23●	8.18	4.36●	7.44○	5.86●
EW	BWVar	7.60	2.34●	7.52	3.99○	6.88	5.58●	7.60●	2.34●	7.52○	3.99●	6.88○	5.58●
EW	MaxPC	8.32	2.97●	8.83●	2.19	7.37◑	5.08●	8.32●	2.97●	8.83◑	2.19●	7.37●	5.08●
EW	MinPC	7.87◑	2.48●	7.20	5.00	6.97◑	6.22	7.87●	2.48●	7.20	5.00●	6.97●	6.22●
EW	Best θ_N^2	7.88	2.72●	7.92	3.82	7.02	6.03●	8.27	2.75●	7.82◑	3.95●	6.90●	5.92●
EWRF		1.42	−4.35	1.26	−3.06	0.80	−0.29	1.42	−4.35	1.26	−3.06	0.80	−0.29
SGMV		1.76	−49.7	4.25	−0.34	−2.28	−4.67	10.3	2.33	7.75	4.02	8.48	6.65
GMVRF		1.81	−26.3	−0.98	−11.4	−12.3	−5.34	4.01	−0.00	1.47	0.14	0.66	0.49
SMV		−4E+05	−3E+11	−1E+06	−6E+07	−8E+08	−8E+07	−1443	−107	−1399	−2360	−1457	−1143
SMV-L_1		15.8●	−108●	11.7●	−579●	−2.91●	0.88●	23.9●	67.8	5.13●	−195●	10.4●	16.4●
SMV-TW		−35.3●	−302●	−16.8●	−105●	−25.9●	−143●	−29.9●	−52.1	−17.8●	−376●	−33.6●	−349●

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 10 portfolio strategies in Table 2, the 11 asset selection rules in Table 3, across the six datasets in Table 1. The table is constructed following the out-of-sample methodology described in Section 5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under each selection rule with the 'All' selection rule as benchmark, using the block bootstrap methodology described in Section 5.4. We also report p-values for SMV-TW and SMV-L_1, using the SMV as benchmark. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

shrinkage covariance matrix, the difference between the annualized net out-of-sample utility of 2F, 3FGMV, and 3FEW implemented on a subset of $\hat{N}^{\star}$ assets versus all M assets. Having addressed our main objective, we now discuss other findings from Table 4.

Figure 5: Difference in net out-of-sample utility relative to investing in all assets



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

Second, we find that applying the MinSR, MaxVar, and MaxPC asset selection rules to the 2F, 3FGMV, and 3FEW strategies delivers consistently positive net out-of-sample utilities that are almost systematically larger than those of the seven other rules. Focusing on the shrinkage covariance matrix, the BWSR, BWVar, MinPC, and Best-θ^2_N selection rules also consistently deliver positive utilities that generally outperform the benchmarks. These results are quite remarkable because the EW portfolio, in particular, is difficult to outperform using mean-variance portfolios when estimation risk, i.e., M/T, is large, which is the case in all our datasets where M/T ranges from 0.78 to 0.9. These results are the consequence of our optimal portfolio size because the 2F, 3FGMV and 3FEW portfolios with the 'All' selection rule ($N = M$) underperform the EW portfolio for four out of six datasets.

Third, we turn to the asset selection rules based on the Sharpe ratio, i.e., MaxSR, MinSR, and BWSR. When these are applied to the EW portfolio, MaxSR consistently outperforms and delivers better performance than the EW portfolio of all M assets. In contrast, MinSR and BWSR underperform the EW portfolio of all assets. Specifically, on average across datasets, the EW portfolio of all assets delivers an annualized net out-of-sample utility of 5.94 percentage points, whereas the EW portfolio with the MaxSR, MinSR, and BWSR selection rules delivers average utilities of 6.93, 5.08, and 5.60, respectively. These results indicate return momentum as forming an EW portfolio of assets with maximum in-sample Sharpe ratios helps improve out-of-sample performance relative to selecting all assets.

However, when we turn to MV optimized portfolios, namely, 2F, 3FGMV and 3FEW, we reach an opposite conclusion. Specifically, MinSR consistently outperforms MaxSR as well as the portfolio implemented on all M assets, and MaxSR is one the worst-performing selection rules. These results mean that when we use optimized portfolio weights, the performance gains from asset selection rules do not originate from selecting assets with the best marginal performance, but from selecting assets that can be best exploited by the portfolio, e.g., with short positions for the MinSR rule. Finally, the BWSR selection rule, which capitalizes on long and short positions by selecting both good and bad assets, achieves its objective as it consistently outperforms MaxSR and MinSR when using a shrinkage covariance matrix.[29]

Fourth, we turn to the asset selection rules based on the variance, i.e., MaxVar, MinVar, and BWVar. When these are applied to the EW portfolio, MinVar consistently outperforms and also delivers better performance than the EW portfolio of all M assets. Specifically, on average across datasets, the EW portfolio of all assets delivers an annualized net out-of-

[29]However, BWSR is often outperformed by MinSR under the sample covariance matrix because, in that case, the long-short positions are too extreme, and thus, do not work well out of sample and increase turnover.

sample utility of 5.94 percentage points, whereas the EW portfolio with the MaxVar, MinVar, and BWVar selection rules delivers average utilities of 5.47, 6.44, and 5.65, respectively. This finding indicates a low-risk anomaly in our datasets. However, as noted above, this does not necessarily mean that MinVar is also the best selection rule for the 2F, 3FGMV and 3FEW strategies. And indeed, Figure 5 shows that for these portfolios, MinVar delivers a worse performance overall than MaxVar and BWVar. This can be explained because the assets with the maximum variances, which are selected by MaxVar and BWVar, often have low or negative average returns as the low-risk anomaly predicts, and the 2F, 3FGMV and 3FEW portfolios are able to profit from this using small or short positions.[30]

Fifth, the MaxPC and MinPC asset selection rules, which account for the dependence between the assets by selecting those that are most or least correlated with the first PC, offer a consistently good performance. In particular, the resulting 2F, 3FGMV, and 3FEW portfolios have positive net out-of-sample utilities in all cases. However, there is no clear trend regarding which of the two selection rules outperforms the other.

Sixth, the Best θ_N^2 selection rule, which attempts to maximize the portfolio Sharpe ratio, outperforms investing in all assets for the 2F, 3FGMV and 3FEW portfolios in all cases but one. Interestingly, although they are quite similar in spirit, the Best θ_N^2 selection rule systematically and significantly outperforms the MaxSR rule under the shrinkage covariance matrix, indicating that considering information about the dependence of assets is crucial in selecting assets that deliver a large portfolio Sharpe ratio.

Seventh, the MaxW selection rule outperforms on average all other rules under the shrinkage covariance matrix. For instance, considering the 3FGMV portfolio, MaxW delivers an annualized net utility of 82.69 percentage points on average across datasets, whereas the

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

second-best selection rule, BWSR, delivers an average utility of 62.32 percentage points. However, MaxW underperforms all other selection rules under the sample covariance matrix. This can be explained because under the sample estimator, the 2F, 3FGMV and 3FEW portfolios are more subject to error maximization (Michaud, 1989) than under the shrinkage estimator. Therefore, selecting the assets with large absolute weights, i.e., identified as most relevant, is more desirable under the shrinkage estimator than under the sample estimator.[31]

Finally, we investigate our three ways of implementing the SMV portfolio: SMV, SMV-L_1, and SMV-TW. First, the plain SMV delivers by far the worst performance because we consider settings with high M/T. Second, SMV-L_1, which selects assets by maximizing the MV utility subject to an L_1-norm constraint, greatly improves out-of-sample performance relative to SMV and SMV-TW. However, in most cases, SMV-L_1 underperforms the optimally restricted 2F, 3FGMV, and 3FEW portfolios. Third, SMV-TW, which keeps only the assets whose SMV portfolio weights are above a threshold selected via cross-validation, outperforms the plain SMV but still delivers negative net utilities. In Section D.5 of the supplementary material, we compare the portfolio size obtained with SMV-L_1 and SMV-TW to that obtained with our theory, $\hat{N}^{\star}_{smv}$. We find that the three approaches set a small N on average, and that our approach displays much less variability across windows.

6 Conclusion

We offer a novel perspective on the optimal *portfolio size* and challenge the conventional wisdom that investors should invest in as many assets as possible to diversify away idiosyncratic risk. Specifically, because of *parameter uncertainty*, we show that it is optimal to invest

[31] To confirm this, we implement in unreported results an opposite selection rule denoted MinW which selects the assets with the lowest absolute weights instead of the highest absolute weights. The idea behind MinW is to exclude assets with extreme portfolio weights because these are most subject to error maximization. As expected, we find that MinW systematically outperforms MaxW by a substantial margin when using the sample covariance matrix, but the opposite happens under the shrinkage covariance matrix.

in a limited number N of assets. For several mean-variance portfolio combination rules, we derive the optimal value $N^\star$ of N maximizing the expected *out-of-sample* performance. This optimal N strikes a tradeoff between accessing additional investment opportunities and limiting the number of parameters and optimal weights to estimate. For typical levels of equity return correlations, we find that the optimal N is slightly below *half the sample size*, which can be quite small. Our approach is flexible in the sense that it disentangles the computation of the optimal N from the choice of which N assets to include in the portfolio.

We empirically test our theory in various environments and analyze the results. To this end, we propose an estimator $\hat{N}^\star$ of $N^\star$ and suggest a set of simple, sensible and intrinsically different selection rules to determine which $\hat{N}^\star$ assets to select from the initial investment universe. We show that our optimally restricted portfolios outperform their counterparts that invest in all available assets. This result prevails in nearly all cases, across different portfolio strategies, datasets, and asset selection rules, and holds with transaction costs. For some asset selection rules, our optimally restricted portfolios also systematically outperform the robust EW and minimum-variance portfolios, which are hard to beat in high dimension.

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

References

Ang, A., Hodrick, R., Xing, Y., and Zhang, X. (2006), “The cross-section of volatility and expected returns.” *The Journal of Finance*, 61(1):259–299.

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

Barroso, P. and Saxena, K. (2021), “Lest we forget: Using out-of-sample forecast errors in portfolio optimization.” *The Review of Financial Studies*, 35(3):1222–1278.

Boyle, P., Garlappi, L., Uppal, R., and Wang, T. (2012), "Keynes meets Markowitz: The trade-off between familiarity and diversification." *Management Science*, 58(2):253–272.

Calvet, L., Campbell, J., and Sodini, P. (2007), "Down or out: Assessing the welfare costs of household investment mistakes." *Journal of Political Economy*, 115(5):707–747.

Campbell, J. (2006), "Household Finance." *Journal of Finance*, 61(4):1553–1604.

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

Chamberlain, G. (1983), "A characterization of the distributions that imply mean—variance utility functions." *Journal of Economic Theory*, 29(1):185–201.

Chung, M., Lee, Y., Kim, J. H., Kim, W. C., and Fabozzi, F. (2022), "The effects of errors in means, variances, and correlations on the mean-variance framework." *Quantitative Finance*, 22(10):1893–1903.

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

DeMiguel, V., Garlappi, L., and Uppal, R. (2009b), "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *The Review of Financial Studies*, 22(5):1915–1953.

DeMiguel, V., Martín-Utrera, A., and Nogales, F. (2015), "Parameter uncertainty in multiperiod portfolio optimization with transaction costs." *Journal of Financial and Quantitative Analysis*, 50(6):1443–1471.

Du, J.-H., Guo, Y., and Wang, X. (2023), "High-dimensional portfolio selection with cardinality constraints." *Journal of the American Statistical Association*, 118(542):779–791.

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

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

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

Fan, J., Zhang, J., and Yu, K. (2012), "Vast portfolio selection with gross-exposure constraints." *Journal of the American Statistical Association*, 107(498):592–606.

Frazzini, A., Israel, R., and Moskowitz, T. J. (2018), "Trading Costs." *SSRN Working Paper*.

Gao, J. and Li, D. (2013), "Optimal cardinality constrained portfolio selection." *Operations Research*, 61(3):745–761.

Goetzmann, W. and Kumar, A. (2008), "Equity portfolio diversification." *Review of Finance*, 12(3):433–446.

Kan, R. and Lassance, N. (2023), "Optimal portfolio choice with fat tails and parameter uncertainty." *SSRN working paper*.

Kan, R. and Wang, X. (2023), "Optimal portfolio choice with unknown benchmark efficiency." *Management Science*, forthcoming.

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

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

Koijen, R. and Yogo, M. (2019), "A demand system approach to asset pricing." *Journal of Political Economy*, 127(4):1475–1515.

Lassance, N. and Martín-Utrera, A. (2023), "Sentiment-based portfolios." *SSRN working paper*.

Lassance, N., Martín-Utrera, A., and Simaan, M. (2024a), "The risk of expected utility under parameter uncertainty." *Management Science*, forthcoming.

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

Ledoit, O. and Wolf, M. (2003), "Honey, I shrunk the sample covariance matrix." *The Journal of Portfolio Management*, 30(4):110–119.

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

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

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

Merton, R. (1980), "On estimating the expected return on the market: An exploratory investigation." *Journal of Financial Economics*, 8(4):323–361.

Michaud, R. (1989), "The Markowitz optimization enigma: Is 'optimized' optimal?" *Financial Analysts Journal*, 45(1):31–42.

Moreira, A. and Muir, T. (2017), "Volatility-managed portfolios." *The Journal of Finance*, 72(4):1611–1643.

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

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

Schuhmacher, F., Kohrs, H., and Auer, B. (2021), "Justifying mean-variance portfolio selection when asset returns are skewed." *Management Science*, 67(12):7812–7824.

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

Yen, Y.-M. (2016), "Sparse weighted-norm minimum variance portfolios." *Review of Finance*, 20(3):1259–1287.

Yuan, M. and Zhou, G. (2024), "Why naïve 1/N diversification is not so naïve, and how to beat it?" *Journal of Financial and Quantitative Analysis*, forthcoming.

Supplementary Material to

Optimal Portfolio Size with Parameter Uncertainty

Nathan Lassance Rodolphe Vanderveken Frédéric Vrins

This supplementary material to the paper contains five sections. In Section A, we present the theory we use to determine the optimal portfolio size for the three-fund rules. In Section B, we study the impact of the sequence of assets on the EU and the optimal portfolio size. In Section C, we explain how we estimate the different parameters on which the combination coefficients and the optimal values of N depend. In Section D, we report and discuss additional empirical results. In Section E, we provide the proofs of all theoretical results.

