

MAXIMIZING THE OUT-OF-SAMPLE SHARPE RATIO

Nathan Lassance

LIDAM Discussion Paper LFIN
2021 / 13

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>

Maximizing the Out-of-Sample Sharpe Ratio*

Nathan Lassance

LFIN/LIDAM, UCLouvain, nathan.lassance@uclouvain.be

December 23, 2021

Maximizing the out-of-sample Sharpe ratio is an important objective for investors. To achieve this, we characterize optimal portfolio combinations maximizing expected out-of-sample Sharpe ratio. When investing in the risk-free asset is allowed and combination coefficients are unconstrained, as in Kan and Zhou (2007), we uncover that combining portfolios to maximize expected out-of-sample utility optimizes expected out-of-sample Sharpe ratio as well. However, the two criteria are not equivalent for portfolios fully invested in risky assets, and in this case we show how to adapt the optimal portfolio combination of Kan, Wang, and Zhou (2021) to maximize expected out-of-sample Sharpe ratio. We find that the proposed mean-variance portfolio combinations calibrated to maximize expected out-of-sample Sharpe ratio generally outperform the considered benchmarks. Relative to the minimum-variance portfolio estimated with a nonlinearly shrunk covariance matrix, the annualized out-of-sample Sharpe ratio increases from 0.988 to 1.110 before transaction costs, and from 0.914 to 1.007 net of transaction costs, on average across four typical datasets.

Key words: mean-variance portfolio, parameter uncertainty, estimation risk, out-of-sample performance.

1. Introduction

When looking for investment strategies that are likely to perform well in the future, the literature warns investors not to blindly rely on historical estimates of return inputs. In particular, optimal mean-variance portfolios estimated using sample estimates of the mean and covariance matrix of asset returns deliver poor out-of-sample performance; see, e.g., Jagannathan and Ma (2003), DeMiguel, Garlappi, and Uppal (2009b), and Barroso and Saxena (2021).¹ Any hope for robust out-of-sample performance is not lost however, because over the years the literature has also developed many effective methods to alleviate the impact of parameter uncertainty on investment performance. In this paper, we contribute to the analytical framework pioneered by Kan and Zhou (2007) who characterize the expected out-of-sample utility of sample mean-variance portfolios under the parametric assumption of iid normal returns, and design a three-fund rule optimizing this quantity. Following Kan and Zhou (2007), many papers exploit expected out-of-sample utility as a robust

* I am very grateful to Victor DeMiguel, Raymond Kan, Felix Kircher, Alberto Martín-Utrera, Majeed Simaan, Xiaolu Wang, and Guofu Zhou for their feedback. This work was supported by the Fonds de la Recherche Scientifique (FNRS) under Grant number J.0115.22.

¹ We focus on mean-variance portfolios in this paper, which are analytically more tractable and less prone to estimation risk than higher-moment strategies. Nonetheless, a large literature exists that incorporates higher moments in portfolio choice; see Martellini and Ziemann (2010) and Lassance, DeMiguel, and Vrins (2021a) for reviews.

portfolio choice criterion; see Zhou (2008), DeMiguel, Garlappi, and Uppal (2009b), Frahm and Memmel (2010), Tu and Zhou (2011), DeMiguel, Martin-Utrera, and Nogales (2013), DeMiguel, Martin-Utrera, and Nogales (2015), Branger, Lučivjanská, and Weissensteiner (2019), Kan, Wang, and Zhou (2021b), Kan and Wang (2021), Lassance, Martin-Utrera, and Simaan (2021b), and Bodnar, Okhrin, and Parolya (2021). A recent fundamental contribution among these papers is Kan, Wang, and Zhou (2021b) who propose a methodology to maximize expected out-of-sample utility in the common setting with portfolios fully invested in risky assets.

It may seem surprising that this literature focuses on maximizing expected out-of-sample *utility*, when another commonly used portfolio performance criterion is the *Sharpe ratio* (Sharpe 1966, 1994). Actually, we can identify three good reasons for that choice. First, according to DeMiguel, Garlappi, and Uppal (2009b, p. 1928), mean-variance utility has a practically useful interpretation of being the certainty-equivalent return, i.e., “*the risk-free rate that an investor is willing to accept rather than adopting a particular risky portfolio strategy.*” Second, when investing in the risk-free asset is allowed, all portfolios on the efficient frontier have equal out-of-sample Sharpe ratio, and thus the utility criterion is needed to identify a single optimal portfolio. Third, while the out-of-sample Sharpe ratio is a *ratio* between two random variables, the out-of-sample utility is a *difference* between two random variables, which makes its expectation easier to evaluate. This third reason is the one put forward by Tu and Zhou (2011, p.208): “*In this paper, we focus on maximizing the expected utility as it is easier to solve than maximizing the Sharpe ratio because the latter is to maximize a highly nonlinear function of the portfolio weights and there are no closed-form solutions available in the presence of estimation errors.*”

Notwithstanding the previously identified reasons for relying on expected out-of-sample utility as a criterion in the presence of estimation errors, the Sharpe ratio is another very important criterion for investors. As Kelly, Malamud, and Zhou (2021) put it: “*Academics and real-world investors rely nearly universally on the unconditional Sharpe ratio when evaluating empirical trading strategies.*” Therefore, in this paper, we complement the literature and derive closed-form solutions for optimal portfolio combinations maximizing expected out-of-sample Sharpe ratio under iid normal returns.² We show that the analytical challenge identified above by Tu and Zhou (2011) can be bypassed because a simple approximation to the expected out-of-sample Sharpe ratio provides very accurate results and delivers closed-form optimal portfolio solutions. We find in particular that when a risk-free asset is available and combination coefficients are unconstrained, the expected out-of-sample Sharpe ratio is not a discriminant criterion, and maximizing expected out-of-sample utility

² While we exploit the framework introduced by Kan and Zhou (2007) to optimize the out-of-sample Sharpe ratio, other approaches exist in the literature. Two recent examples in the setting with a risk-free asset are Ao, Li, and Zheng (2019) who use high-dimensional sparse regression techniques, and Rapponi, Uppal, and Zaffaroni (2021) who use estimation methods that are robust to model misspecification.

optimizes expected out-of-sample Sharpe ratio as well. However, for portfolios fully invested in risky assets, the two criteria are not equivalent and we show how to adapt the optimal portfolio combination of Kan, Wang, and Zhou (2021b) to maximize expected out-of-sample Sharpe ratio.

We are not the first to propose robust investment strategies maximizing expected out-of-sample Sharpe ratio.³ However, to the best of our knowledge, we are the first to provide analytical expressions for the resulting optimal portfolios.⁴ Closed-form expressions are important because they allow for precise theoretical insights that one cannot obtain with data-driven approaches. Moreover, they have the additional benefit of delivering computationally efficient solutions.

We make several contributions to the literature on optimal portfolio choice in the presence of parameter uncertainty, which we organize in five main sections. First, in Section 3, we introduce a Taylor-series approximation to the expected out-of-sample Sharpe ratio that is more accurate than the approximation proposed by DeMiguel, Martin-Utrera, and Nogales (2013). Focusing on sample mean-variance portfolios, we derive a closed-form expression for the novel approximation and find it is very accurate. Moreover, we validate the approximation put forward by DeMiguel, Martin-Utrera, and Nogales (2013) because we show that it barely differs from the refined approximation and delivers a similar optimal combination coefficient. Therefore, we use their approximation throughout the paper, which allows for analytical insights.