A Optimal portfolio size for three-fund rules

In Section 3, we study the optimal portfolio size N for the SMV portfolio and two-fund rule. We now study the optimal N for the *three-fund rules* that add another portfolio rule to further improve the EU. Specifically, we consider in Section A.1 a three-fund rule based on the GMV portfolio as in, e.g., Kan and Zhou (2007), DeMiguel et al. (2015), and Yuan and Zhou (2024), which we call the *GMV-three-fund rule*. In Section A.2, we consider a three-fund rule based on the EW portfolio as in, e.g., Tu and Zhou (2011), Kan and Wang (2023), and Lassance et al. (2024b), which we call the *EW-three-fund rule*.

A.1 Three-fund rule with the global minimum-variance portfolio

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

$$\boldsymbol{w}_g = \frac{\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}. \tag{SM1}$$

We introduce the following parameters:

$$\mu_{g,N} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N},\ \sigma_{g,N}^2 = \frac{1}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N},\ \lambda_{g,N} = \frac{\mu_{g,N}}{\sigma_{g,N}^2},\ \theta_{g,N}^2 = \frac{\mu_{g,N}^2}{\sigma_{g,N}^2},\ \psi_{g,N}^2 = \theta_N^2 - \theta_{g,N}^2, \tag{SM2}$$

which stand for the mean return, variance, price of risk, squared Sharpe ratio, and inefficiency of the GMV portfolio $\boldsymbol{w}_g$ computed on N assets.

Under parameter uncertainty, we define the three-fund rule that invests in the SMV portfolio, the sample GMV (SGMV) portfolio, and the risk-free asset as

$$\hat{\boldsymbol{w}}(\boldsymbol{\alpha}) = \frac{\alpha_1}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N + \frac{\alpha_2}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\mathbf{1}_N, \tag{SM3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{R}^2$ is the vector of combination coefficients. We call $\hat{\boldsymbol{w}}(\boldsymbol{\alpha})$ the *GMV-three-fund rule*. In the next proposition, we exploit Proposition 2 and Kan and Lassance (2023, Proposition 8) to derive the EU of the GMV-three-fund rule in (SM3) when asset returns are elliptically distributed, as well as the resulting optimal combination coefficients $\boldsymbol{\alpha}^\star$ and which EU they deliver.

Proposition SM.1 *Let $T > N+4$ and Assumption 2 hold. Then, the expected out-of-sample utility of the GMV-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\alpha})$ in* (SM3) *is*

$$\begin{aligned} EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha})) = \frac{1}{2\gamma}\frac{T}{T-N-2}\Bigg[&2\kappa_{N,1}\left(\alpha_1\theta_N^2 + \alpha_2\lambda_{g,N}\right) - \frac{c_N T}{T-N-2} \\ &\times\left(\alpha_1^2\left(\kappa_{N,2}\theta_N^2 + \kappa_{N,3}\frac{N}{T}\right) + \frac{\alpha_2^2\kappa_{N,2}}{\sigma_{g,N}^2} + 2\alpha_1\alpha_2\kappa_{N,2}\lambda_{g,N}\right)\Bigg]. \end{aligned} \tag{SM4}$$

Moreover, the optimal combination coefficients $\boldsymbol{\alpha}^\star = (\alpha_1^\star, \alpha_2^\star)$ maximizing (SM4) *are*

$$(\alpha_1^\star, \alpha_2^\star) = \left(\frac{T-N-2}{c_N T}\frac{\kappa_{N,1}\psi_{g,N}^2}{\kappa_{N,2}\psi_{g,N}^2 + \kappa_{N,3}\frac{N}{T}}, \frac{T-N-2}{c_N T}\frac{\kappa_{N,3}}{\kappa_{N,2}}\frac{\kappa_{N,1}\frac{N}{T}\mu_{g,N}}{\kappa_{N,2}\psi_{g,N}^2 + \kappa_{N,3}\frac{N}{T}}\right), \tag{SM5}$$

and the resulting expected out-of-sample utility is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha}^\star)) = \frac{\kappa_{N,1}^2}{2\gamma c_N}\frac{\theta_N^2\psi_{g,N}^2 + \frac{\kappa_{N,3}}{\kappa_{N,2}}\frac{N}{T}\theta_{g,N}^2}{\kappa_{N,2}\psi_{g,N}^2 + \kappa_{N,3}\frac{N}{T}}. \tag{SM6}$$

Proposition SM.1 shows that for an investor holding the optimal GMV-three-fund rule, the optimal portfolio size N is found by maximizing the EU in (SM6). We proceed by expressing θ_N^2 and $\theta_{g,N}^2$ as explicit functions of N by assuming that the covariance matrix $\boldsymbol{\Sigma}_N$

complies with Assumption 1. This is done in Proposition 1 for θ_N^2, and we now derive the expression for $\theta_{g,N}^2$.

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

$$\theta_{g,N}^2 = \frac{N}{1-\rho} r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N), \quad r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N) = \frac{\left(\bar{\lambda}_N - \frac{N\rho}{1-\rho+N\rho}\bar{\theta}_{N,1}\bar{\sigma}_{N,-1}\right)^2}{\bar{\sigma}_{N,-2} - \frac{N\rho}{1-\rho+N\rho}\bar{\sigma}_{N,-1}^2}, \tag{SM7}$$

where $\bar{\theta}_{N,1}$ *is defined in* (5), $\bar{\boldsymbol{\sigma}}_N = (\bar{\sigma}_{N,-1}, \bar{\sigma}_{N,-2})$, *and*

$$\bar{\lambda}_N = \frac{1}{N}\sum_{i=1}^{N} \lambda_i, \quad \bar{\sigma}_{N,k} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^k, \tag{SM8}$$

with $\lambda_i = \mu_i/\sigma_i^2$ *the price of risk of asset* i.

Using Propositions 1 and SM.2, the EU of the GMV-three-fund rule in (SM6) becomes, under Assumption 1,

$$\begin{aligned} & EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha}^\star)) \\ &= \frac{N\kappa_{N,1}^2}{2\gamma c_N(1-\rho)} \frac{\delta_N(\bar{\boldsymbol{\theta}}_N)\left(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)\right) + \frac{\kappa_{N,3}}{\kappa_{N,2}}\frac{1-\rho}{T} r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)}{\kappa_{N,2}\left(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{g,N}(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)\right) + \kappa_{N,3}\frac{1-\rho}{T}} && \text{(SM9)} \\ &=: EU_{3f,g}(N, T, \rho, \tau, \bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N). && \text{(SM10)} \end{aligned}$$

Then, as in Section 3, we replace $(\bar{\theta}_{N,1}, \bar{\lambda}_N, \bar{\boldsymbol{\sigma}}_N)$ by $(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$ and find the optimal N as

$$N_{3f,g}^\star = \underset{N\in\{1,\ldots,\min(M,T-5)\}}{\text{argmax}} EU_{3f,g}(N, T, \rho, \tau, \bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M). \tag{SM11}$$

Note that, similar to the two-fund rule, the objective function $EU_{3f,g}$ is proportional to N/c_N, which as shown in (23)–(24) is maximized by N slightly below $T/2$.[SM1] The simulations

[SM1] That the EU of both the optimal two-fund rule and the GMV-three-fund rule is proportional to N/c_N can be explained as follows: the proportionality to $1/c_N$ is because both the SMV portfolio and the SGMV portfolio are proportional to $\hat{\boldsymbol{\Sigma}}_N^{-1}$, and the proportionality to N is because both the squared Sharpe ratio of the MV portfolio (θ_N^2) and the GMV portfolio ($\theta_{g,N}^2$) are proportional to N under Assumption 1.

in Section 4 show that $N^{\star}_{3f,g}$ is similar to $N^{\star}_{2f}$ and is slightly below $T/2$ when ρ is not too small, as it is the case with equity data.

A.2 Three-fund rule with the equally weighted portfolio

We now consider the EW portfolio as an additional portfolio rule to invest in,

$$\boldsymbol{w}_{ew} = \frac{\mathbf{1}_N}{N}, \tag{SM12}$$

which benefits from no parameter uncertainty. We introduce the following parameters similar to the case with the GMV-three-fund rule:

$$\mu_{ew,N} = \boldsymbol{w}'_{ew}\boldsymbol{\mu}_N,\ \sigma^2_{ew,N} = \boldsymbol{w}'_{ew}\boldsymbol{\Sigma}_N\boldsymbol{w}_{ew},\ \lambda_{ew,N} = \frac{\mu_{ew,N}}{\sigma^2_{ew,N}},\ \theta^2_{ew,N} = \frac{\mu^2_{ew,N}}{\sigma^2_{ew,N}},\ \psi^2_{ew,N} = \theta^2_N - \theta^2_{ew,N}. \tag{SM13}$$

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

$$\hat{\boldsymbol{w}}(\boldsymbol{\beta}) = \frac{\beta_1}{\gamma}\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\boldsymbol{\mu}}_N + \frac{\beta_2}{\gamma}\boldsymbol{w}_{ew}, \tag{SM14}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2) \in \mathbb{R}^2$ is the vector of combination coefficients. We call $\hat{\boldsymbol{w}}(\boldsymbol{\beta})$ the *EW-three-fund rule*. In the next proposition, we exploit Proposition 2 to derive the EU of the EW-three-fund rule (SM3) when asset returns are elliptically distributed, as well as the resulting optimal combination coefficients $\boldsymbol{\alpha}^{\star}$ and which EU they deliver.

Proposition SM.3 *Let $T > N+4$ and Assumption 2 hold. Then, the expected out-of-sample utility of the EW-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\beta})$ in* (SM14) *is*

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\beta})) = \frac{1}{2\gamma} \times \left[\frac{2\beta_1\kappa_{N,1}\theta^2_N T}{T-N-2} + 2\beta_2\mu_{ew,N}\right.$$

$$
- \frac{\beta_1^2 c_N T^2}{(T-N-2)^2} \left(\kappa_{N,2} \theta_N^2 + \kappa_{N,3} \frac{N}{T} \right) - \beta_2^2 \sigma_{ew,N}^2 - \frac{2\beta_1 \beta_2 \kappa_{N,1} \mu_{ew,N} T}{T-N-2} \Bigg]. \tag{SM15}
$$

Moreover, the optimal combination coefficients $\boldsymbol{\beta}^\star = (\beta_1^\star, \beta_2^\star)$ *maximizing* (SM15) *are*

$$
(\beta_1^\star, \beta_2^\star) = \left(\frac{T-N-2}{T} \frac{\kappa_{N,1} \psi_{ew,N}^2}{\kappa_{N,1}^2 \psi_{ew,N}^2 + d_N}, \frac{d_N \lambda_{ew,N}}{\kappa_{N,1}^2 \psi_{ew,N}^2 + d_N} \right), \tag{SM16}
$$

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*

$$
EU(\hat{\boldsymbol{w}}(\boldsymbol{\beta}^\star)) = \frac{1}{2\gamma} \frac{\kappa_{N,1}^2 \theta_N^2 \psi_{ew,N}^2 + d_N \theta_{ew,N}^2}{\kappa_{N,1}^2 \psi_{ew,N}^2 + d_N}. \tag{SM17}
$$

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

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

$$
\theta_{ew,N}^2 = \frac{N}{1-\rho} r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}), \quad r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) = \frac{(1-\rho)\bar{\mu}_N^2}{(1-\rho)\bar{\sigma}_{N,2} + \rho N \bar{\sigma}_{N,1}^2}, \tag{SM18}
$$

where $\bar{\sigma}_{N,1}$ *and* $\bar{\sigma}_{N,2}$ *are defined in* (SM8) *and*

$$
\bar{\mu}_N = \frac{1}{N} \sum_{i=1}^N \mu_i. \tag{SM19}
$$

Using Propositions 1 and SM.4, the EU of the EW-three-fund rule in (SM17) becomes, under Assumption 1,

$$
EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha}^\star)) = \frac{N}{2\gamma(1-\rho)} \times
$$

$$
\frac{\kappa_{N,1}^2 \delta_N(\bar{\boldsymbol{\theta}}_N) \Big(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) \Big) + \Big(\frac{(1-\rho) c_N \kappa_{N,3}}{T} + (c_N \kappa_{N,2} - \kappa_{N,1}^2) \delta_N(\bar{\boldsymbol{\theta}}_N) \Big) r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2})}{\kappa_{N,1}^2 \Big(\delta_N(\bar{\boldsymbol{\theta}}_N) - r_{ew,N}(\bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}) \Big) + \Big(\frac{(1-\rho) c_N \kappa_{N,3}}{T} + (c_N \kappa_{N,2} - \kappa_{N,1}^2) \delta_N(\bar{\boldsymbol{\theta}}_N) \Big)} \tag{SM20}
$$

$$=: EU_{3f,ew}(N, T, \rho, \tau, \bar{\mu}_N, \bar{\sigma}_{N,1}, \bar{\sigma}_{N,2}). \quad \text{(SM21)}$$

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

$$N^{\star}_{3f,ew} = \underset{N \in \{1,\dots,\min(M,T-5)\}}{\operatorname{argmax}} EU_{3f,ew}(N, T, \rho, \tau, \bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2}). \quad \text{(SM22)}$$

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 $N^{\star}_{3f,ew}$ typically hovers slightly below $T/2$.

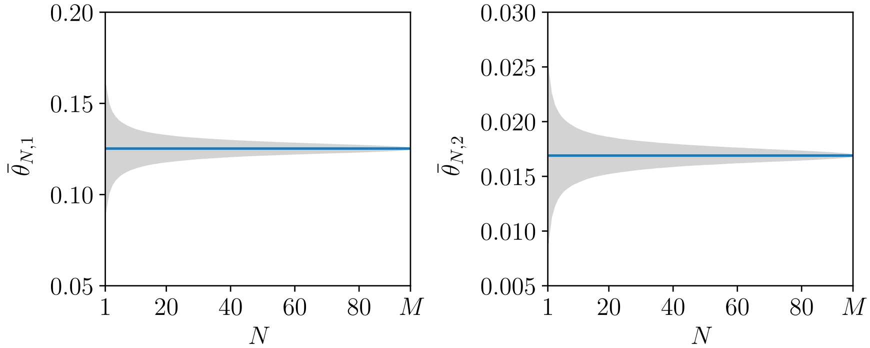
B Impact of the sequence of assets on expected utility

In this section, we study the impact of the sequence in which the assets are picked on $\bar{\boldsymbol{\theta}}_N$, which is defined in Proposition 1, and on the EU of the two-fund rule, $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ in (20). Specifically, we study how close $\bar{\boldsymbol{\theta}}_N$ is to $\bar{\boldsymbol{\theta}}_M$ and how close our EU approximation $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ in (22) is to $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ as a function of N.

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

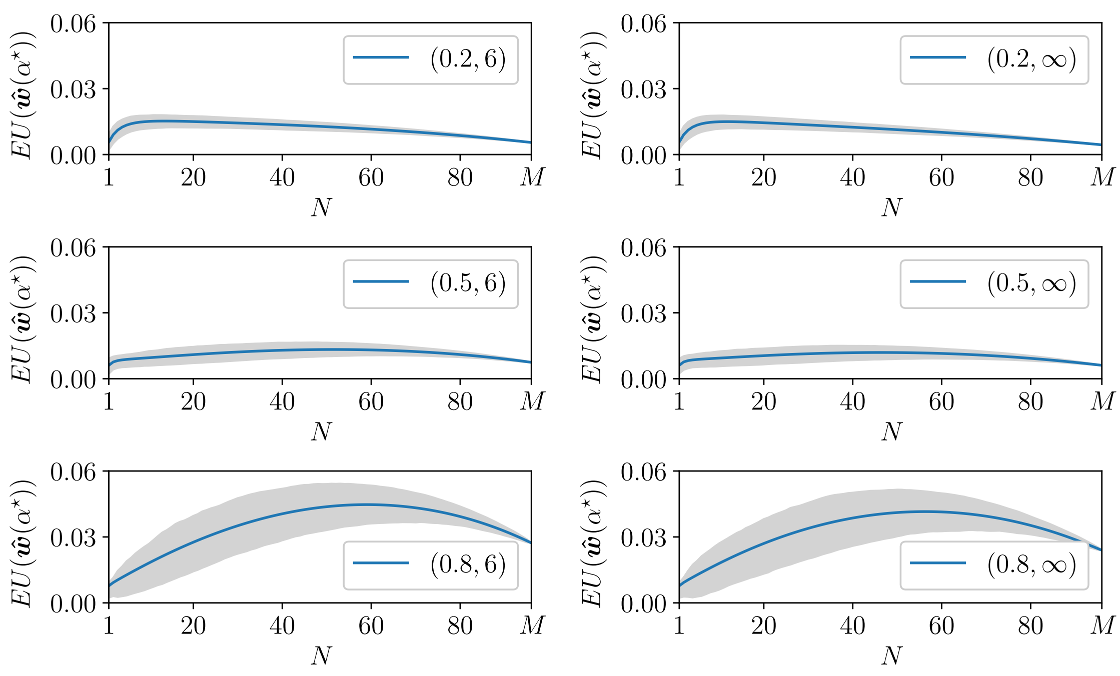
The results from this experiment are depicted in two figures. First, in Figure SM.1, we depict $\bar{\boldsymbol{\theta}}_M$ in solid blue and a one standard deviation interval of the 10,000 values of $\bar{\boldsymbol{\theta}}_N$ in shaded gray, as a function of N. We observe that the interval quickly shrinks as N increases. Second, in Figure SM.2, we depict, for a sample size $T = 120$, $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ in solid

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



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

Figure SM.2: Impact of the sequence of assets on the EU of the optimal two-fund rule



Notes. This figure depicts our approximation of the EU of the optimal two-fund rule $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$ in (22) as a function of the portfolio size N (solid blue line). We also depict in shaded gray the one standard deviation interval of the 10,000 values of $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ obtained with random sequences of assets. We calibrate $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3})$ to the t-distribution with $\nu \in \{6, \infty\}$ degrees of freedom, and consider $\rho \in \{0.2, 0.5, 0.8\}$. Each panel corresponds to a choice of (ρ, ν). We set $\gamma = 1$. We calibrate this experiment to a dataset of $M = 96$ portfolios sorted on size and book-to-market (96S-BM) spanning July 1963 to August 2023.

blue and a one standard deviation interval of the 10,000 values of $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_N)$ in shaded gray, as a function of N. We can observe that the gray area closely follows our curve for determining the optimal $N^\star_{2f}$, $EU_{2f}(N, T, \rho, \tau, \bar{\boldsymbol{\theta}}_M)$.

C Estimation of parameters

In this section, we explain how we estimate the different parameters that are needed as inputs in our different portfolio rules to determine the optimal portfolio size (i.e., $N^\star_{smv}$, $N^\star_{2f}$, $N^\star_{3f,g}$, $N^\star_{3f,ew}$) and the optimal combination coefficients (i.e., $\alpha^\star$, $\alpha^\star_1$, $\alpha^\star_2$, $\beta^\star_1$, $\beta^\star_2$).

C.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 (13)–(15). To speed up the computation, we use the high-dimensional approximation of these parameters. Specifically, from El Karoui (2010, 2013), we have that as $N, T \to \infty$ and $N/T \to \phi \in (0, 1)$, $(\kappa_{N,1}, \kappa_{N,2}, \kappa_{N,3}) \to (\tilde{\kappa}_{N,1}, \tilde{\kappa}_{N,2}, \tilde{\kappa}_{N,1})$, where $\tilde{\kappa}_{N,1} \geq 1$ is the unique positive solution to

$$\mathbb{E}\left[(1 - \phi + \phi\tilde{\kappa}_{N,1}\tau)^{-1}\right] = 1, \tag{SM23}$$

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

$$\tilde{\kappa}_{N,2} = (1 - \phi)\left(\tilde{\kappa}^{-2}_{N,1} - \mathbb{E}\left[\frac{\phi\tau^2}{(1 - \phi + \phi\tilde{\kappa}_{N,1}\tau)^2}\right]\right)^{-1}. \tag{SM24}$$

Kan and Lassance (2023) 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 (2023). First, we assume that the asset returns follow a t-distribution, i.e., $\tau \sim (\nu - 2)/\chi^2_\nu$, and we estimate the number of degrees of freedom ν by maximum likelihood from a sample of T historical returns $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)$, giving us $\hat{\nu}$. Then, we can use the closed-form expression for $\tilde{\kappa}_{N,1}$ and $\tilde{\kappa}_{N,2}$ when asset returns are t-distributed in Kan and Lassance (2023, Proposition 6), which yields that the estimate of $\tilde{\kappa}_{N,1}$, denoted $\tilde{\kappa}^{\nu}_{N,1}$, is the unique positive solution to

$$y e^{y} E_{\hat{\nu}/2}(y) = \phi_N \quad \text{with} \quad y = \frac{(\hat{\nu}-2)\phi_N \tilde{\kappa}^{\nu}_{N,1}}{2(1-\phi_N)} \quad \text{and} \quad \phi_N = \frac{N}{T}, \tag{SM25}$$

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

$$\tilde{\kappa}^{\nu}_{N,2} = \frac{2(\tilde{\kappa}^{\nu}_{N,1})^2(1-\phi_N)}{\hat{\nu} - \tilde{\kappa}^{\nu}_{N,1}(\hat{\nu}-2)}. \tag{SM26}$$

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

The second method we use to estimate $\tilde{\kappa}_{N,1}$ and $\tilde{\kappa}_{N,2}$ does not specify a particular parametric distribution for τ, but instead, relies on the following sample estimate of the distribution of τ proposed by El Karoui (2010, 2013):

$$\hat{\tau}_{N,t} = \frac{(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)'(\boldsymbol{r}_t - \hat{\boldsymbol{\mu}}_N)}{\frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_N)'(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_N)}, \quad t = 1, \ldots, T, \tag{SM27}$$

which is consistent as $N \to \infty$. 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 $\tilde{\kappa}^{s}_{N,1}$, is the unique positive solution to

$$\frac{1}{T}\sum_{t=1}^{T}\left(1 - \phi_N + \phi_N \tilde{\kappa}^{s}_{N,1}\hat{\tau}_{N,t}\right)^{-1} = 1, \tag{SM28}$$

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

$$\tilde{\kappa}^{s}_{N,2} = (1-\phi_N)\left((\tilde{\kappa}^{s}_{N,1})^{-2} - \frac{1}{T}\sum_{t=1}^{T}\frac{\phi_N \hat{\tau}^2_{N,t}}{(1-\phi_N+\phi_N\tilde{\kappa}^{s}_{N,1}\hat{\tau}_{N,t})^2}\right)^{-1}. \tag{SM29}$$

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.

C.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 (SM2), $\psi_{ew,N}^2$ in (SM13), $\mu_{g,N}$ in (SM2), and $\lambda_{ew,N}$ in (SM13).

For $\mu_{g,N}$ and $\lambda_{ew,N}$, we rely on the estimates that are unbiased when asset returns are normally distributed[SM2], i.e.,

$$\hat{\mu}_{g,N} = \frac{\mathbf{1}_N' \hat{\mathbf{\Sigma}}_N^{-1} \hat{\boldsymbol{\mu}}_N}{\mathbf{1}_N' \hat{\mathbf{\Sigma}}_N^{-1} \mathbf{1}_N}, \tag{SM30}$$

$$\hat{\lambda}_{ew,N} = \frac{T-3}{T} \frac{\boldsymbol{w}_{ew}' \hat{\boldsymbol{\mu}}_N}{\boldsymbol{w}_{ew}' \hat{\mathbf{\Sigma}}_N \boldsymbol{w}_{ew}}. \tag{SM31}$$

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

$$\hat{\theta}_{N,a}^2 = \frac{(T-N-2)\hat{\theta}_N^2 - N}{T} + \frac{2(\hat{\theta}_N^2)^{\frac{N}{2}} (1+\hat{\theta}_N^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\theta}_N^2/(1+\hat{\theta}_N^2)}\left(\frac{N}{2}, \frac{T-N}{2}\right)}, \tag{SM32}$$

$$\hat{\psi}_{g,N,a}^2 = \frac{(T-N-1)\hat{\psi}_{g,N}^2 - (N-1)}{T} + \frac{2(\hat{\psi}_{g,N}^2)^{\frac{N-1}{2}} (1+\hat{\psi}_{g,N}^2)^{\frac{2-T}{2}}}{T \times B_{\hat{\psi}_{g,N}^2/(1+\hat{\psi}_{g,N}^2)}\left(\frac{N-1}{2}, \frac{T-N+1}{2}\right)}, \tag{SM33}$$

$$\hat{\psi}_{ew,N,a}^2 = \frac{(T-N-2)\hat{\psi}_{ew,N}^2 - (N-1)(1+\hat{\theta}_{ew,N}^2)}{T} + \frac{2(1+\hat{\theta}_{ew,N}^2)^{\frac{T-N}{2}} (\hat{\psi}_{ew,N}^2)^{\frac{N-1}{2}} (1+\hat{\theta}_N^2)^{\frac{3-T}{2}}}{T \times B_{\hat{\psi}_{ew,N}^2/(1+\hat{\theta}_N^2)}\left(\frac{N-1}{2}, \frac{T-N}{2}\right)}, \tag{SM34}$$

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

[SM2] We can show that when asset returns are elliptically distributed as in Assumption 2, $\hat{\mu}_{g,N}$ and $\hat{\lambda}_{ew,N}$ are also unbiased in the high-dimensional asymptotic regime as $N, T \to \infty$ and $N/T \to \phi \in (0,1)$.

C.3 Estimation of assets' correlation and marginal performance

The third and last set of parameters are those that control the dependence between the assets, i.e., ρ under Assumption 1, and the marginal performance of the assets, i.e., the three functions $\delta_N(\bar{\boldsymbol{\theta}}_M)$ in (4), $r_{g,N}(\bar{\theta}_{M,1}, \bar{\lambda}_M, \bar{\boldsymbol{\sigma}}_M)$ in (SM7), and $r_{ew,N}(\bar{\mu}_M, \bar{\sigma}_{M,1}, \bar{\sigma}_{M,2})$ in (SM18). Recall that the parameters in these functions are computed on all M assets to find the optimal portfolio size N. We use an estimator $\hat{\rho}$ simply given by the average of all sample pairwise correlations $\hat{\rho}_{ij}$,

$$\hat{\rho} = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \hat{\rho}_{ij}. \tag{SM35}$$

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

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

$$\mathbb{E}[\hat{\delta}_N] - \delta_N = \frac{1-\rho}{T} \times \frac{1 - \frac{N}{M}\rho + N\rho}{1-\rho+N\rho}, \tag{SM36}$$

$$\mathbb{E}[\hat{r}_{g,N}] - r_{g,N} = \frac{1-\rho}{MT} \times \frac{\bar{\sigma}_{M,-2} - \frac{N\rho}{1-\rho+N\rho}\left(1 + \frac{(1-\rho)(1-\frac{M}{N})}{1-\rho+N\rho}\right)\bar{\sigma}^2_{M,-1}}{\bar{\sigma}_{M,-2} - \frac{N\rho}{1-\rho+N\rho}\bar{\sigma}^2_{M,-1}}, \tag{SM37}$$

$$\mathbb{E}[\hat{r}_{ew,N}] - r_{ew,N} = \frac{1-\rho}{MT} \times \frac{\bar{\sigma}_{M,2} + \frac{\rho}{1-\rho}M\bar{\sigma}^2_{M,1}}{\bar{\sigma}_{M,2} + \frac{\rho}{1-\rho}N\bar{\sigma}^2_{M,1}}. \tag{SM38}$$