Second, in Section 4, we consider the setting in which a risk-free investment is allowed. We show that the mathematics of combining portfolios to trade off between expected out-of-sample mean return and variance is similar to the mathematics of classical portfolio theory. In particular, when combination coefficients are unconstrained, the expected out-of-sample Sharpe ratio cannot identify a unique optimal portfolio combination and is automatically maximized when optimizing expected out-of-sample utility. We illustrate this result for the popular three-fund rule proposed by Kan and Zhou (2007) that combines the risk-free asset, the tangent portfolio, and the GMV portfolio to maximize expected out-of-sample utility.⁵

³ DeMiguel, Martin-Utrera, and Nogales (2013) propose several criteria to calibrate shrinkage portfolios including the expected out-of-sample Sharpe ratio. However, they do not evaluate the accuracy of the approximation to the expected out-of-sample Sharpe ratio they propose, and they do not provide analytical solutions for the optimal portfolios. Yuan and Zhou (2021) derive the combination of the sample mean-variance and equally weighted portfolios that maximizes out-of-sample Sharpe ratio in the high-dimensional regime, which contrasts with our finite-sample setting. Chakrabarti (2021) and Kircher and Rosch (2021) study optimal mean-variance portfolios maximizing expected out-of-sample Sharpe ratio via data-driven approaches based on robust optimization and bootstrap, respectively, which differ from the analytical parametric approach introduced by Kan and Zhou (2007).

⁴ Kan, Wang, and Zheng (2021a) derive the distribution (and the expectation) of out-of-sample Sharpe ratio of sample efficient portfolios in the presence of a risk-free asset. However, they use this result to explain the poor out-of-sample performance of multi-factor asset-pricing models, not to optimally combine portfolios.

⁵ Tu and Zhou (2011) combine the risk-free asset, the tangent portfolio, and the equally weighted portfolio to maximize expected out-of-sample utility. However, they constrain the combination coefficients on the tangent and equally weighted portfolios to sum to one, and thus their portfolio combination does not maximize expected out-of-sample Sharpe ratio. The results for this portfolio combination are available in Section EC.3 of the electronic companion.

Third, in Section 5, we turn to the setting with portfolios fully invested in risky assets. In that case, the combination coefficients have to sum to one, and thus the expected out-of-sample Sharpe ratio yields a unique optimal portfolio combination, which is not equivalent to that maximizing expected out-of-sample utility. As in Kan, Wang, and Zhou (2021b), we focus on combining fully invested mean-variance portfolios, and derive an analytical expression for the combination coefficient maximizing expected out-of-sample Sharpe ratio. When the covariance matrix is known, the optimal intensity with which the sample optimal portfolio is shrunk toward the GMV portfolio is the same under the expected out-of-sample utility and expected out-of-sample Sharpe ratio criteria. However, this equivalence no longer holds when the covariance matrix is unknown, in which case the expected out-of-sample Sharpe ratio implies a smaller shrinkage toward the GMV portfolio. Therefore, a naive application of the results in Kan, Wang, and Zhou (2021b), which are derived for the utility criterion, does not deliver an optimal expected Sharpe ratio out of sample.

Fourth, in Section 6, we follow Lassance, Martin-Utrera, and Simaan (2021b) and study the portfolio combination that delivers an optimal tradeoff between the expectation and standard deviation of out-of-sample Sharpe ratio, which allows us to alleviate estimation errors in combination coefficients. We characterize an approximation to the standard deviation of out-of-sample Sharpe ratio and show that penalizing this standard deviation is equivalent to shrinking the covariance matrix of estimated portfolio weights toward another matrix, which is similar to the equivalence between norm constraints and shrinkage in DeMiguel, Garlappi, Nogales, and Uppal (2009a). For the case of fully invested mean-variance portfolios, this penalization acts as an additional source of shrinkage because it is reduced by shrinking the tangent portfolio toward the GMV portfolio more intensively. This “double shrinkage” is reminiscent of Candelon, Hurlin, and Tokpavi (2012) and we find that it is useful to further improve out-of-sample performance for realistic sample sizes.

Finally, in Section 7, we evaluate the out-of-sample performance of the optimal portfolios introduced in the paper in the setting with fully invested portfolios. We first consider simulated normal data and show that, in most cases, the portfolio combination maximizing expected out-of-sample Sharpe ratio outperforms several benchmarks: the equally weighted portfolio, the reward-to-risk timing portfolio of Kirby and Ostdiek (2012), the GMV portfolio, the tangent portfolio, and the portfolio combination maximizing expected out-of-sample utility as in Kan, Wang, and Zhou (2021b). This outperformance holds when considering both the sample covariance matrix and the nonlinearly shrunk covariance matrix of Ledoit and Wolf (2020). Moreover, for realistic values of the sample size, the robust portfolio combination that accounts for the standard deviation of out-of-sample Sharpe ratio outperforms that maximizing only its expectation. We turn next to real data and find that our approach delivers larger out-of-sample Sharpe ratios than the GMV portfolio in most cases, even when it is estimated via the nonlinear shrinkage estimator of the covariance

matrix of Ledoit and Wolf (2020). Notably, it outperforms for the dataset of 30 industry portfolios and a sample size of 10 years, which is a challenging setting because, as noted by Kirby and Ost-diek (2012), industry portfolios do not display significant cross-sectional variation in mean returns. Moreover, our approach delivers larger Sharpe ratios than the aforementioned benchmarks and than the portfolio combination of Kircher and Rosch (2021) that maximizes expected out-of-sample Sharpe ratio via bootstrap. In particular, although bootstrap generates good Sharpe ratios, our analytical approach delivers portfolio combinations that are less exposed to the tangent portfolio, have lower turnover, and outperform in most cases, while being computationally more efficient.

The proofs of all results in the paper are available in Section EC.7 of the electronic companion.

2. Notation, assumptions, and definitions

We consider an investment universe of $N \geq 2$ assets. At the end of period t , the investor cares about her next-period portfolio (excess) return, $w'R_{t+1}$, where w is the vector of portfolio weights on the risky assets and R_{t+1} is the random vector of asset excess returns. We assume for tractability that next-period returns are multivariate normal, $R_{t+1} \sim \mathcal{N}(\mu, \Sigma)$, and the covariance matrix Σ is positive definite, which are common assumptions in the literature (Kan and Zhou 2007, Kan and Smith 2008, Ao, Li, and Zheng 2019). Given a chosen portfolio w , a quantity $w'\iota$ is invested in risky assets and $1 - w'\iota$ in the risk-free asset, where ι is an N -dimensional vector of ones. When the portfolio is fully invested in risky assets as in Section 5, the weights sum to one, $w'\iota = 1$.

Because parameters are unknown, we estimate μ and Σ at time t using a sample of $T > N$ historical return observations, $(r_{t-T+1}, \dots, r_t)$, assumed iid, via maximum likelihood:

$$\hat{\mu} = \frac{1}{T} \sum_{m=1}^T r_{t-T+m} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{T} \sum_{m=1}^T (r_{t-T+m} - \hat{\mu})(r_{t-T+m} - \hat{\mu})'. \quad (1)$$

The mean-return vector μ is always unknown and estimated by $\hat{\mu}$, while we consider both cases in which the covariance matrix Σ is assumed known or estimated by $\hat{\Sigma}$.

It is well known that, without parameter uncertainty, the maximum Sharpe ratio is

$$\theta = \sqrt{\mu' \Sigma^{-1} \mu}. \quad (2)$$

Moreover, the global-minimum-variance (GMV) portfolio is

$$w_g = \Sigma^{-1} \iota (\iota' \Sigma^{-1} \iota)^{-1}. \quad (3)$$

Finally, for further references, we define the return mean, variance, and Sharpe ratio of the GMV portfolio, and the difference between the maximum square Sharpe ratio and that of GMV, as

$$\mu_g = \iota' \Sigma^{-1} \mu (\iota' \Sigma^{-1} \iota)^{-1}, \quad \sigma_g^2 = (\iota' \Sigma^{-1} \iota)^{-1}, \quad \theta_g = \mu_g / \sigma_g, \quad \text{and} \quad \psi^2 = \theta^2 - \theta_g^2. \quad (4)$$

3. Approximating the expected out-of-sample Sharpe ratio

The maximum Sharpe ratio, θ in (2), is unattainable in the presence of parameter uncertainty, which calls for a criterion that accounts for the impact of estimation errors on out-of-sample performance. In this paper, we aim to optimize the *out-of-sample Sharpe ratio*, which for a given estimated portfolio $\hat{w}$ is defined as $\hat{w}'\mu/\sqrt{\hat{w}'\Sigma\hat{w}}$. The out-of-sample Sharpe ratio is a random variable and thus we follow the literature by optimizing the *expected* out-of-sample Sharpe ratio,

$$\mathcal{S}(\hat{w}) = \mathbb{E} \left[\frac{\hat{w}'\mu}{\sqrt{\hat{w}'\Sigma\hat{w}}} \right], \quad (5)$$

where the expectation is taken with respect to the normal distribution of asset returns with moments μ and Σ . While there are specific cases in which (5) can be evaluated exactly (Kan, Wang, and Zheng 2021a), in most cases it is difficult to evaluate analytically.⁶ Therefore, to allow for a wider application of the expected out-of-sample Sharpe ratio criterion, it is useful to use mathematically tractable approximations. In this section, we compare two approximations to $\mathcal{S}(\hat{w})$.

The first approximation is that proposed by DeMiguel, Martin-Utrera, and Nogales (2013), which can be interpreted as a zero-order Taylor-series expansion of $\mathcal{S}(\hat{w})$:

$$\tilde{\mathcal{S}}_0(\hat{w}) = \frac{\mathbb{E}[\hat{w}'\mu]}{\sqrt{\mathbb{E}[\hat{w}'\Sigma\hat{w}]}}. \quad (6)$$

Then, we introduce a second approximation that consists in taking a first-order Taylor-series expansion of $\mathcal{S}(\hat{w})$ around $\tilde{\mathcal{S}}_0(\hat{w})$, which gives⁷

$$\tilde{\mathcal{S}}_1(\hat{w}) = \tilde{\mathcal{S}}_0(\hat{w}) + \frac{3\mathbb{E}[\hat{w}'\mu]\mathbb{V}[\hat{w}'\Sigma\hat{w}]}{8(\mathbb{E}[\hat{w}'\Sigma\hat{w}])^{5/2}} - \frac{\text{Cov}[\hat{w}'\mu, \hat{w}'\Sigma\hat{w}]}{2(\mathbb{E}[\hat{w}'\Sigma\hat{w}])^{3/2}}. \quad (7)$$

Being a first-order Taylor-series expansion, it holds that $\mathcal{S}(\hat{w}) = \tilde{\mathcal{S}}_1(\hat{w}) + O(1/T^2)$. Also, both $\tilde{\mathcal{S}}_0(\hat{w})$ and $\tilde{\mathcal{S}}_1(\hat{w})$ converge to $\mathcal{S}(\hat{w})$ as $T \rightarrow \infty$. Therefore, asymptotically, the two approximations deliver the optimal portfolio combination maximizing the true expected out-of-sample Sharpe ratio.

In Section EC.2 of the e-companion, we consider the setting with fully invested mean-variance portfolios as in Section 5 and derive analytical expressions for the two approximations. We find that the novel approximation, $\tilde{\mathcal{S}}_1$, is very accurate and thus can be used to find a precise approximation to the optimal combination coefficient. However, it is a more complex function of this coefficient, which hinders us from deriving closed-form expressions for the optimal portfolio combination. In

⁶ In Section EC.1 of the e-companion, we provide an exact semi-analytical expression of $\mathcal{S}(\hat{w})$ in terms of a one-dimensional integral for the special case where $\hat{w}$ is normally distributed.

⁷ The first-order Taylor-series expansion of $\mathbb{E}[f(X, Y)]$ around $(\mathbb{E}[X], \mathbb{E}[Y])$ is $\mathbb{E}[f(X, Y)] \approx f(\mathbb{E}[X], \mathbb{E}[Y]) + \frac{1}{2}(f''_{xx}(\mathbb{E}[X], \mathbb{E}[Y])\mathbb{V}[X] + \frac{1}{2}(f''_{yy}(\mathbb{E}[X], \mathbb{E}[Y])\mathbb{V}[Y] + f''_{xy}(\mathbb{E}[X], \mathbb{E}[Y])\text{Cov}(X, Y))$. Applying this formula to $f(x, y) = x/\sqrt{y}$ yields Equation (7).

contrast, the simpler approximation, $\tilde{\mathcal{S}}_0$, is only slightly less accurate⁸ and delivers a similar optimal combination coefficient, which moreover is obtained in closed form. Therefore, in the rest of the paper, we use the approximation $\tilde{\mathcal{S}}_0$ proposed by DeMiguel, Martin-Utrera, and Nogales (2013), which allows us to obtain analytical results.⁹ We will call $\tilde{\mathcal{S}}_0$ the expected out-of-sample Sharpe ratio from now on for brevity, but it is important to keep in mind that it is an approximation.

4. Expected out-of-sample Sharpe ratio with a risk-free asset

We begin our theoretical investigation with the setting in which investing in the risk-free asset is allowed. We show that the mathematics of combining portfolios to maximize expected out-of-sample utility and Sharpe ratio is similar to the mathematics of combining assets without parameter uncertainty in classical portfolio theory. As a result, when a risk-free investment is available and the combination coefficients are unconstrained, maximizing expected out-of-sample utility automatically optimizes expected out-of-sample Sharpe ratio as well.