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

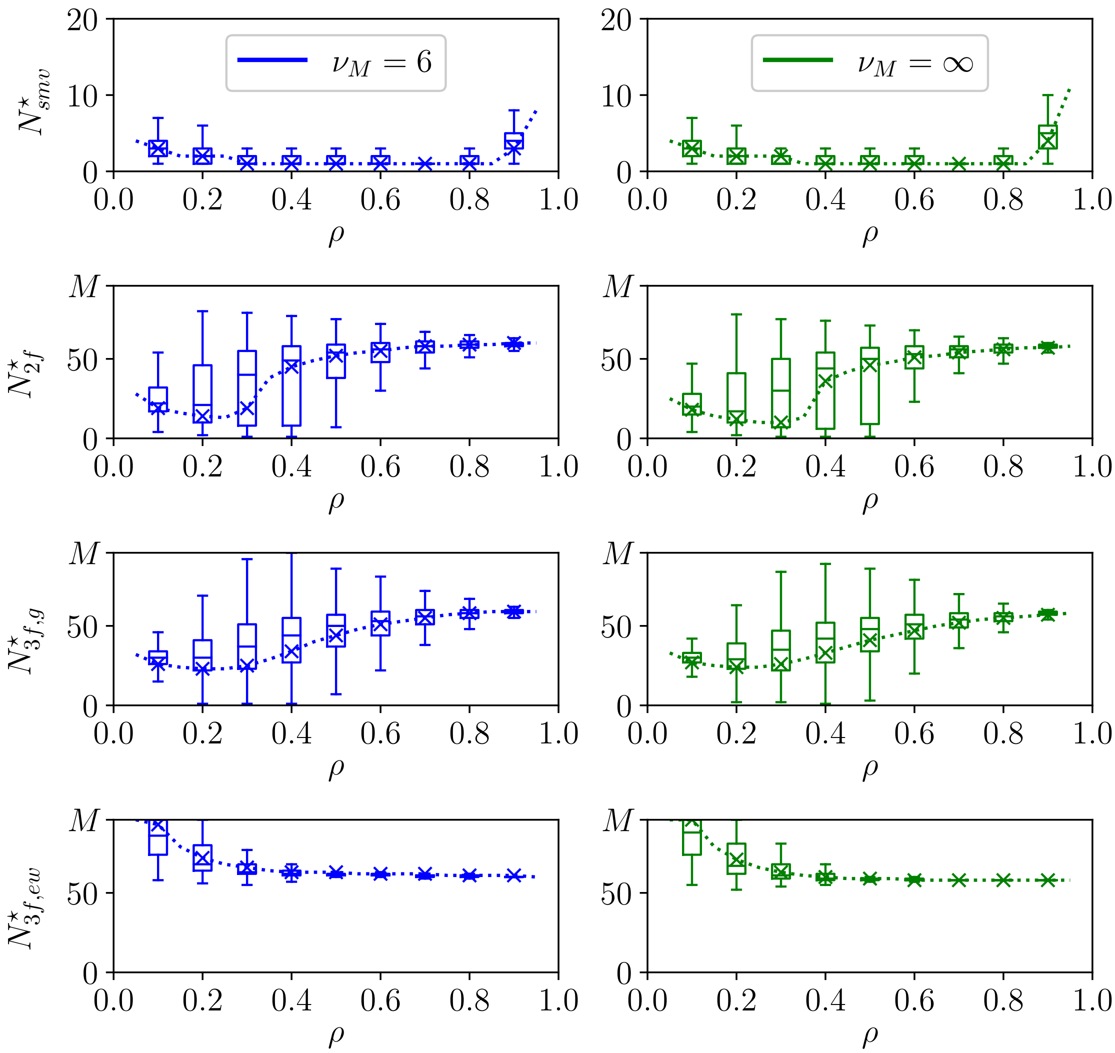
D Additional simulation and empirical results

We now report simulation and empirical results that complement those in the main body of the paper. In the simulation setting, Section D.1 looks at the impact of the correlation on the optimal portfolio size and Section D.2 considers a block-correlation matrix. In the empirical setting, Section D.3 looks at the equicorrelation and elliptical distribution assumptions, Section D.4 considers a nonlinear shrinkage estimator of the covariance matrix, Section D.5 applies our different asset selection rules to the SMV portfolio, Section D.6 reports the out-of-sample Sharpe ratio of the considered portfolios, Section D.7 evaluates the portfolio turnover and how much of it is driven by changes to the selected assets over time, and Section D.8 investigates the overlap between the selection rules implemented in the main text.

D.1 Impact of correlation on the optimal portfolio size

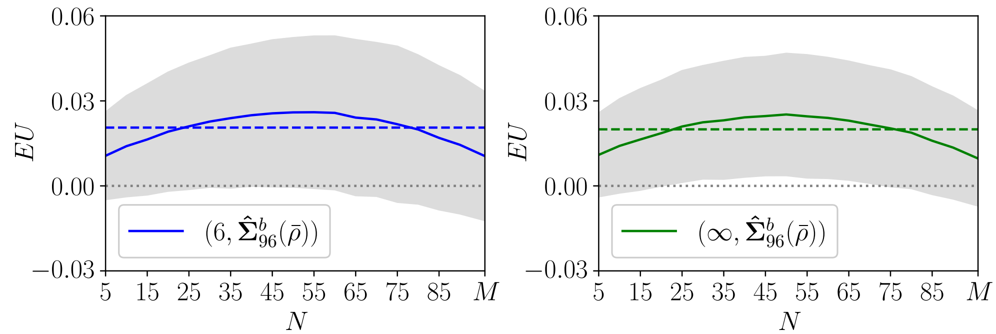
In this section, we run the same experiment as that in Section 4.1 except that we vary the correlation ρ. Specifically, we let $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\rho)$ and consider values of ρ between 0.1 and 0.9 with a step size of 0.1. Because $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$, we can compare the estimated optimal N with the oracle optimal N. We depict the results in Figure SM.3. We observe that the oracle optimal N for the two-fund and three-fund rules is slightly below $T/2$ when ρ is close enough to one, as observed in Figure 2 where $\rho = \bar{\rho} = 0.74$. In this case of high correlation, typical for equity data, the estimated optimal N is remarkably robust to variability in the parameters as the boxplots are quite thin. However, as ρ decreases away from one, the optimal N moves away from $T/2$ and the boxplots widen as N becomes more sensitive to the parameters. Finally, we also observe that overall the boxplots for the estimated optimal N are centered close to the oracle N.

Figure SM.3: Impact of correlation ρ on optimal portfolio size in simulated data



Notes. This figure depicts boxplots of the estimated optimal portfolio size N for the SMV portfolio ($\hat{N}^\star_{smv}$), two-fund rule ($\hat{N}^\star_{2f}$), GMV-three-fund rule ($\hat{N}^\star_{3f,g}$), and EW-three-fund rule ($\hat{N}^\star_{3f,ew}$). These are the estimated counterparts of $N^\star_{smv}$ in (21), $N^\star_{2f}$ in (22), $N^\star_{3f,g}$ in (SM11), and $N^\star_{3f,ew}$ in (SM22) following the estimation methodology in Section C of the supplementary material. The boxplots are obtained by simulating 10,000 times $T = 120$ t-distributed returns. Using the 96S-BM dataset, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\boldsymbol{\Sigma}_M = \hat{\boldsymbol{\Sigma}}_{96}(\rho)$, where ρ varies between 0.1 and 0.9 with a step size of 0.1., and consider $\nu \in \{6, \infty\}$ degrees of freedom. Each boxplot corresponds to a different choice of (ν, ρ). We depict with dotted lines and '×' markers the oracle value of the optimal N known under Assumption 1 because $\boldsymbol{\Sigma}_M$ is of the form $\boldsymbol{\Sigma}_M(\rho)$.

Figure SM.4: EU of the two-fund rule in simulated data with a block correlation matrix



Notes. This figure depicts the expected out-of-sample utility (EU) of the two-fund rule as a function of the portfolio size N. We simulate 10,000 samples of t-distributed returns with $\nu \in \{6, \infty\}$ degrees of freedom. For each N, we conduct a rolling window exercise as described in Section 4.2. Using the 96S-BM dataset, we set $\boldsymbol{\mu}_M = \hat{\boldsymbol{\mu}}_{96}$ and $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}^b_{96}(\bar{\rho})$, where $\bar{\rho} = 0.74$ is the average of all correlations in $\hat{\mathbf{\Sigma}}_{96}$ and the construction of the block-correlation matrix is described in Section D.2. Each panel corresponds to a different choice of ν. In each panel, the solid line depicts the EU in (25), the shaded gray area depicts the one standard deviation interval around the EU across all simulations, and the dashed horizontal line depicts the EU obtained when using the estimated optimal N for the two-fund rule, i.e., the estimate of $N^{\star}_{2f}$ in (22) according to the estimation method in Section C of the supplementary material. The dotted horizontal gray line depicts the zero EU level. We set a risk-aversion coefficient of $\gamma = 1$.

D.2 Block-correlation matrix

In this section, we run the same experiment as in Section 4.2 except that we use a different specification for the correlation matrix of asset returns. In Section 4.2, we consider two correlation structures: equicorrelation, i.e., we set $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}_{96}(\bar{\rho})$, and the empirical correlations we observe in the data, i.e., we set $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}_{96}$. To test the robustness of our approach, we now consider a third specification in which correlations form a block structure. Specifically, we set the population covariance matrix $\mathbf{\Sigma}_M = \hat{\mathbf{\Sigma}}^b_{96}(\bar{\rho}) = \hat{\boldsymbol{D}}_{96}\boldsymbol{P}^b_{96}(\bar{\rho})\hat{\boldsymbol{D}}'_{96}$ where $\boldsymbol{P}^b_{96}(\bar{\rho})$ is a block-correlation matrix. We consider two blocks of size $M/2$ for which the intra-correlation, i.e., correlation between assets of the same block, is $\bar{\rho} = 0.74$, and the inter-correlation, i.e., correlation between assets of different blocks, is zero. The two blocks are selected from the first and last 48 assets in the 96S-BM dataset. We depict the results in Figure SM.4. We observe that in this extreme setup, our estimated optimal N no longer delivers the best out-of-sample performance, but still largely outperforms investing in all or too few assets.

D.3 Equicorrelation and elliptical distribution in empirical data

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

Regarding Assumption 1, we now define a measure of distance between the correlation matrix observed in the data and the equicorrelation matrix. Specifically, we estimate ρ as the average of sample correlations,

$$\bar{\rho} = \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} \hat{\rho}_{ij}, \tag{SM39}$$

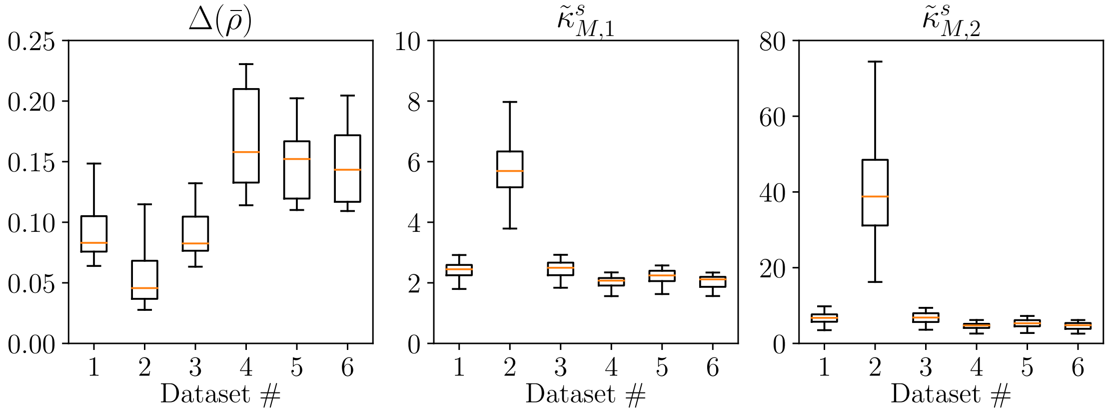
where $\hat{\rho}_{ij}$ is the sample correlation between assets i and j, and we define $\Delta(\bar{\rho})$ as the root mean squared error between the sample correlation matrix and the equicorrelation matrix obtained by setting $\rho = \bar{\rho}$, i.e.,

$$\Delta(\bar{\rho}) := \sqrt{\frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{j=i+1}^{M} (\hat{\rho}_{ij} - \bar{\rho})^2}. \tag{SM40}$$

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

Regarding Assumption 2, we look at the three parameters $(\kappa_{M,1}, \kappa_{M,2}, \kappa_{M,3})$ defined in (13)–(15) that control the impact of elliptical fat tails on the EU and that appear in the

Figure SM.5: Assumptions of equicorrelation and elliptical distribution in empirical data



Notes. This figure depicts boxplots of the estimated values of $\Delta(\bar{\rho})$ in (SM40), which measures the distance between the sample correlation matrix and the equicorrelation matrix, and $\tilde{\kappa}^s_{M,1}$ and $\tilde{\kappa}^s_{M,2}$ in (SM28)–(SM29), 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.

formulas for the optimal combination coefficients in the two-fund and three-fund rules. As explained in Section C.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 $\tilde{\kappa}^s_{M,1}$ in (SM28) and $\kappa_{M,2}$ with $\tilde{\kappa}^s_{M,2}$ in (SM29). The middle and right panels of Figure SM.5 depict boxplots of $\tilde{\kappa}^s_{M,1}$ and $\tilde{\kappa}^s_{M,2}$ across all estimation windows. Recall that the closer they are to one, the closer the data is to being normally distributed. We observe that $\tilde{\kappa}^s_{M,1}$ and $\tilde{\kappa}^s_{M,2}$ 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.

D.4 Nonlinear shrinkage covariance matrix

In Table SM.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.