4.1. Nondiscriminacy and equivalence with expected out-of-sample utility

Consider a portfolio combination that combines the risk-free asset with K portfolios of risky assets,

$$\hat{w}(\kappa) = \hat{w}(\kappa_1, \dots, \kappa_K) = \sum_{i=1}^K \kappa_i \hat{w}_i, \quad (8)$$

where κ_i is the combination coefficient on portfolio $\hat{w}_i$, and the weight on the risk-free asset is $1 - \hat{w}(\kappa)' \iota$. Then, the expected out-of-sample mean return of this portfolio combination is

$$\mathbb{E}[\hat{w}(\kappa)' \mu] = \kappa' \eta \quad \text{with} \quad \eta_i = \mathbb{E}[\hat{w}_i' \mu], \quad (9)$$

and the expected out-of-sample variance is¹⁰

$$\mathbb{E}[\hat{w}(\kappa)' \Sigma \hat{w}(\kappa)] = \kappa' \mathbf{S} \kappa \quad \text{with} \quad \mathbf{S}_{ij} = \mathbb{E}[\hat{w}_i' \Sigma \hat{w}_j], \quad (10)$$

assuming the expectations in η and $\mathbf{S}$ exist. Equations (9) and (10) show that combining portfolios to trade off between expected out-of-sample mean return and variance, $\kappa' \eta$ and $\kappa' \mathbf{S} \kappa$, is similar to combining assets to trade off between mean return and variance, $w' \mu$ and $w' \Sigma w$. In particular, when a risk-free asset is available and the combination coefficients are unconstrained, we recover

⁸ Consider for example the case where μ and Σ are calibrated from a dataset of monthly returns on the 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021, and the sample size is $T = 120$. The expected out-of-sample Sharpe ratio of the portfolio that combines the sample GMV and sample tangent portfolios, with a weight of 50% each, is 0.208 according to $\tilde{\mathcal{S}}_0$ while the true value is 0.211; see Figure EC.2 in the e-companion.

⁹ In Section EC.6 of the e-companion, we show our empirical results are very similar when using $\tilde{\mathcal{S}}_1$ instead of $\tilde{\mathcal{S}}_0$.

¹⁰ The matrix $\mathbf{S}$ is positive definite and thus invertible. This is because the covariance matrix Σ is assumed positive definite, and thus any expected out-of-sample variance is strictly positive, $\kappa' \mathbf{S} \kappa > 0$ for any $\kappa \neq 0$.

the well-known result that maximizing utility maximizes Sharpe ratio as well. Specifically, define the expected out-of-sample utility, introduced by Kan and Zhou (2007), as

$$\mathcal{U}(\hat{w}) = \mathbb{E}[\hat{w}'\mu] - \frac{\gamma}{2}\mathbb{E}[\hat{w}'\Sigma\hat{w}], \quad (11)$$

where $\gamma > 0$ is the risk-aversion coefficient. Then, the vector of combination coefficients, κ , maximizing expected out-of-sample utility of the portfolio combination in (8) is

$$\kappa_\gamma = \arg \max_{\kappa} \mathcal{U}(\hat{w}(\kappa)) = \arg \max_{\kappa} \kappa'\eta - \frac{\gamma}{2}\kappa'\mathbf{S}\kappa = \frac{1}{\gamma}\mathbf{S}^{-1}\eta, \quad (12)$$

and delivers the maximum expected out-of-sample Sharpe ratio according to $\tilde{\mathcal{S}}_0$ in (6):

$$\max_{\kappa} \tilde{\mathcal{S}}_0(\hat{w}(\kappa)) = \max_{\kappa} \frac{\kappa'\eta}{\sqrt{\kappa'\mathbf{S}\kappa}} = \sqrt{\eta'\mathbf{S}^{-1}\eta} = \tilde{\mathcal{S}}_0(\hat{w}(\kappa_\gamma)). \quad (13)$$

In summary, we uncover two important results in the setting with a risk-free asset. First, the expected out-of-sample Sharpe ratio is not a discriminant criterion because, for any $\gamma > 0$, the vector of combination coefficients in (12) delivers the maximum expected out-of-sample Sharpe ratio. Second, maximizing expected out-of-sample utility automatically optimizes expected out-of-sample Sharpe ratio as well. This second important result is formalized in the next proposition.

PROPOSITION 1. *Consider the portfolio in (8) that combines the risk-free asset with K portfolios $\hat{w}_i$. Assume the vector η and the matrix $\mathbf{S}$ in (9)–(10) exist, and the combination coefficients κ are unconstrained. Then, the combination coefficients maximizing expected out-of-sample utility in (12) also maximize expected out-of-sample Sharpe ratio, $\tilde{\mathcal{S}}_0$ in (6).*

The assumption that combination coefficients are unconstrained is crucial. For example, Tu and Zhou (2011) combine the risk-free asset, the tangent portfolio, and the equally weighted portfolio to maximize expected out-of-sample utility. However, they constrain the two combination coefficients on the tangent and equally weighted portfolios to sum to one, and thus their portfolio does not maximize expected out-of-sample Sharpe ratio, which leads them to write (p.208): “When the true parameters are known, maximizing the Sharpe ratio and maximizing the expected utility are equivalent, a well-known fact. However, once the true parameters are unknown, the two are different.” Proposition 1 shows however that this result is no longer true when combination coefficients are unconstrained. We analyze the setting of Tu and Zhou (2011) in Section EC.3 of the e-companion.

4.2. Example: three-fund portfolio of Kan and Zhou (2007)

We now illustrate the finding of Section 4.1 with the popular three-fund portfolio of Kan and Zhou (2007) that combines the risk-free asset, the tangent portfolio, and the GMV portfolio. We assume the covariance matrix Σ is known to simplify the illustration. This three-fund portfolio is

$$\hat{w}(\kappa) = \kappa_1\hat{w}_1 + \kappa_2\hat{w}_2, \quad \text{where} \quad \hat{w}_1 = \Sigma^{-1}\hat{\mu} \quad \text{and} \quad \hat{w}_2 = \Sigma^{-1}\iota.$$

The vector η in (9) and the matrix $\mathbf{S}$ in (10) are

$$\eta = \begin{pmatrix} \theta^2 \\ \mu_g/\sigma_g^2 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \theta^2 + \frac{N}{T} & \mu_g/\sigma_g^2 \\ \mu_g/\sigma_g^2 & 1/\sigma_g^2 \end{pmatrix}.$$

As a result, the unconstrained combination coefficients maximizing expected-out-of-sample utility in (11) are

$$\kappa_\gamma = \frac{1}{\gamma} \mathbf{S}^{-1} \eta = \left(\frac{1}{\gamma} \frac{\psi^2}{\psi^2 + N/T}, \frac{\mu_g}{\gamma} \frac{N/T}{\psi^2 + N/T} \right).$$

Moreover, for any risk-aversion coefficient $\gamma > 0$, these two combination coefficients deliver the maximum expected-out-of-sample Sharpe ratio:

$$\max_{\kappa} \tilde{\mathcal{S}}_0(\hat{w}(\kappa)) = \tilde{\mathcal{S}}_0(\hat{w}(\kappa_\gamma)) = \sqrt{\eta' \mathbf{S}^{-1} \eta} = \sqrt{\theta_g^2 + (\theta^2 - \theta_g^2) \frac{\psi^2}{\psi^2 + N/T}}.$$

That is, as established in Proposition 1, the optimal three-fund portfolio of Kan and Zhou (2007) optimizes both expected out-of-sample utility and Sharpe ratio.

5. Expected out-of-sample Sharpe ratio with fully invested portfolios

As explained in Section 4, when a risk-free investment is allowed and combination coefficients are unconstrained, the expected out-of-sample Sharpe ratio is not a discriminant criterion and is maximized by optimizing expected out-of-sample utility. In this section, we consider the setting with portfolios fully invested in risky assets, in which case the combination coefficients must sum to one. As a result, the expected out-of-sample Sharpe ratio *is* a discriminant criterion and is *not* equivalent to the expected out-of-sample utility. As in Kan, Wang, and Zhou (2021b), we then consider the case of combination of fully invested sample mean-variance portfolios.

5.1. Uniqueness and distinction from expected out-of-sample utility

Consider a combination of K portfolios fully invested in risky assets,

$$\hat{w}(\kappa) = \hat{w}(\kappa_1, \dots, \kappa_K) = \sum_{i=1}^K \kappa_i \hat{w}_i, \quad (14)$$

where $\hat{w}_i' \iota = 1$ for every i . To ensure that $\hat{w}(\kappa)$ is fully invested as well, we need $\kappa' \iota = 1$. Due to the constraint that the combination coefficients sum to one, the theoretical insights from Section 4.1 change in two main ways, similar to classical portfolio theory. First, the combination coefficients maximizing expected out-of-sample Sharpe ratio are unique and given by

$$\kappa_{sr} = \arg \max_{\kappa: \kappa' \iota = 1} \tilde{\mathcal{S}}_0(\hat{w}(\kappa)) = \frac{\mathbf{S}^{-1} \eta}{\iota' \mathbf{S}^{-1} \eta} \quad (15)$$

provided the denominator $\iota' \mathbf{S}^{-1} \eta > 0$, where the vector η and the matrix $\mathbf{S}$ are defined in (9)–(10). Second, provided the entries of the vector η are not all equal, the combination coefficients maximizing expected out-of-sample utility in (11) are different from κ_{sr} , except if $\gamma = \iota' \mathbf{S}^{-1} \eta$. The next proposition formalizes these two results.

PROPOSITION 2. Consider the portfolio in (14) that combines K portfolios $\hat{w}_i$ fully invested in risky assets. Assume the vector η and the matrix $\mathbf{S}$ in (9)–(10) exist, $\iota'\mathbf{S}^{-1}\eta > 0$, and there are no other constraints on the combination coefficients κ besides $\kappa'\iota = 1$. Then, the combination coefficients maximizing expected out-of-sample Sharpe ratio, $\tilde{\mathcal{S}}_0$ in (6), are unique and given by (15). Moreover, if the expected out-of-sample mean returns, $\mathbb{E}[\hat{w}_i'\mu]$, are not all equal, these combination coefficients are different from those maximizing expected out-of-sample utility in (11), except for a single value of the risk-aversion coefficient of $\gamma = \iota'\mathbf{S}^{-1}\eta$.

In the next sections, we use the expected out-of-sample Sharpe ratio to combine fully invested sample mean-variance portfolios, as in Kan, Wang, and Zhou (2021b).

5.2. No parameter uncertainty

When the mean-return vector μ and covariance matrix Σ are known to the investor, the fully invested mean-variance portfolio maximizing the Sharpe ratio is the tangent portfolio,

$$w_{sr}^* = \frac{1}{\gamma_{sr}} \Sigma^{-1} \mu, \quad (16)$$

where

$$\gamma_{sr} = \iota' \Sigma^{-1} \mu = \mu_g / \sigma_g^2 \quad (17)$$

is the risk-aversion coefficient implied by the Sharpe ratio performance criterion. We assume $\gamma_{sr} > 0$, which is a mild assumption for typical datasets.¹¹ For forthcoming results, it is useful to rewrite this portfolio as

$$w_{sr}^* = w_g + \frac{1}{\gamma_{sr}} w_z, \quad (18)$$

where w_z is a zero-cost portfolio (i.e., $w_z'\iota = 1$) defined as

$$w_z = \mathbf{B}\mu \quad \text{with} \quad \mathbf{B} = \Sigma^{-1}(\mathbf{I} - \iota w_g'). \quad (19)$$

A useful property of the matrix $\mathbf{B}$ is that $\mu'\mathbf{B}\mu = \psi^2$, where ψ^2 is defined in (4).

5.3. Uncertainty in mean returns

We now consider that the covariance matrix Σ is known, and thus estimation errors only come from sample mean returns $\hat{\mu}$, as in Garlappi, Uppal, and Wang (2007) and Paulsen and Söhl (2020), for example. Though unrealistic, this assumption is useful because it allows us to derive the optimal mean-variance portfolio when estimation errors in the covariance matrix are negligible relative to those in sample mean returns. In a real setting, the covariance matrix Σ can then be estimated by state-of-the-art approaches such as nonlinear shrinkage by Ledoit and Wolf (2017, 2020). We find in

¹¹ This assumption is equivalent to the mean return of the GMV portfolio, μ_g , being strictly positive. If γ_{sr} is negative, then the maximum-Sharpe-ratio portfolio does not exist, and w_{sr}^* in (16) is the portfolio with minimum Sharpe ratio.

the empirical results of Section 7 that this approach delivers better out-of-sample performance than the optimal mean-variance portfolio derived under the sample covariance matrix in Section 5.4.

As in Kan, Wang, and Zhou (2021b), we combine fully invested sample mean-variance portfolios. Specifically, we combine the GMV portfolio, w_g , with the mean-variance portfolio having unit exposure to the estimated zero-cost portfolio, $\hat{w}_z = \mathbf{B}\hat{\mu}$.¹² This gives:

$$\hat{w}_{mv}^\sigma(\kappa) = (1 - \kappa)w_g + \kappa(w_g + \hat{w}_z) = w_g + \kappa\mathbf{B}\hat{\mu}, \quad (20)$$

where w_g and $\mathbf{B}$ are known because we assume the covariance matrix Σ is known. From now on, we call the quantity $\gamma_{sr}\kappa$ the *shrinkage intensity* because it quantifies the intensity with which the GMV portfolio is shrunk toward the tangent portfolio:

$$\hat{w}_{mv}^\sigma(\kappa) = (1 - \gamma_{sr}\kappa)w_g + \gamma_{sr}\kappa\hat{w}_{sr}^* \quad (21)$$

with $\hat{w}_{sr}^* = \gamma_{sr}^{-1}\Sigma^{-1}\hat{\mu}$. Using the results in Kan, Wang, and Zhou (2021b), the combination coefficient κ maximizing expected out-of-sample utility in (11) is

$$\kappa_\gamma^\sigma = \arg \max_{\kappa} \mathcal{U}(\hat{w}_{mv}^\sigma(\kappa)) = \frac{1}{\gamma} \frac{\psi^2}{\psi^2 + (N-1)/T}. \quad (22)$$

In the next proposition, we analytically characterize the combination coefficient maximizing expected out-of-sample Sharpe ratio instead.¹³

PROPOSITION 3. *When the covariance matrix Σ is known, the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ in (20) satisfies the following:*

1. *The expected out-of-sample Sharpe ratio of $\hat{w}_{mv}^\sigma(\kappa)$, according to $\tilde{\mathcal{S}}_0$ in (6), is*

$$\tilde{\mathcal{S}}_0(\hat{w}_{mv}^\sigma(\kappa)) = \frac{\mu_g + \kappa\psi^2}{\sqrt{\sigma_g^2 + \kappa^2(\psi^2 + \frac{N-1}{T})}}. \quad (23)$$

2. *The combination coefficient maximizing expected out-of-sample Sharpe ratio in (23) is*

$$\kappa_{sr}^\sigma = \frac{1}{\gamma_{sr}} \frac{\psi^2}{\psi^2 + (N-1)/T}. \quad (24)$$