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

Portfolio strat.	Asset select. rule	Sample covariance matrix: Dataset 96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV	Nonlinear shrinkage covariance matrix: Dataset 96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV
2F	All	−0.31	−59.6	−15.3	99.4	−18.8	56.8	4.64	−3E+31	1.05	44.4	2.19	16.4
2F	MaxSR	1.38	11.6●	−7.48	23.6○	−10.5	−27.1◑	1.86○	−3E+04●	−5.41	28.7◑	14.2●	10.5
2F	MinSR	25.4●	73.2●	7.10●	146	24.9●	139◑	19.1●	−2E+05●	5.54◑	104●	32.7●	68.5●
2F	BWSR	7.39	84.4●	−4.42	133	11.7○	242●	23.7●	−6E+05●	7.02	150●	39.1●	102●
2F	MaxVar	25.0◑	124●	17.9●	114	21.7●	83.8	24.9●	−2E+05●	9.04◑	126●	41.6●	72.1●
2F	MinVar	2.47	12.5●	−13.1	219◑	−1.88	225●	4.93	−3E+04●	−3.89	150●	21.2●	80.9●
2F	BWVar	13.0	89.3●	−6.76	182○	0.14	180●	22.6●	−5E+05●	5.16	141●	30.1●	93.3●
2F	MaxPC	8.80	18.3●	3.67◑	169○	9.62●	237●	15.0●	−9E+04●	5.66	125●	26.7●	89.1●
2F	MinPC	27.4●	112●	−6.23	115	17.9◑	94.1	26.3●	−5E+05●	−2.38	118●	40.5●	50.3●
2F	Best θ^2_N	3.77	14.7●	−8.77	182○	−7.87	137◑	19.1●	−5E+05●	−0.27	143●	30.5●	80.6●
2F	MaxW	−127●	−305●	−179●	−581●	−129●	−366●	33.4●	−2E+05●	9.79	205●	42.3●	132●
3FGMV	All	0.47	−60.5	−12.5	99.1	−19.7	55.4	5.45	−2E+31	1.64	44.7	3.13	16.5
3FGMV	MaxSR	−4.06	10.8●	−1.93○	15.6◑	−6.02	−27.4◑	2.99	−3E+04●	3.98	30.0◑	24.3●	11.3
3FGMV	MinSR	36.0●	71.6●	12.9●	150	31.5●	141◑	25.7●	−2E+05●	8.83◑	106●	37.9●	71.0●
3FGMV	BWSR	10.9	82.4●	1.02	143	19.9◑	253●	30.3●	−6E+05●	16.5●	152●	39.5●	104●
3FGMV	MaxVar	25.1◑	126●	24.9●	117	21.7●	85.0	27.3●	−2E+05●	12.9◑	128●	34.2●	74.2●
3FGMV	MinVar	6.13	13.3●	−14.9	222◑	0.35○	219●	7.25	−4E+04●	1.63	150●	25.2●	80.4●
3FGMV	BWVar	27.4◑	89.0●	2.74	183○	11.2◑	190●	28.4●	−5E+05●	13.1◑	145●	31.9●	94.6●
3FGMV	MaxPC	12.2	16.7●	10.8●	176○	7.50◑	225●	18.6●	−1E+05●	13.6●	127●	29.1●	90.7●
3FGMV	MinPC	33.3●	113●	6.47◑	115	25.9●	91.3	28.8●	−6E+05●	9.59○	120●	36.4●	52.7●
3FGMV	Best θ^2_N	12.3	7.39●	−0.08	122	1.33○	165●	22.9●	−6E+05●	10.2◑	147●	27.7●	81.0●
3FGMV	MaxW	−136●	−300●	−147●	−650●	−110●	−325●	38.3●	−2E+05●	20.0●	210●	43.7●	132●
3FEW	All	1.83	−62.8	−10.8	102	−16.8	60.8	6.82	−3E+31	2.95	43.2	3.06	17.2
3FEW	MaxSR	3.68	12.6●	−0.61	23.6◑	−12.1	−20.4◑	3.60	−3E+04●	2.85	28.1◑	6.34	11.6
3FEW	MinSR	24.8●	70.8●	11.3●	145	24.8●	146◑	18.6●	−2E+05●	7.41	102●	25.9●	68.0●
3FEW	BWSR	8.43	84.4●	3.41	135	17.9◑	237●	24.8●	−5E+05●	11.4◑	147●	34.8●	101●
3FEW	MaxVar	26.0◑	122●	25.3●	124	16.0●	98.9	23.5●	−2E+05●	12.2◑	126●	31.1●	71.4●
3FEW	MinVar	4.93	12.3●	−5.87	220◑	0.40	230●	7.97	−3E+04●	4.06	150●	15.6●	82.2●
3FEW	BWVar	14.2	88.6●	−3.30	195○	−0.15	187●	21.9●	−5E+05●	9.71◑	141●	21.3●	93.6●
3FEW	MaxPC	9.42	18.1●	11.9●	162	5.87◑	247●	17.4●	−9E+04●	10.7◑	120●	20.7●	89.1●
3FEW	MinPC	26.0◑	114●	2.26	120	15.9◑	103	23.9●	−5E+05●	6.26	120●	31.4●	52.8●
3FEW	Best θ^2_N	12.5	4.58●	−4.45	136	−8.83	166●	19.1●	−6E+05●	6.75	148●	21.8●	78.8●
3FEW	MaxW	−124●	−305●	−157●	−650●	−139●	−361●	23.5●	−2E+05●	8.38○	211●	12.2◑	131●
EW	All	8.11	2.80	8.00	3.98	7.13	5.61	8.11	2.80	8.00	3.98	7.13	5.61
EW	MaxSR	8.77	4.72●	8.49	5.97◑	7.40	6.25◑	8.77	4.72●	8.49	5.97◑	7.40	6.25
EW	MinSR	7.39●	0.83●	7.55●	2.75	6.77●	5.18◑	7.39●	0.83●	7.55	2.75●	6.77●	5.18●
EW	BWSR	7.94	2.42●	7.72	2.86	7.35◑	5.28●	7.94●	2.42●	7.72◑	2.86●	7.35●	5.28●
EW	MaxVar	7.95◑	1.38●	7.97●	3.52	6.68●	5.31	7.95●	1.38●	7.97◑	3.52●	6.68●	5.31●
EW	MinVar	8.60	4.23●	8.18	4.36◑	7.44	5.86●	8.60	4.23●	8.18	4.36●	7.44●	5.86●
EW	BWVar	7.60	2.34●	7.52	3.99○	6.88	5.58●	7.60●	2.34●	7.52◑	3.99●	6.88●	5.58●
EW	MaxPC	8.32	2.97●	8.83●	2.19	7.37◑	5.08●	8.32●	2.97●	8.83◑	2.19●	7.37●	5.08●
EW	MinPC	7.87◑	2.48●	7.20	5.00	6.97◑	6.22	7.87●	2.48●	7.20	5.00●	6.97●	6.22●
EW	Best θ^2_N	7.88	2.72●	7.92	3.82	7.02	6.03●	8.08●	2.79●	8.16	3.89●	6.79●	5.55●
EWRF		1.42	−4.35	1.26	−3.06	0.80	−0.29	1.42	−4.35	1.26	−3.06	0.80	−0.29
SGMV		1.76	−49.7	4.25	−0.34	−2.28	−4.67	10.1	2.38	6.52	3.88	7.66	6.04
GMVRF		1.81	−26.3	−0.98	−11.4	−12.3	−5.34	3.81	−3E+34	1.25	0.19	2.65	0.82
SMV		−4E+05	−3E+11	−1E+06	−6E+07	−8E+08	−8E+07	−994	−1E+44	−974	−2115	−2744	−747
SMV-L_1		15.8●	−108●	11.7●	−579●	−2.91●	0.88●	21.6●	−349●	7.03●	−243●	−32.2●	9.12●
SMV-TW		−35.3●	−302●	−16.8●	−105●	−25.9●	−143●	−28.0●	−4E+32●	−29.6●	−421●	−288●	−84.8●

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 10 portfolio strategies in Table 2, the 11 asset selection rules in Table 3, across the six datasets in Table 1. The table is constructed following the empirical methodology described in Section 5.4. We construct the portfolios either with the sample or the nonlinear shrinkage covariance matrix of Ledoit and Wolf (2020). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under the 'All' selection rule versus the 10 other selection rules, using the block bootstrap methodology described in Section 5.4. We also report p-values for SMV-TW and SMV-L_1 versus SMV. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

Table SM.2: Annualized net out-of-sample utility of the restricted SMV portfolio

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	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV
SMV	All	−4E+05	−3E+11	−1E+06	−6E+07	−8E+08	−8E+07	−1443	−107	−1399	−2360	−1457	−1143
SMV	MaxSR	−17.3•	−25.3•	−2011•	−38.5•	−27.2•	−21.5•	−15.8•	0.93	−34.6•	−37.6•	−25.7•	−19.8•
SMV	MinSR	−16.0•	−2.04•	−2007•	−43.4•	−12.8•	−17.9•	−14.5•	18.2	−31.7•	−13.2•	−13.1•	−17.8•
SMV	BWSR	−50.9•	18.5•	−2050•	−62.2•	−54.0•	−54.6•	−33.3•	95.1○	−63.4•	−8.63•	−44.0•	−34.3•
SMV	MaxVar	−5.94•	−57.3•	−1977•	−26.4•	−8.54•	−18.4•	−3.90•	47.1	−9.97•	−23.6•	−9.36•	−18.6•
SMV	MinVar	−22.1•	−35.1•	−1998•	−52.4•	0.22•	14.7•	−16.8•	4.38	−13.0•	30.3•	7.30•	25.7•
SMV	BWVar	0.73•	−83.2•	−1986•	−50.3•	−14.3•	−7.33•	2.36•	−1.50	−8.42•	−13.2•	−8.47•	−0.28•
SMV	MaxPC	−14.9•	26.8•	−2009•	−68.5•	−4.96•	−41.6•	−8.01•	30.8	−20.0•	11.6•	1.06•	−1.16•
SMV	MinPC	−19.3•	−49.7•	−2004•	−12.7•	−10.9•	−2.23•	−16.2•	−11.9	−30.8•	−11.1•	−10.4•	−2.07•
SMV	Best θ^2_N	−0.98•	34.2•	−1996•	22.1•	−44.8•	11.4•	9.97•	34.7	−30.6•	39.9•	−16.1•	3.14•
SMV	MaxW	−15.3•	−82.5•	−2021•	−76.2•	−16.3•	−78.4•	−17.5•	50.9	−45.4•	22.5•	−0.62•	−14.5•
SMV-L_1		15.8•	−108•	11.7•	−579•	−2.91•	0.88•	23.9•	67.8	5.13•	−195•	10.4•	16.4•
SMV-TW		−35.3•	−302•	−16.8•	−105•	−25.9•	−143•	−29.9•	−52.1	−17.8•	−376•	−33.6•	−349•

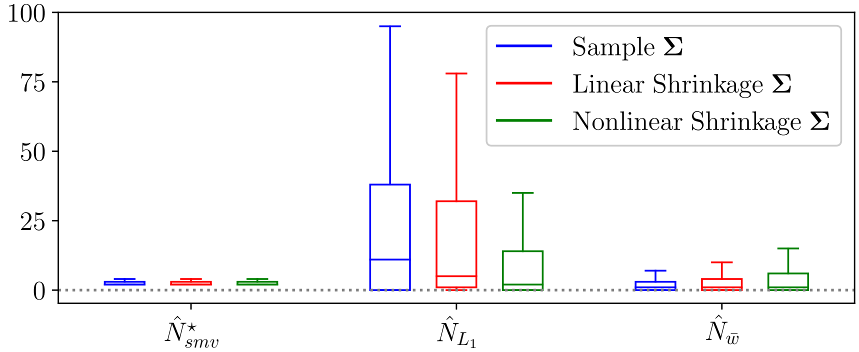
Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the sample mean-variance portfolio across the six datasets in Table 1 and the 11 asset selection rules in Table 3, as well as two additional ways to implement the SMV portfolio described in Section 5.2: the norm constraint (L_1) and the threshold on weights (TW). The table is constructed following the empirical methodology described in Section 5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$.

D.5 Performance of optimally restricted SMV portfolio

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

We now compare the obtained portfolio sizes for the three different ways of restricting

Figure SM.6: Estimated portfolio sizes for the SMV portfolio



Notes. This figure depicts boxplots of the portfolio sizes obtained in empirical data for the SMV portfolio with three different strategies detailed in Section 5.2. First, our estimated portfolio size for the SMV portfolio, $\hat{N}^{\star}_{smv}$. Second, the implicit portfolio size obtained with L_1 norm regularization as given by (28), $\hat{N}_{L_1}$. Third, the portfolio size obtained using a cross-validated threshold on the weights as given by (29), $\hat{N}_{\bar{w}}$. We consider three different estimators of the covariance matrix, the sample covariance matrix given by (6) in blue and the linear and nonlinear shrinkage estimators of Ledoit and Wolf (2004, 2017), in red and green respectively. We average values obtained across the six datasets considered in the Empirical Analysis of Section 5.

the size of the SMV portfolio we consider in Table SM.2. First, using our estimated portfolio size $\hat{N}^{\star}_{smv}$ that we estimate following the methodology in Section C of the supplementary material. Second, the implicit portfolio size obtained by implementing L_1-norm regularization as detailed in Section 5.2. Specifically, at each month t, we solve (28) by finding the optimal δ with cross-validation, which yields the vector of weights $\hat{\boldsymbol{w}}_{L_1,t}$, from which we can compute the implicit portfolio size $\hat{N}_{L_1,t}$,

$$\hat{N}_{L_1,t} = \sum_{i=1}^{M} \mathbb{1}_{\{|\hat{w}_{L_1,i,t}|>0\}}. \tag{SM41}$$

Third, we consider the portfolio size $\hat{N}_{\bar{w}}$ obtained implicitly by finding the optimal $\bar{w}$ with cross-validation, as detailed in Section 5.2. We depict the results in Figure SM.6 and observe that the portfolio sizes for the SMV portfolio are small on average compared to M, and that $\hat{N}_{L_1}$ displays a much higher variance than $\hat{N}^{\star}_{smv}$ and $\hat{N}_{\bar{w}}$.

D.6 Out-of-sample Sharpe ratio

Table 4 in the main body of the paper reports the out-of-sample utility as a performance measure because it is the metric optimized by the portfolio combination rules in our theory. Another important mean-variance portfolio performance measure is the Sharpe ratio. Therefore, in Table SM.3, we report the annualized out-of-sample Sharpe ratio net of proportional transaction costs of the portfolio strategies considered in Table 4,

$$SR_k = \sqrt{12} \times \frac{\hat{\mu}_k}{\hat{\sigma}_k}, \tag{SM42}$$

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 (30). 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.

D.7 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 $N \leq M$. The portfolio turnover is impacted by the changes to the investment universe over time implied by these selection rules. The larger the number of changes in selected assets in a given period, the lower the *stability* of the selection rule, and the higher the portfolio turnover.

The impact of a selection rule on portfolio turnover also depends the *selection frequency*, which is the frequency at which we decide to change the optimal N and the selected assets. In the empirical analysis of Section 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,

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

		Sample covariance matrix						Linear shrinkage covariance matrix					
		Dataset						Dataset					
Port-folio strat.	Asset select. rule	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV
2F	All	24.7	−150	−26.8	149	−6.46	113	89.4	214	40.1	244	72.6	209
2F	MaxSR	19.4	59.5●	−1.83	97.7◑	−0.85	44.7●	19.8●	162●	−2.01◑	81.5●	12.5●	55.4●
2F	MinSR	72.3●	121●	38.3●	171	70.6●	168◑	86.0	215	44.4	195●	77.5	180○
2F	BWSR	52.5◑	130●	33.4●	177	65.9●	220●	69.6○	215	32.1	223◑	67.1	200
2F	MaxVar	71.4●	163●	60.7●	157	72.7●	135	83.8	238●	50.7	197●	74.1	159●
2F	MinVar	32.0	65.4●	−13.6	211●	23.7○	212●	38.1●	179●	−2.61◑	225○	39.1◑	206
2F	BWVar	56.3◑	134●	26.7●	200◑	37.2◑	198●	75.8	217	33.0	214●	52.7	190○
2F	MaxPC	46.7	68.8●	33.2●	184○	46.9●	218●	71.8	194○	31.0	228○	61.0	236◑
2F	MinPC	74.2●	150●	12.3◑	167	71.9●	140	89.9	228○	1.89◑	176●	79.4	137●
2F	Best θ_N^2	42.6	75.4●	13.6◑	193○	35.8◑	169◑	83.9	211	16.0	203●	73.0	173◑
2F	MaxW	51.5●	3.14●	0.47◑	171	29.4●	177●	87.5	233●	28.7	238	63.2	210
3FGMV	All	26.9	−151	−14.2	149	−3.63	112	88.5	215	57.0	246	80.5	208
3FGMV	MaxSR	9.59	59.1●	8.19	94.7●	15.6	46.8●	18.2●	161●	27.7○	83.3●	35.7●	53.0●
3FGMV	MinSR	89.9●	120●	50.8●	173	79.3●	169◑	104	215	55.8	200●	91.4	183
3FGMV	BWSR	59.8●	128●	39.1●	181○	74.5●	225●	77.9	215	57.3	226○	80.1	198
3FGMV	MaxVar	71.1●	164●	70.7●	159	74.0●	136	89.8	238●	62.2	197●	76.8	156●
3FGMV	MinVar	41.4	66.2●	−18.3	213●	32.6◑	209●	42.5●	179●	5.01●	225○	48.6◑	202
3FGMV	BWVar	74.8●	133●	37.9●	201◑	53.3●	202●	87.8	216	59.0	218◑	65.2	186○
3FGMV	MaxPC	52.1○	67.4●	47.8●	188○	44.5●	212●	74.7	192◑	64.9	230	61.4	234○
3FGMV	MinPC	81.6●	151●	40.5●	167	80.9●	139	88.2	229	45.7	177●	88.3	136●
3FGMV	Best θ_N^2	53.7○	71.0●	29.6●	172	42.7●	182●	61.1◑	196◑	59.5	227○	73.7	162●
3FGMV	MaxW	47.3◑	3.59●	13.5◑	165	35.7●	182●	93.8	233●	56.7	242	80.8	208
3FEW	All	38.9	−117	12.1	151	15.9	116	40.2	5.52	32.7	135	31.8	66.6
3FEW	MaxSR	41.8	62.3●	33.9●	98.6●	27.0	51.2●	40.2	85.9●	36.9	67.3●	38.3○	49.6○
3FEW	MinSR	70.5●	119●	49.2●	170	70.9●	171◑	61.4●	175●	38.4	187◑	57.3◑	167●
3FEW	BWSR	58.4◑	130●	45.7●	178	73.7●	218●	64.4●	190●	44.5○	214●	66.8●	181●
3FEW	MaxVar	73.3●	161●	71.6●	162	69.8●	144	69.1●	212●	50.7●	191◑	60.5●	148●
3FEW	MinVar	45.6	66.5●	23.1	211●	44.2◑	215●	45.7	118●	34.7	217●	44.9●	172●
3FEW	BWVar	60.8○	133●	37.3◑	204◑	43.2○	201●	62.0●	191●	44.6○	210●	44.8○	175●
3FEW	MaxPC	53.8	69.3●	51.9●	180	47.9◑	223●	58.5●	135●	45.4◑	211●	48.2●	203●
3FEW	MinPC	75.3●	152●	38.2◑	170	72.4●	146	66.8●	198●	36.6	176○	65.7●	120●
3FEW	Best θ_N^2	56.9○	69.4●	29.6○	178	37.5○	182●	50.8○	157●	45.5◑	218●	51.7●	143●
3FEW	MaxW	50.6	2.91●	10.4	168	24.9	175●	70.7●	154●	39.8	237●	34.1	205●
EW	All	53.0	24.1	52.7	31.9	50.4	41.5	53.0	24.1	52.7	31.9	50.4	41.5
EW	MaxSR	57.6	35.0●	55.6●	44.6●	52.9	45.6●	57.6	35.0●	55.6	44.6●	52.9○	45.6○
EW	MinSR	48.1●	14.4●	49.7●	24.5	47.2●	38.5◑	48.1●	14.4●	49.7	24.5◑	47.2◑	38.5●
EW	BWSR	52.0◑	22.2●	50.9●	25.1	51.4●	39.4●	52.0●	22.2●	50.9○	25.1●	51.4●	39.4●
EW	MaxVar	49.5●	17.2●	49.9●	28.3	45.0●	38.1	49.5●	17.2●	49.9●	28.3◑	45.0●	38.1●
EW	MinVar	58.9	32.9●	56.7	35.4●	55.6◑	45.0●	58.9	32.9●	56.7	35.4●	55.6●	45.0●
EW	BWVar	50.0○	21.8●	49.7◑	32.1◑	49.0○	41.4●	50.0●	21.8●	49.7○	32.1●	49.0○	41.4●
EW	MaxPC	52.8	24.9●	56.0●	21.1	50.9◑	37.8●	52.8●	24.9●	56.0◑	21.1●	50.9●	37.8●
EW	MinPC	52.4●	22.5●	48.9◑	38.9	50.0●	45.5	52.4●	22.5●	48.9	38.9○	50.0●	45.5●
EW	Best θ_N^2	51.4○	23.7●	52.2○	31.1	49.7○	44.1●	53.9○	23.8●	51.6◑	31.7●	49.2●	43.5●
EWRF		31.4	1.11	31.0	8.93	30.1	21.1	31.4	1.11	31.0	8.93	30.1	21.1
SGMV		19.3	−193	29.6	7.82	6.33	−6.63	82.0	31.1	68.0	42.4	77.5	59.6
GMVRF		23.8	−177	9.52	−27.6	−10.4	0.81	65.6	−1.26	53.0	5.35	50.9	14.0
SMV		−50.2	−29.7	−14.7	−91.5	−16.5	−76.9	91.2	240	40.0	243	72.1	206
SMV-L_1		77.5●	167●	64.0●	163●	78.9●	177●	77.0	175●	43.8	157●	65.5	147●
SMV-TW		36.0●	−97.3●	29.6◑	24.7●	20.7	46.4●	21.2●	103●	18.8	128●	15.6●	108●

Notes. This table reports the annualized net out-of-sample Sharpe ratio, in percentage points, for the 10 portfolio strategies in Table 2, the 11 asset selection rules in Table 3, across the six datasets in Table 1. The table is constructed following the empirical methodology described in Section 5.4. We construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004). The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F, 3FGMV, and 3FEW portfolios under the 'All' selection rule versus the 10 other selection rules, using the block bootstrap methodology described in Section 5.4. We also report p-values for SMV-TW and SMV-L_1 versus SMV. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

and thus, it is important to test the robustness of our results to this choice.

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

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

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

$$\phi_{i,k,t} = \sum_{j=1}^{M} \mathbb{1}_{\{\delta_{i,k,t} \leq \delta_{j,k,t}\}}. \tag{SM43}$$

Third, we select assets depending on their rank $\phi_{i,k,t}$ and the type of selection rule. For that purpose, we denote $y_{i,k,t}$ the binary variable indicating whether, for selection rule k, asset i is selected at time t: $y_{i,k,t} = 1$ if asset i is selected, and 0 otherwise. We consider three types of asset selection rules in Table 3. For the MaxSR, MaxVar, MaxPC, and MaxW rules that select the N assets with the highest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} \leq N\}}$. For the MinSR, MinVar and MinPC rules that select the N assets with the lowest $\delta_{i,k,t}$ values, we have $y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} > M-N\}}$. Finally, for the BWSR and BWVar rules that select the $\lceil N/2 \rceil$ assets with the highest $\delta_{i,k,t}$ values and the $\lfloor N/2 \rfloor$ assets with the lowest $\delta_{i,k,t}$ values, we have

$y_{i,k,t} = \mathbb{1}_{\{\phi_{i,k,t} \leq \lceil N/2 \rceil \text{ or } \phi_{i,k,t} > M - \lfloor N/2 \rfloor\}}$.[SM3]

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

$$\text{selectionturnover}_{k,t} = \sum_{i=1}^{M} |y_{i,k,t} - y_{i,k,t-1}|, \quad t = T+1, \ldots, \tau. \tag{SM44}$$

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

$$\text{turnover}_{k,t}^{\text{selection}} = \sum_{i:\, y_{i,k,t} \neq y_{i,k,t-1}} |w_{i,k,t} - w_{i,k,(t-1)^+}|, \quad t = T+1, \ldots, \tau, \tag{SM45}$$

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 (31) 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:[SM4]

$$\text{turnover}_{k,t}^{\text{selection/all}} = \frac{\text{turnover}_{k,t}^{\text{selection}}}{\text{turnover}_{k,t}} \leq 1, \tag{SM46}$$

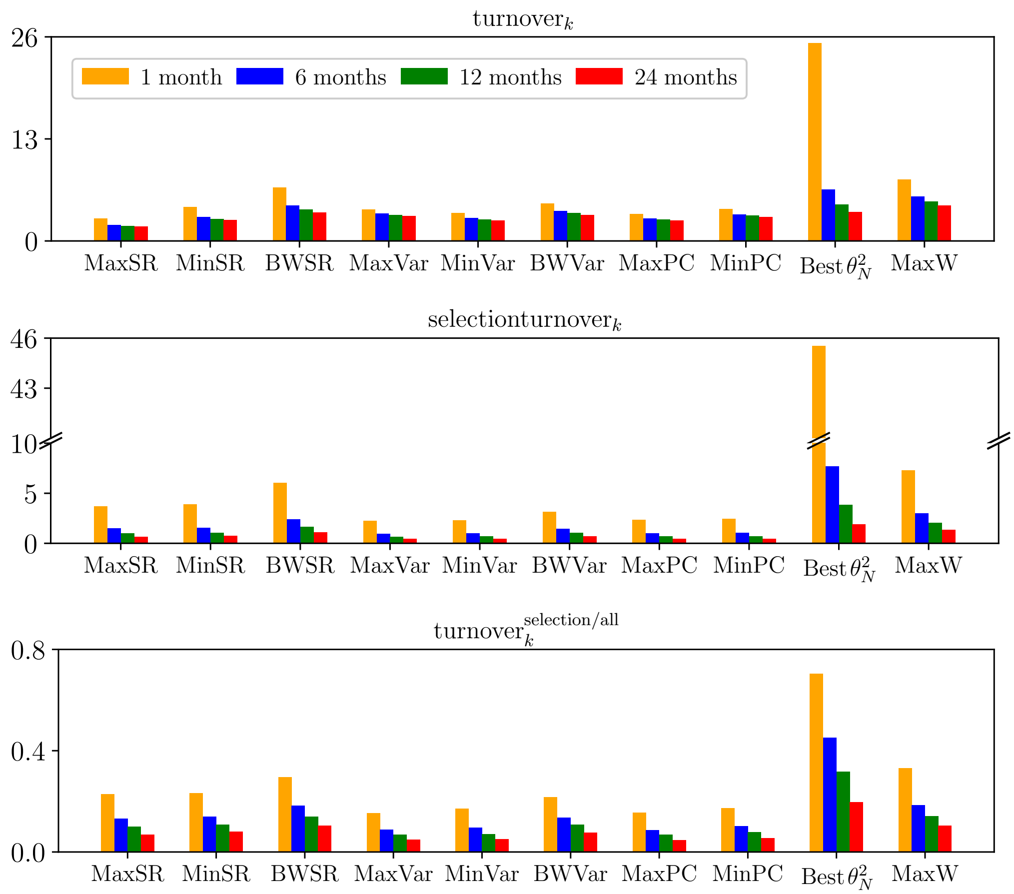
which we expect to increase with $\text{selectionturnover}_{k,t}$.

In Figure SM.7, we depict the portfolio turnover in (31), the selection turnover in (SM44), and the proportion of portfolio turnover explained by changes in the selected assets in (SM46).

[SM3] 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, $y_{i,k,t} = 0$.

[SM4] 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 SM.7: Portfolio turnover with asset selection rules and different selection frequencies



Notes. This figure depicts different turnover metrics for the two-fund portfolio rule across the 10 asset selection rules in Table 3 (we exclude the 'All' rule). The top panel depicts the average of the monthly portfolio turnover, defined in (31). The middle panel depicts the average of the monthly selection turnover, defined in (SM44), 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 (SM45). These metrics are averaged across the 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 construct the two-fund rule with the linear shrinkage covariance matrix of Ledoit and Wolf (2004) and a risk-aversion coefficient of $\gamma = 1$. The figure is constructed following the out-of-sample methodology described in Section 5.4. The middle panel includes '//' markers indicating an axis break for better visibility.

We average these measures across all months t and across the six datasets described in Table 1. We report results for each of the four considered selection frequencies, i.e., every 1, 6, 12, and 24 months. We only consider the 2F portfolio combination rule for conciseness; the results for the 3FGMV and 3FEW portfolios are similar. Moreover, we only report the results for the case in which we use the linear shrinkage covariance matrix of Ledoit and Wolf (2004) because, as discussed in Section 5.5, it leads to better out-of-sample performance, and also lower portfolio turnover. We make several observations from Figure SM.7.

First, we look at the stability of the different asset selection rules. The middle panel of Figure SM.7 shows that the Best θ_N^2 selection rule has the largest selection turnover, especially for a monthly selection frequency. This is expected as this rule is based on random selections. Then, the ex-post selection rule MaxW has the second largest selection turnover because it ranks assets according to portfolio weights and these weights can vary substantially from one period to another. Then, selection rules that rank assets based on the marginal Sharpe ratio are also quite unstable because they depend on estimates of expected returns that are notoriously noisy.[SM5] In contrast, selection rules built on the variance and PCA are remarkably stable, with the set of selected assets changing by only one or two assets on average each month when we use a monthly selection frequency. Finally, we also observe that the best-worst asset selection rules, BWSR and BWVar, are more unstable because they select the assets with the most extreme rankings, which are also more subject to more estimation errors.

Second, the bottom panel of Figure SM.7 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.

[SM5]The number of changes in selected assets can be substantial. For instance, on average across the six datasets, the selection turnover of BWSR in Figure SM.7 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.

Third, the bottom panel of Figure SM.7 shows that, except for Best θ_N^2, the proportion of the total portfolio turnover explained by assets that enter and exit the set of selected assets each month is around 10% to 30% on average when we use a monthly selection frequency but, as expected, decreases substantially as we decrease the selection frequency. This proportion of 10% to 30% with a monthly selection frequency is not negligible, particularly given that the selection turnover in the middle panel of Figure SM.7 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 SM.7 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 SM.4 the annualized net out-of-sample utility of the 2F portfolio implemented with the 11 asset selection rules in Table 3 and a selection frequency of 1, 6, 12, and 24 months. The results for the 3FGMV and 3FEW portfolio combination rules are similar. We make three main observations from Table SM.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 our use of a 12-month frequency in the main body of the paper is robust to other possible choices.

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

D.8 Overlap between selection rules

We show in Section 5 of the main text that reducing the portfolio size from M to N is beneficial in most cases by implementing different selection rules listed in Table 3. These selection rules are different by construction (marginal information vs. information on dependence, ex-ante vs. ex-post approach) to test whether our results hold for different types of selection rule. Thus, we raise the following question: do these rules indeed select dissimilar assets in practice?

To answer this question, we must measure the *overlap* between two selection rules, i.e., the number of assets jointly selected by both selection rules. Following Section D.7, we denote by $\boldsymbol{y}_{k,t} = (y_{1,k,t}, y_{2,k,t}, \ldots, y_{M,k,t})$ the $(M \times 1)$ selection vector for any selection rule k: $y_{i,k,t} = 1$ if asset i is selected in rule k at time t, and 0 otherwise. The overlap between rules k and l is then given by $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \sum_{i=1}^{M} y_{i,k,t}y_{i,l,t}$.