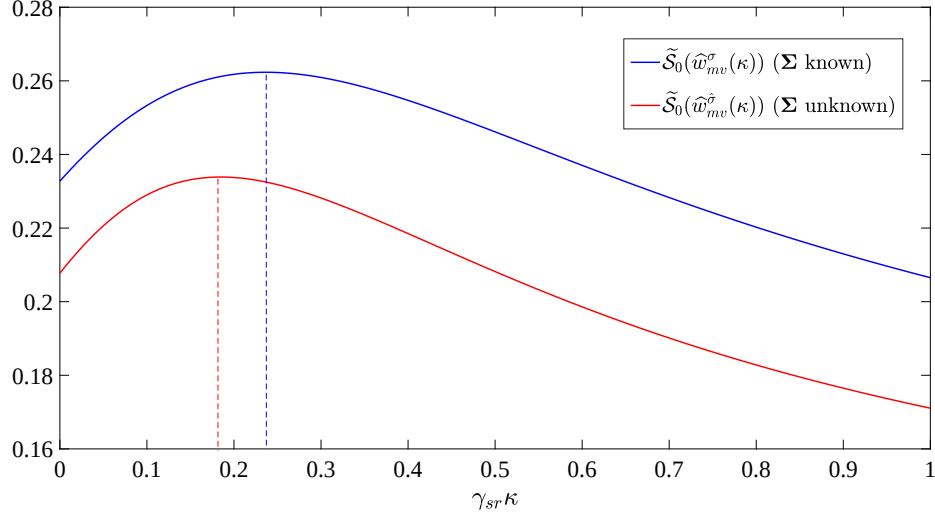
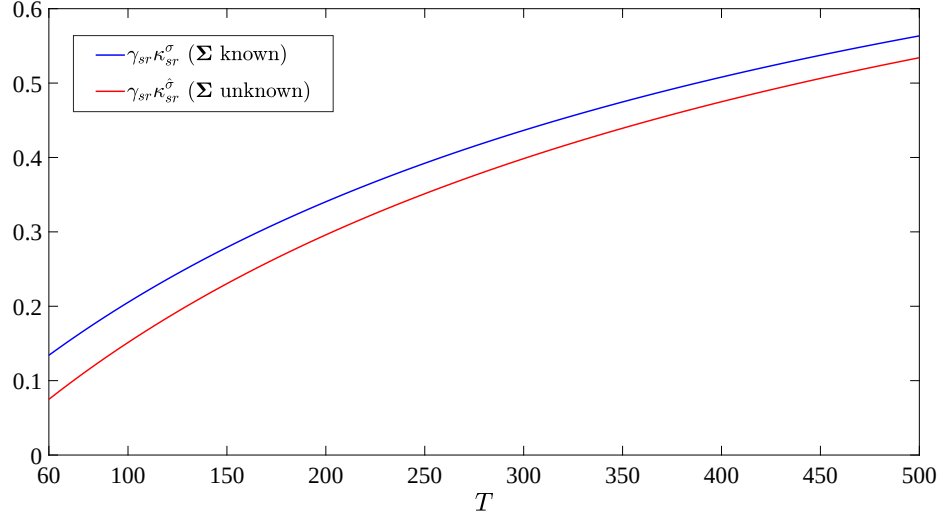
Moreover, $\kappa_{sr}^\sigma \rightarrow 1/\gamma_{sr}$ as $T \rightarrow \infty$, and thus $\hat{w}_{mv}^\sigma(\kappa_{sr}^\sigma)$ is a consistent estimator of w_{sr}^* .

3. *The optimal value of the expected out-of-sample Sharpe ratio is*

$$\tilde{\mathcal{S}}_0(\hat{w}_{mv}^\sigma(\kappa_{sr}^\sigma)) = \sqrt{\theta_g^2 + (\theta^2 - \theta_g^2) \frac{\psi^2}{\psi^2 + (N-1)/T}}. \quad (25)$$

¹² We also consider adding the equally weighted portfolio in the combination, similar to Tu and Zhou (2011). However, the resulting performance improvement in our empirical data is limited, and thus we omit these results for brevity.

¹³ Given that $\hat{w}_{mv}^\sigma(\kappa)$ is normally distributed, its expected out-of-sample Sharpe ratio can be evaluated exactly using the semi-analytical integral expression provided in Section EC.1 of the e-companion.

Figure 1 Expected out-of-sample Sharpe ratio with fully invested portfoliosa) Expected out-of-sample Sharpe ratio as a function of the shrinkage intensity ($\theta = 0.341$)b) Optimal shrinkage intensity as a function of the sample size ($\gamma_{sr} = 5.35$)

Note. Both panels are constructed with population values of the mean and covariance matrix of asset returns calibrated from a dataset of monthly returns on 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021. The maximum Sharpe ratio without parameter uncertainty is $\theta = 0.341$. Panel a) depicts the expected out-of-sample Sharpe ratio, according to the approximation $\tilde{\mathcal{S}}_0$ in (6), of fully invested sample mean-variance portfolios. The portfolio $\hat{w}_{mv}^\sigma(\kappa)$ in (20) assumes the covariance matrix is known, and the portfolio $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ in (26) uses the sample covariance matrix. Both portfolios are estimated with a sample size of $T = 120$ monthly observations. The horizontal axis is the intensity with which the global-minimum-variance portfolio is shrunk toward the sample tangent portfolio obtained for a coefficient $\kappa = 1/\gamma_{sr}$ defined in (17). Panel b) depicts the optimal shrinkage intensity maximizing expected out-of-sample Sharpe ratio as a function of the sample size T between 60 and 500 months. The combination coefficient κ_{sr}^σ (24) assumes the covariance matrix is known, and the coefficient $\kappa_{sr}^{\hat{\sigma}}$ in (29) uses the sample covariance matrix.

A number of comments are in order in relation to the results presented in Proposition 3. First, to maximize expected out-of-sample Sharpe ratio, one can simply use the combination coefficient maximizing expected out-of-sample utility in (22) and replace the risk-aversion coefficient γ by γ_{sr} in (17). Referring to the result in Proposition 2, this is because the quantity $\iota' \mathbf{S}^{-1} \eta = \gamma_{sr}$ for the considered portfolio combination in (20). Therefore, when the covariance matrix is known, a naive application of the expected out-of-sample utility solution is sufficient to maximize expected out-of-sample Sharpe ratio.

Second, the quantity $\psi^2 / (\psi^2 + (N - 1)/T)$ plays two roles: it is the optimal intensity with which the sample tangent portfolio must be shrunk toward the GMV portfolio to maximize expected out-of-sample Sharpe ratio, and from Equation (25), it is the percentage of the maximum square Sharpe ratio on top of that of the GMV portfolio that is achievable out of sample.

Third, the optimal combination coefficient κ_{sr}^σ is asymptotically optimal because it corresponds to $1/\gamma_{sr}$ when the sample size $T \rightarrow \infty$, which is the coefficient that delivers the tangent portfolio and this portfolio is optimal without parameter uncertainty.

In Figure 1, we illustrate the results contained in Proposition 3. We calibrate the moments of asset returns, μ and Σ , from a dataset of monthly returns on the 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021. Panel a) considers a sample size $T = 120$ and depicts the expected out-of-sample Sharpe ratio of the portfolio $\hat{w}_{mv}^\sigma(\kappa)$ as a function of the shrinkage intensity, $\gamma_{sr}\kappa$. The panel shows that by appropriately choosing the shrinkage intensity, the investor can obtain a larger expected out-of-sample Sharpe ratio than both the GMV and tangent portfolios. Also, compared to the maximum Sharpe ratio without parameter uncertainty, $\theta = 0.341$, the maximum expected out-of-sample Sharpe ratio is 0.262, which is 23% smaller. Then, Panel b) depicts the optimal shrinkage intensity, $\gamma_{sr}\kappa_{sr}^\sigma$, as a function of the sample size T . The panel shows that the weight allocated to the tangent portfolio increases with T , which is because estimation errors in mean returns decrease with the sample size.