We then compare the overlap between the selection rules listed in Table 3 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

Table SM.4: Annualized net out-of-sample utility in empirical data with different selection frequencies

		Sample covariance matrix						Linear shrinkage covariance matrix					
		Dataset						Dataset					
Asset select. rule	Selection Frequency	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV	96 S-BM	108 CHA	100 S-OP	94 IN-NV	107 IN-CHA	98IN-CHA-NV
All	n.a.	−0.31	−59.6	−15.3	99.4	−18.8	56.8	5.11	1.20	1.09	41.6	1.09	17.7
MaxSR	1 month	−5.80	−30.0○	−7.72	2.25◑	−21.7	−35.2●	−1.98●	19.8●	−1.94	21.4●	0.66	7.66◑
MaxSR	6 months	0.77	3.62●	−5.15	31.7	−9.86	−31.1●	0.66●	22.3●	−2.71	26.5◑	3.72	10.3○
MaxSR	12 months	1.38	11.6●	−7.48	23.6○	−10.5	−27.1◑	1.93○	23.1●	−6.10	23.1●	0.69	10.8○
MaxSR	24 months	2.02	16.2●	−4.14○	34.8○	−13.3	−18.8◑	2.56○	22.8●	−1.54	22.4●	−0.05	8.59◑
MinSR	1 month	19.3◑	25.9●	6.84●	123	19.2●	112○	16.0●	50.7●	5.21○	97.9●	12.0●	67.3●
MinSR	6 months	25.5●	70.1●	8.49●	145	26.8●	130◑	17.4●	53.4●	6.54●	98.3●	14.8●	70.3●
MinSR	12 months	25.4●	73.2●	7.10●	146	24.9●	139◑	18.0●	53.2●	5.21○	101●	16.0●	72.1●
MinSR	24 months	25.4●	85.3●	9.09●	167	19.3●	150●	19.6●	53.9●	6.28◑	107●	15.2●	74.7●
BWSR	1 month	−2.88	14.3●	−14.5	161	−8.83	168●	16.5○	61.8●	3.02	140●	18.8●	88.8●
BWSR	6 months	0.55	66.1●	−2.22	154	−4.83	232●	18.1◑	63.1●	8.21	143●	20.1●	95.8●
BWSR	12 months	7.39	84.4●	−4.42	133	11.7○	242●	19.2◑	63.8●	4.74	143●	18.8◑	99.5●
BWSR	24 months	16.0	80.7●	−0.23	96.1	10.8○	219●	21.7●	64.0●	8.18	133●	13.0	103●
MaxVar	1 month	27.7●	112●	10.0●	112	20.9●	76.9	26.4●	69.1●	9.46◑	127●	20.5●	77.6●
MaxVar	6 months	26.5◑	127●	17.0●	118	21.9●	73.3	24.1●	70.0●	10.6◑	125●	22.6●	78.0●
MaxVar	12 months	25.0◑	124●	17.9●	114	21.7●	83.8	23.2●	70.5●	9.99◑	122●	21.1●	77.7●
MaxVar	24 months	29.8●	129●	18.6●	113	19.5●	91.9	25.1●	69.8●	12.0●	120●	18.0●	78.5●
MinVar	1 month	1.97	−18.5◑	−14.5	215◑	−0.88○	199●	5.17	32.6●	−2.38	136●	7.27◑	80.1●
MinVar	6 months	5.32	7.83●	−10.5	213◑	0.57○	222●	6.30	32.7●	−1.07	141●	6.66◑	78.8●
MinVar	12 months	2.47	12.5●	−13.1	219◑	−1.88	225●	6.08	33.1●	−5.25○	140●	5.40○	81.4●
MinVar	24 months	2.04	20.0●	−12.6	227●	−3.33	226●	5.68	33.8●	−4.89○	140●	4.11	81.8●
BWVar	1 month	19.4◑	75.0●	−0.95○	189◑	−4.76	151◑	21.2●	62.9●	9.34○	127●	11.1●	85.7●
BWVar	6 months	17.5○	90.4●	3.84◑	206◑	−0.03	168●	21.2●	63.6●	7.90	132●	12.3●	88.9●
BWVar	12 months	13.0	89.3●	−6.76	182○	0.14	180●	19.8●	63.4●	5.39	133●	10.6◑	89.9●
BWVar	24 months	20.8◑	101●	1.00○	190○	−0.89	190●	21.7●	62.8●	7.48	133●	9.12○	89.0●
MaxPC	1 month	16.7◑	8.82●	7.52●	170○	0.52○	222●	16.4●	35.5●	9.04●	119●	9.72●	91.0●
MaxPC	6 months	8.48	24.5●	6.84●	173○	7.58◑	229●	14.8●	37.0●	6.74○	119●	10.7●	90.1●
MaxPC	12 months	8.80	18.3●	3.67◑	169○	9.62●	237●	14.7●	36.9●	4.67	119●	9.81●	91.7●
MaxPC	24 months	14.6○	14.8●	9.20●	168○	9.07◑	238●	17.6●	36.8●	7.01○	120●	9.75●	94.9●
MinPC	1 month	25.7●	92.8●	−8.43	86.8	17.0●	88.9	24.2●	66.2●	−1.14	108●	22.1●	49.2●
MinPC	6 months	30.9●	105●	−7.68	89.2	27.8●	98.3	25.4●	67.8●	−2.59	105●	22.9●	52.6●
MinPC	12 months	27.4●	112●	−6.23	115	17.9◑	94.1	24.9●	67.7●	−2.55	109●	22.8●	50.5●
MinPC	24 months	20.8○	106●	−4.26	101	22.3●	106○	22.8●	67.9●	2.32	107●	24.4●	50.7●
Best θ^2_N	1 month	−40.2●	−541●	−50.8●	−99.4●	−73.1●	−104●	4.40	29.1●	−9.28●	79.2●	−3.92	33.3◑
Best θ^2_N	6 months	4.06	−47.3	−3.26	159	−1.21	102	19.6●	52.5●	5.82	120●	16.6●	72.9●
Best θ^2_N	12 months	3.77	14.7●	−8.77	182○	−7.87	137◑	20.7●	55.1●	0.49	119●	17.6●	73.2●
Best θ^2_N	24 months	14.6	50.1●	−2.44	206◑	−2.04	143◑	21.7●	56.0●	4.13	134●	20.7●	87.4●
MaxW	1 month	−307●	−798●	−350●	−740●	−300●	−543●	30.3●	79.2●	9.19	199●	17.9○	130●
MaxW	6 months	−171●	−421●	−234●	−749●	−151●	−456●	29.2●	79.4●	3.28	203●	19.4○	131●
MaxW	12 months	−127●	−305●	−179●	−581●	−129●	−366●	32.3●	79.0●	2.19	197●	19.8○	129●
MaxW	24 months	−97.6●	−74.8	−115●	−416●	−56.3○	−244●	32.2●	77.4●	3.59	195●	16.0	126●

Notes. This table reports the annualized net out-of-sample utility, in percentage points, for the 2F portfolio implemented with the 11 asset selection rules in Table 3, across the six 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 construct the portfolios either with the sample or the linear shrinkage covariance matrix of Ledoit and Wolf (2004) and rebalance the portfolio weights every month. The net out-of-sample utility is computed using rolling windows, a sample size $T = 120$ months, and proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$. We compute two-sided p-values for the statistical test of the difference between the utility of the 2F portfolio under the 'All' selection rule versus the 10 other selection rules, for all selection frequencies, using the block bootstrap methodology described in Section 5.4. The symbols ○, ◑, and ● indicate that the p-value is less than 10%, 5%, and 1%, respectively.

means that the selected assets are indeed dissimilar.

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

Proposition SM.6 *Consider two independent selection rules that randomly select N_t assets out of M available ones. Then, the expected overlap between the two rules is*

$$\bar{N}_{t,M} = M \frac{\binom{M-1}{N_t-1}^2}{\binom{M}{N_t}^2}. \tag{SM47}$$

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

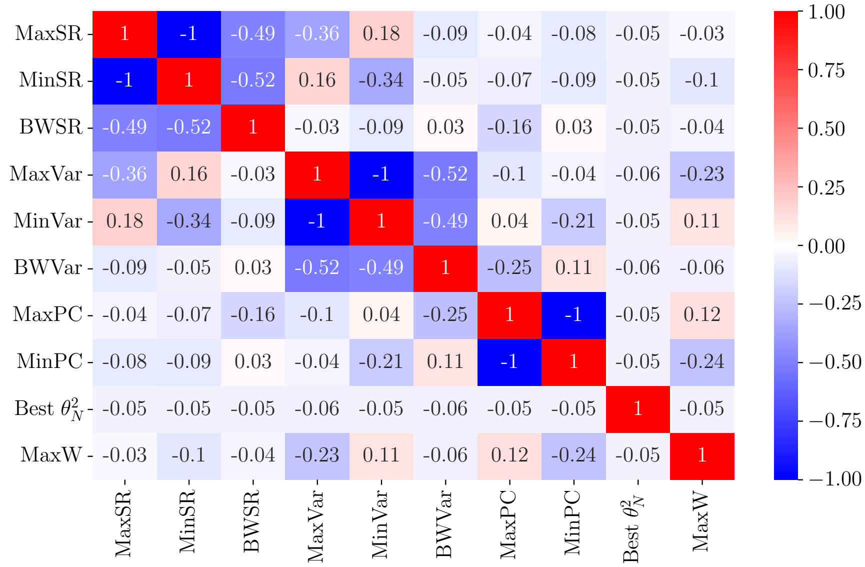
Recall that our objective is to identify how a pair of selection rules (k, l) selects assets relative to two independent random selection rules. Therefore, we use the expected overlap in (SM47) to build a measure of the *excess overlap* of any (k, l) pair relative to that of a pair of independent random selection rules. We denote this measure $\mathrm{eo}_{k,l,t}$ and define it as

$$\mathrm{eo}_{k,l,t} = (\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} - \bar{N}_{t,M}) \times \left(\frac{\mathbb{1}_{\{\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} - \bar{N}_{t,M} \geq 0\}}}{(N_t - \bar{N}_{t,M})} - \frac{\mathbb{1}_{\{\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} - \bar{N}_{t,M} < 0\}}}{(\max(0, 2N_t - M) - \bar{N}_{t,M})} \right). \tag{SM48}$$

Several comments are in order. First, if $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \bar{N}_{t,M}$, then $\mathrm{eo}_{k,l,t} = 0$. 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., $\mathrm{eo}_{k,l,t} > 0$ if $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} > \bar{N}_{t,M}$ and vice-versa. Third, the measure of excess overlap is bounded between -1 and 1. A value of 1 indicates the *maximum* overlap and is obtained if and only if k and l are *identical*, i.e., $\boldsymbol{y}_{k,t} = \boldsymbol{y}_{l,t}$ such that $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = N_t$. A value of -1 indicates the *minimum* overlap, i.e., $\boldsymbol{y}'_{k,t}\boldsymbol{y}_{l,t} = \max(0, 2N_t - M)$.[SM6]

[SM6] If $M \geq 2N_t$, then it is possible for two rules to have zero overlap as each rule can select N_t different assets. However, if $M < 2N_t$, then the minimum possible overlap will be strictly positive and will be of $2N_t - M$ assets. For instance, if $M = 90$ and $N_t = 50$, two rules must have an overlap of at least $2N_t - M = 10$ assets.

Figure SM.8: Average excess overlap between selection rules



Notes. This figure depicts the average excess overlap, defined in (SM48), for each pair of selection rule listed in Table 3 (we exclude the 'All' rule). The excess overlap is averaged over the 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.

In Figure SM.8, we depict a heatmap of the excess overlap in (SM48) for all selection rules pairs (k, l) in Table 3. We average it across all months t and across the six datasets described in Table 1. We only consider the optimal portfolio size of the two-fund rule for conciseness, but the results are similar for the three-fund rules. Moreover, because selection rules are based on sample estimates, we report results under the sample covariance matrix.

Two main observations can be made from Figure SM.8. First, the results match our expectations. As expected, opposite rules (MaxSR and MinSR, MaxVar and MinVar, MaxPC and MinPC) have an excess overlap of -1. Also, the sign of the excess overlap is coherent. For instance, the average excess overlap is positive between MaxVar and MinSR (0.16) because both rules tend to select assets with high marginal variance, while it is negative between MaxVar and MaxSR (-0.36) because they tend to select different assets.

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 assets differently.

E Proofs of all results

Proof of Proposition 1

Denoting $\boldsymbol{D}_N = \text{diag}(\sigma_1, \ldots, \sigma_N)$, $\boldsymbol{\Sigma}_N = \boldsymbol{D}_N \boldsymbol{P}_N(\rho) \boldsymbol{D}_N$ and its inverse is given by

$$\boldsymbol{\Sigma}_N^{-1} = \boldsymbol{D}_N^{-1} \boldsymbol{P}_N(\rho)^{-1} \boldsymbol{D}_N^{-1}, \tag{SM49}$$

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

$$\boldsymbol{P}_N(\rho)^{-1} = \frac{1}{1-\rho} \left(\boldsymbol{I}_N - \frac{\rho}{1-\rho+N\rho} \mathbf{1}_N \mathbf{1}_N' \right). \tag{SM50}$$

Combining (SM49) and (SM50) yields

$$\boldsymbol{\Sigma}_N^{-1} = \frac{1}{1-\rho} \left(\boldsymbol{D}_N^{-2} - \frac{\rho}{1-\rho+N\rho} \boldsymbol{D}_N^{-1} \mathbf{1}_N \mathbf{1}_N' \boldsymbol{D}_N^{-1} \right), \tag{SM51}$$

and thus the maximum utility becomes

$$U(\boldsymbol{w}^\star) = \frac{\boldsymbol{\mu}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\mu}_N}{2\gamma} = \frac{1}{2\gamma(1-\rho)} \left[\boldsymbol{\mu}_N' \boldsymbol{D}_N^{-2} \boldsymbol{\mu}_N - \frac{\rho}{1-\rho+N\rho} (\boldsymbol{\mu}_N' \boldsymbol{D}_N^{-1} \mathbf{1}_N)^2 \right], \tag{SM52}$$

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

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

$$\begin{aligned} &(1-\rho)(\theta_{N+1}^2 - \theta_N^2) \\ &= \left[\sum_{i=1}^{N+1} s_i^2 - \frac{\rho}{1-\rho+(N+1)\rho} \left(\sum_{i=1}^{N+1} s_i \right)^2 \right] - \left[\sum_{i=1}^{N} s_i^2 - \frac{\rho}{1-\rho+N\rho} \left(\sum_{i=1}^{N} s_i \right)^2 \right] \end{aligned} \tag{SM53}$$

Decomposing $\left(\sum_{i=1}^{N+1} s_i\right)^2$ as $\left(\sum_{i=1}^{N} s_i\right)^2 + s_{N+1}^2 + 2s_{N+1}\sum_{i=1}^{N} s_i$, (SM53) becomes

$$
\begin{aligned}
(1-\rho)(\theta_{N+1}^2 - \theta_N^2) &= \frac{1-\rho+N\rho}{1-\rho+(N+1)\rho} s_{N+1}^2 + \frac{\rho^2\left(\sum_{i=1}^{N} s_i\right)^2}{(1-\rho+N\rho)(1-\rho+(N+1)\rho)} \\
&\quad - \frac{2\rho s_{N+1}\sum_{i=1}^{N} s_i}{1-\rho+(N+1)\rho} \\
&= \frac{\left[\rho N\bar{\theta}_{N,1} - (1-\rho+N\rho)s_{N+1}\right]^2}{(1-\rho+N\rho)(1-\rho+(N+1)\rho)},
\end{aligned} \tag{SM54}
$$

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

Proof of Proposition 2

Equation (11) is a direct extension of Kan and Lassance (2023, Proposition 7) for a general α instead of $\alpha = 1$. It is then easy to show that the α maximizing (11) is equal to (16), which also corresponds to Kan and Lassance (2023, Equation (50)). Finally, after some developments, plugging (16) into (11) yields Equation (17), which concludes the proof.