5.4. Uncertainty in mean returns and covariance matrix

We now turn to the case in which there are estimation errors in the covariance matrix in addition to those in mean returns, i.e., Σ is estimated via the sample estimator $\hat{\Sigma}$ in (1). In this setting, the portfolio combination in (20) becomes

$$\hat{w}_{mv}^{\hat{\sigma}}(\kappa) = (1 - \kappa)\hat{w}_g + \kappa(\hat{w}_g + \hat{\mathbf{B}}\hat{\mu}) = \hat{w}_g + \kappa\hat{\mathbf{B}}\hat{\mu}, \quad (26)$$

where the sample GMV portfolio $\hat{w}_g$ and the matrix $\hat{\mathbf{B}}$ are obtained after replacing Σ by $\hat{\Sigma}$. Kan, Wang, and Zhou (2021b) show that the combination coefficient κ maximizing expected out-of-sample utility in (11) is

$$\kappa_{\hat{\gamma}}^{\hat{\sigma}} = \arg \max_{\kappa} \mathcal{U}(\hat{w}_{mv}^{\hat{\sigma}}(\kappa)) = \frac{1}{\gamma} \frac{(T - N)(T - N - 3)}{T(T - 2)} \frac{\psi^2}{\psi^2 + (N - 1)/T}. \quad (27)$$

In the next proposition, we analytically characterize the combination coefficient maximizing expected out-of-sample Sharpe ratio instead.

PROPOSITION 4. *Let $T \geq N + 4$. When the covariance matrix Σ is estimated by $\hat{\Sigma}$ in (1), the sample mean-variance portfolio $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ in (26) satisfies the following:*

1. *The expected out-of-sample Sharpe ratio of $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$, according to $\tilde{\mathcal{S}}_0$ in (6), is*

$$\tilde{\mathcal{S}}_0(\hat{w}_{mv}^{\hat{\sigma}}(\kappa)) = \frac{\mu_g + \kappa \frac{T}{T-N-1} \psi^2}{\sqrt{\frac{T-2}{T-N-1} \sigma_g^2 + \kappa^2 \frac{T^2(T-2)}{(T-N)(T-N-1)(T-N-3)} (\psi^2 + \frac{N-1}{T})}}. \quad (28)$$

2. *The combination coefficient maximizing expected out-of-sample Sharpe ratio in (28) is*

$$\kappa_{sr}^{\hat{\sigma}} = \frac{1}{\gamma_{sr}} \frac{(T-N)(T-N-3)}{T(T-N-1)} \frac{\psi^2}{\psi^2 + (N-1)/T}, \quad (29)$$

which is strictly smaller than that when the covariance matrix is known in (24). Moreover, $\kappa_{sr}^{\hat{\sigma}} \rightarrow 1/\gamma_{sr}$ as $T \rightarrow \infty$, and thus $\hat{w}_{mv}^{\hat{\sigma}}(\kappa_{sr}^{\hat{\sigma}})$ is a consistent estimator of w_{sr}^* .

3. *The optimal value of the expected out-of-sample Sharpe ratio is*

$$\tilde{\mathcal{S}}_0(\hat{w}_{mv}^{\hat{\sigma}}(\kappa_{sr}^{\hat{\sigma}})) = \sqrt{\theta_g^2 \frac{T-N-1}{T-2} + (\theta^2 - \theta_g^2) \frac{(T-N)(T-N-3)}{(T-2)(T-N-1)} \frac{\psi^2}{\psi^2 + (N-1)/T}}, \quad (30)$$

which is strictly smaller than that when the covariance matrix is known in (25).

Several comments are in order in relation to the results presented in Proposition 4. First, to maximize expected out-of-sample Sharpe ratio, one *cannot* simply use the combination coefficient maximizing expected out-of-sample utility in (27) and replace the risk-aversion coefficient γ by γ_{sr} , contrary to the case where the covariance matrix is known in Section 5.3. Instead, to obtain the optimal combination coefficient in (29), γ must be replaced by $\gamma_{sr} \frac{T-N-1}{T-2}$, which is the quantity $\iota' \mathbf{S}^{-1} \eta$ in Proposition 2. In particular, one must shrink less intensively toward the sample GMV portfolio to maximize expected out-of-sample Sharpe ratio instead of expected out-of-sample utility. Therefore, a naive application of the results in Kan, Wang, and Zhou (2021b), which are derived for the utility criterion, does not deliver an optimal expected Sharpe ratio out of sample.

Second, introducing parameter uncertainty in the covariance matrix deteriorates the maximum expected Sharpe ratio that is achievable out of sample, which can be seen by comparing (30) with (25). For example, when the sample size is $T = 120$ in Panel a) of Figure 1, the maximum expected out-of-sample Sharpe ratio is 0.234 when the covariance matrix is unknown versus 0.262 when it is known. For a smaller sample size of $T = 60$ observations, it decreases to 0.191 versus 0.250.

Third, relative to the case in which the covariance matrix is known, a more intensive shrinkage toward the GMV portfolio is optimal in the presence of estimation errors in the covariance matrix. For example, when $T = 120$ in Panel a) of Figure 1, the optimal weight on the GMV portfolio

is 0.815 when the covariance matrix is unknown, versus 0.764 when it is known. This is because estimation errors in the covariance matrix amplify the impact of estimation errors in mean returns, and thus it is optimal to shrink more intensively toward the GMV portfolio even though this portfolio is no longer perfectly estimated.

6. Alleviating estimation errors in combination coefficients

The optimal combination coefficients in Propositions 3 and 4 are oracle estimators because they depend on the true moments of asset returns via the parameters γ_{sr} in (17) and ψ^2 in (4). In practice these parameters are estimated,¹⁴ and estimation errors in combination coefficients deteriorate out-of-sample performance (Kan and Wang 2021, Lassance, Martin-Utrera, and Simaan 2021b).

To alleviate this problem, we follow Lassance, Martin-Utrera, and Simaan (2021b) who propose a methodology to trade off between expected out-of-sample utility and out-of-sample utility *standard deviation*. As they show, this is equivalent to maximizing expected out-of-sample utility via a min-max robust optimization problem that accounts for estimation errors in expected out-of-sample utility; see their section 4.5 for more details. We use their methodology and propose an optimization problem that trades off between the expectation and standard deviation of out-of-sample Sharpe ratio. We show that solving this problem amounts to shrinking more intensively toward the GMV portfolio, and our empirical results in Section 7 suggest that this additional source of shrinkage helps improve expected out-of-sample Sharpe ratio for realistic sample sizes.