Proof of Corollary 1

Under Assumption 1, the maximum squared Sharpe ratio θ_N^2 is given by (4). Plugging it into (11) yields the EU of the SMV portfolio $\hat{\boldsymbol{w}}^\star$ in Equation (18), and plugging it into (17) yields the EU of the optimal two-fund rule $\hat{\boldsymbol{w}}(\alpha^\star)$ in (20).

Proof of Proposition SM.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 $\hat{\boldsymbol{w}}(\boldsymbol{\alpha})$ in (SM3) is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\alpha})) = \frac{1}{2\gamma}\left(2\alpha_1\tilde{\mu}_1 + 2\alpha_2\tilde{\mu}_2 - \alpha_1^2\tilde{\sigma}_1^2 - \alpha_2^2\tilde{\sigma}_2^2 - 2\alpha_1\alpha_2\tilde{\sigma}_{12}\right). \tag{SM55}$$

Kan and Lassance (2023, Equations (IA80)–(IA84)) show that

$$\tilde{\mu}_1 = \frac{\kappa_{N,1}\theta_N^2 T}{T-N-2}, \tag{SM56}$$

$$\tilde{\mu}_2 = \frac{\kappa_{N,1}\lambda_{g,N} T}{T-N-2}, \tag{SM57}$$

$$\tilde{\sigma}_1^2 = \frac{c_N T^2}{(T-N-2)^2}\left(\kappa_{N,2}\theta_N^2 + \kappa_{N,3}\frac{N}{T}\right), \tag{SM58}$$

$$\tilde{\sigma}_2^2 = \frac{c_N \kappa_{N,2} T^2}{\sigma_{g,N}^2 (T-N-2)^2}, \tag{SM59}$$

$$\tilde{\sigma}_{12} = \frac{c_N \kappa_{N,2}\lambda_{g,N} T^2}{(T-N-2)^2}, \tag{SM60}$$

and plugging them into (SM55) yields the desired result in Equation (SM4). It is then easy to show that the $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ maximizing (SM4) is equal to (SM5), which also corresponds to Kan and Lassance (2023, Equations (51)–(52)). Finally, after some developments, plugging (SM5) into (SM4) yields Equation (SM6), which concludes the proof.

Proof of Proposition SM.2

The squared Sharpe ratio of the GMV portfolio depends on $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ as

$$\theta_{g,N}^2 = \frac{(\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N)^2}{\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N}. \tag{SM61}$$

Now, under Assumption 1, $\boldsymbol{\Sigma}_N^{-1}$ is given by (SM51), and thus,

$$\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N = \frac{N}{1-\rho}\left(\bar{\lambda}_N - \frac{N\rho}{1-\rho+N\rho}\bar{\theta}_{N,1}\bar{\sigma}_{N,-1}\right), \tag{SM62}$$

$$\mathbf{1}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{1}_N = \frac{N}{1-\rho}\left(\bar{\sigma}_{N,-2} - \frac{N\rho}{1-\rho+N\rho}\bar{\sigma}_{N,-1}^2\right). \tag{SM63}$$

Plugging (SM62)–(SM63) into (SM61) yields the desired result in Equation (SM7), which concludes the proof.

Proof of Proposition SM.3

The EU of the EW-three-fund rule $\hat{\boldsymbol{w}}(\boldsymbol{\beta})$ in (SM14) is

$$EU(\hat{\boldsymbol{w}}(\boldsymbol{\beta})) = \frac{1}{2\gamma}\left(2\beta_1\tilde{\mu}_1 + 2\beta_2\mu_{ew,N} - \beta_1^2\tilde{\sigma}_1^2 - \beta_2^2\sigma_{ew,N}^2 - 2\mathbb{E}[\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\boldsymbol{w}_{ew}]\right), \tag{SM64}$$

where $\tilde{\mu}_1$ is given by (SM56), $\tilde{\sigma}_1^2$ by (SM58), and $\mathbb{E}[\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\boldsymbol{w}_{ew}] = \frac{\kappa_{N,1}\mu_{ew,N}T}{T-N-2}$, which yields the desired result in Equation (SM15). It is then easy to show that the $\boldsymbol{\beta} = (\beta_1, \beta_2)$ maximizing (SM15) is equal to (SM16). Finally, after some developments, plugging (SM16) into (SM15) yields Equation (SM17), which concludes the proof.

Proof of Proposition SM.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,

$$\sigma_{ew,N}^2 = \frac{1}{N^2}\mathbf{1}_N'\boldsymbol{\Sigma}\mathbf{1}_N = \frac{1}{N}\left(\bar{\sigma}_{N,2} + \frac{\rho}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_i\sigma_j\right). \tag{SM65}$$

This yields the desired result in Equation (SM18) after noticing that

$$\frac{\rho}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sigma_i\sigma_j = \rho N\bar{\sigma}_{N,1}^2 - \rho\bar{\sigma}_{N,2}, \tag{SM66}$$

which concludes the proof.

Proof of Proposition SM.5

Throughout the proof, we denote

$$f(N,\rho) := \frac{N\rho}{1-\rho+N\rho}. \tag{SM67}$$

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

$$\mathbb{E}[\hat{\delta}_N] = \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[\hat{s}_i^2] - \frac{f(N,\rho)}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^{M}\hat{s}_i\right)^2\right]. \tag{SM68}$$

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

$$\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[\hat{s}_i^2] = \bar{\theta}_{M,2} + \frac{1}{T}. \tag{SM69}$$

Moreover,

$$\frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^{M}\hat{s}_i\right)^2\right] = \bar{\theta}_{M,1}^2 + \frac{1}{M^2}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{M}\hat{s}_i\right], \tag{SM70}$$

where

$$\frac{1}{M^2}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{M}\hat{s}_i\right] = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbb{C}\mathrm{ov}[\hat{s}_i,\hat{s}_j] = \frac{1}{T}\left(\frac{1}{M} + \frac{M-1}{M}\rho\right). \tag{SM71}$$

Plugging (SM69)–(SM71) into (SM68) yields

$$\mathbb{E}[\hat{\delta}_N] = \delta_N + \frac{1}{T}\left(1 - \frac{f(N,\rho)(1-\rho+M\rho)}{M}\right), \tag{SM72}$$

which corresponds to (SM36), as desired.

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

$$\mathbb{E}[\hat{r}_{g,N}] = \frac{\mathbb{E}\left[\left(\frac{1}{M}\sum_{i=1}^{M}\hat{\lambda}_i - \frac{f(N,\rho)}{M}\bar{\sigma}_{M,-1}\sum_{i=1}^{M}\hat{s}_i\right)^2\right]}{\bar{\sigma}_{M,-2} - f(N,\rho)\bar{\sigma}_{M,-1}^2}. \tag{SM73}$$

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

$$\mathbb{E}[\hat{r}_{g,N}] = r_{g,N} +$$

$$\frac{1}{M^2}\frac{\mathbb{V}\text{ar}\left[\sum_{i=1}^M \hat{\lambda}_i\right] + f(N,\rho)^2\bar{\sigma}_{M,-1}^2\mathbb{V}\text{ar}\left[\sum_{i=1}^M \hat{s}_i\right] - 2f(N,\rho)\bar{\sigma}_{M,-1}\mathbb{C}\text{ov}\left[\sum_{i=1}^M \hat{\lambda}_i, \sum_{i=1}^M \hat{s}_i\right]}{\bar{\sigma}_{M,-2} - f(N,\rho)\bar{\sigma}_{M,-1}^2}, \tag{SM74}$$

where $\frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^M \hat{s}_i\right]$ is given by (SM71) and

$$\frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^M \hat{\lambda}_i\right] = \frac{1}{M^2}\sum_{i=1}^M\sum_{j=1}^M \mathbb{C}\text{ov}[\hat{\lambda}_i, \hat{\lambda}_j] = \frac{1}{T}\left(\frac{1-\rho}{M}\bar{\sigma}_{M,-2} + \rho\bar{\sigma}_{M,-1}^2\right), \tag{SM75}$$

$$\frac{1}{M^2}\mathbb{C}\text{ov}\left[\sum_{i=1}^M \hat{\lambda}_i, \sum_{i=1}^M \hat{s}_i\right] = \frac{1}{M^2}\sum_{i=1}^M\sum_{j=1}^M \mathbb{C}\text{ov}[\hat{\lambda}_i, \hat{s}_j] = \frac{1-\rho+M\rho}{MT}\bar{\sigma}_{M,-1}. \tag{SM76}$$

Plugging (SM71) and (SM75)–(SM76) into (SM74) yields

$$\mathbb{E}[\hat{r}_{g,N}] = r_{g,N} + \frac{1}{T}\frac{\frac{1-\rho}{M}\bar{\sigma}_{M,-2} + \left(\rho + \frac{f(N,\rho)^2(1-\rho+M\rho)}{M} - \frac{2f(N,\rho)(1-\rho+M\rho)}{M}\right)\bar{\sigma}_{M,-1}^2}{\bar{\sigma}_{M,-2} - f(N,\rho)\bar{\sigma}_{M,-1}^2}, \tag{SM77}$$

which corresponds to (SM37), as desired.

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

$$\mathbb{E}[\hat{r}_{ew,N}] = \frac{(1-\rho)}{(1-\rho)\bar{\sigma}_{M,2} + \rho N\bar{\sigma}_{M,1}^2}\frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^M \hat{\mu}_i\right)^2\right]. \tag{SM78}$$

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

$$\frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=1}^M \hat{\mu}_i\right)^2\right] = \bar{\mu}_M^2 + \frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^M \hat{\mu}_i\right]. \tag{SM79}$$

Moreover,

$$\frac{1}{M^2}\mathbb{V}\text{ar}\left[\sum_{i=1}^M \hat{\mu}_i\right] = \frac{1}{M^2}\sum_{i=1}^M\sum_{j=1}^M \mathbb{C}\text{ov}[\hat{\mu}_i, \hat{\mu}_j] = \frac{(1-\rho)\bar{\sigma}_{M,2} + \rho M\bar{\sigma}_{M,1}^2}{MT}. \tag{SM80}$$

Plugging (SM79)–(SM80) into (SM78) yields

$$\mathbb{E}[\hat{r}_{ew,N}] = r_{ew,N} + \frac{1-\rho}{MT} \times \frac{(1-\rho)\bar{\sigma}_{M,2} + \rho M\bar{\sigma}_{M,1}^2}{(1-\rho)\bar{\sigma}_{M,2} + \rho N\bar{\sigma}_{M,1}^2}, \tag{SM81}$$

which corresponds to (SM38) and concludes the proof.

Proof of Proposition SM.6

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

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

$$\begin{aligned}
\bar{N}_{t,M} &= \mathbb{E}\left[\sum_{i=1}^{M} y_{i,k,t}y_{i,l,t} \middle| \sum_{i=1}^{M} y_{i,k,t} = \sum_{i=1}^{M} y_{i,l,t} = N_t\right] \\
&= \sum_{i=1}^{M} \mathbb{P}\left[y_{i,k,t} = 1 \left| \boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right.\right] \times \mathbb{P}[y_{i,l,t} = 1 | \boldsymbol{y}'_{l,t}\mathbf{1}_M = N_t]. \qquad \text{(SM82)}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}} \left| \boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right.\right] &= \frac{\mathbb{P}\left[\boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t \middle| \boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}}\right] \times \mathbb{P}\left[\boldsymbol{y}_{k,t} = \tilde{\boldsymbol{y}}\right]}{\mathbb{P}\left[\boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right]} \\
&= \frac{\mathbb{1}_{\{n=N_t\}} \times p^n(1-p)^{M-n}}{\binom{M}{N_t} p^{N_t}(1-p)^{M-N_t}} = \frac{\mathbb{1}_{\{n=N_t\}}}{\binom{M}{N_t}} \qquad \text{(SM83)}
\end{aligned}$$

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

$$\mathbb{P}\left[y_{i,k,t} = 1 \left| \boldsymbol{y}'_{k,t}\mathbf{1}_M = N_t\right.\right] = \mathbb{P}\left[y_{i,k,t} = 1 \middle| \sum_{j \neq i} y_{j,k,t} = N_t - 1\right]$$

$$
\begin{aligned}
&= \sum_{h \neq i} \frac{\mathbb{1}_{\{\sum_{j \neq h} b_j = N_t - 1\}}}{\binom{M}{N_t}} \\
&= \frac{\binom{M-1}{N_t-1}}{\binom{M}{N_t}}. \qquad \text{(SM84)}
\end{aligned}
$$

Plugging (SM84) into (SM82) yields the desired result and concludes the proof.

References in Supplementary Material

DeMiguel, V., Martín-Utrera, A., and Nogales, F. (2015), "Parameter uncertainty in multiperiod portfolio optimization with transaction costs." *Journal of Financial and Quantitative Analysis*, 50(6):1443–1471.

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

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

Kan, R. and Lassance, N. (2023), "Optimal portfolio choice with fat tails and parameter uncertainty." *SSRN working paper*.

Kan, R. and Wang, X. (2023), "Optimal portfolio choice with unknown benchmark efficiency." *Management Science*, forthcoming.

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

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

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

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

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

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

Yuan, M. and Zhou, G. (2024), "Why naïve 1/N diversification is not so naïve, and how to beat it?" *Journal of Financial and Quantitative Analysis*, forthcoming.