6.1. General principle

To operationalize this approach, we need to characterize the variance of out-of-sample Sharpe ratio, which is defined as

$$\mathcal{V}(\hat{w}) = \mathbb{V} \left[\frac{\hat{w}'\mu}{\sqrt{\hat{w}'\Sigma\hat{w}}} \right]. \quad (31)$$

Just like the expected out-of-sample Sharpe ratio, the variance of out-of-sample Sharpe ratio is difficult to evaluate exactly, in general.¹⁵ Therefore, we employ the following approximation, denoted by $\tilde{\mathcal{V}}_0(\hat{w})$, which is similar to that use for the expectation, $\tilde{\mathcal{S}}_0(\hat{w})$ in (6), and is mathematically tractable:

$$\mathcal{V}(\hat{w}) = \mathbb{E} \left[\frac{(\hat{w}'\mu)^2}{\hat{w}'\Sigma\hat{w}} \right] - \left(\mathbb{E} \left[\frac{\hat{w}'\mu}{\sqrt{\hat{w}'\Sigma\hat{w}}} \right] \right)^2 \approx \frac{\mathbb{E}[(\hat{w}'\mu)^2]}{\mathbb{E}[\hat{w}'\Sigma\hat{w}]} - (\tilde{\mathcal{S}}_0(\hat{w}))^2 = \frac{\mathbb{V}[\hat{w}'\mu]}{\mathbb{E}[\hat{w}'\Sigma\hat{w}]} = \tilde{\mathcal{V}}_0(\hat{w}). \quad (32)$$

In Section EC.4 of the e-companion, we consider fully invested sample mean-variance portfolios as in Section 5 and show that the approximation $\tilde{\mathcal{V}}_0$ overestimates the true variance because it

¹⁴ We explain how we estimate the parameters γ_{sr} and ψ^2 in Section 7.1.

¹⁵ In Section EC.1 of the e-companion, we provide an exact semi-analytical expression of $\mathcal{V}(\hat{w})$ in terms of a one-dimensional integral for the special case where $\hat{w}$ is normally distributed.

ignores the correlation between out-of-sample mean return and variance. Moreover, we show that it is not too accurate when the combination coefficient κ is not small. As a result, for the reader who wishes to rely on a more accurate approximation, in Section EC.4 we also provide closed-form expressions for a Taylor-series approximation that is a more complex function of κ but that is also very accurate. We rely on $\tilde{\mathcal{V}}_0$ in the sequel because it allows for analytical insights, but we also provide empirical results using the refined approximation in Section EC.6 of the e-companion.

Exploiting the approximation $\tilde{\mathcal{V}}_0$ to the variance of out-of-sample Sharpe ratio, we follow Lassance, Martin-Utrera, and Simaan (2021b) and optimize the difference between the expectation and a multiple of the standard deviation of out-of-sample Sharpe ratio. For a generic combination of K estimated portfolios as in (8), this amounts to solving

$$\kappa_\lambda = \arg \max_{\kappa \in \mathcal{K}} \tilde{\mathcal{S}}_0(\hat{w}(\kappa)) - \lambda \sqrt{\tilde{\mathcal{V}}_0(\hat{w}(\kappa))} = \arg \max_{\kappa \in \mathcal{K}} \frac{\kappa' \eta - \lambda \sqrt{\kappa' \mathbf{V} \kappa}}{\sqrt{\kappa' \mathbf{S} \kappa}}, \quad (33)$$

where $\lambda \geq 0$, $\mathcal{K}$ is a set of constraints on the vector of combination coefficients κ (e.g., $\kappa' \iota = 1$), the vector η is defined in (9), the matrix $\mathbf{S}$ in (10), and the matrix $\mathbf{V}$ is defined as

$$\mathbf{V}_{ij} = \mu' \text{Cov}[\hat{w}_i, \hat{w}_j] \mu, \quad (34)$$

where $\text{Cov}[\hat{w}_i, \hat{w}_j] = \mathbb{E}[\hat{w}_i \hat{w}_j'] - \mathbb{E}[\hat{w}_i] \mathbb{E}[\hat{w}_j']$ is the cross-covariance matrix of $\hat{w}_i$ and $\hat{w}_j$.¹⁶

Note that the objective function in (33) is homogeneous of degree zero and thus, similar to the expected out-of-sample Sharpe ratio, it is not a discriminant criterion when combination coefficients are unconstrained; if κ_λ is a solution, then any $a\kappa_\lambda$ with $a > 0$ is a solution as well. In the next proposition, we impose the constraint $\kappa' \iota = 1$ and show that adding a penalty on the standard deviation of out-of-sample Sharpe ratio in (33) can be interpreted as shrinking the covariance matrix of estimated portfolios, $\mathbf{S}$, toward $\mathbf{V}$.

PROPOSITION 5. *Consider the portfolio in (8) that combines K portfolios $\hat{w}_i$. Let the combination coefficients sum to one, $\kappa' \iota = 1$. Then, the endogenous solution to the coefficients maximizing problem (33) corresponds to that maximizing expected out-of-sample Sharpe ratio, $\kappa_{sr} = \frac{\mathbf{S}^{-1} \eta}{\iota' \mathbf{S}^{-1} \eta}$, with the matrix $\mathbf{S}$ replaced by*

$$(1 - \delta) \mathbf{S} + \delta \mathbf{V} \quad \text{with} \quad \delta = \left(1 + \frac{v}{s^2} \left(\frac{n}{\sqrt{\lambda}} - v \right) \right)^{-1}, \quad (35)$$

where $n = \kappa' \eta$, $s^2 = \kappa' \mathbf{S} \kappa$, and $v^2 = \kappa' \mathbf{V} \kappa$.

¹⁶ The matrix $\mathbf{V}$ is positive semi-definite. However, it is not positive definite if some portfolios $\hat{w}_i$ are not random, because we can then have $\kappa' \mathbf{V} \kappa = 0$ even if $\kappa \neq 0$ (it suffices to take $\kappa_i \neq 0$ and $\kappa_j = 0$ for all $j \neq i$).

Several comments are in order in relation to the results in Proposition 5. First, when $\lambda = 0$, there is no penalty and the matrix $\mathbf{S}$ is not shrunk: $\delta = 0$ when $\lambda = 0$. In contrast, as $\lambda \rightarrow \infty$, we have $\delta = 1$ and $\mathbf{S}$ is fully shrunk toward $\mathbf{V}$. Second, here we shrink the covariance matrix of estimated portfolio weights, $\mathbf{S}$ in (10), rather than the covariance matrix of asset returns as pioneered by Ledoit and Wolf (2004). Third, the result in Proposition 5 is similar to DeMiguel, Garlappi, Nogales, and Uppal (2009a, Proposition 2) who show that penalizing the norm of portfolio weights is equivalent to shrinking the covariance matrix of asset returns. Here, we show that penalizing the standard deviation of out-of-sample Sharpe ratio is equivalent to shrinking the covariance matrix of estimated portfolio weights, $\mathbf{S}$, toward $\mathbf{V}$.

In the next section, we consider the case of fully invested mean-variance portfolios and show that shrinking $\mathbf{S}$ toward $\mathbf{V}$ has the effect of shrinking portfolio weights more intensively toward the GMV portfolio.

6.2. The case of fully invested mean-variance portfolios

We now focus on the fully invested sample mean-variance portfolio combinations considered in Section 5, i.e., $\hat{w}_{mv}^\sigma(\kappa)$ in (20) when the covariance matrix Σ is assumed known, and $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ in (26) when using the sample covariance matrix. In the next proposition, we characterize their variance of out-of-sample Sharpe ratio according to the approximation $\tilde{\mathcal{V}}_0$ in (32).

PROPOSITION 6. *The variance of out-of-sample Sharpe ratio of sample mean-variance portfolios can be approximated as follows:*

1. *When the covariance matrix Σ is known,*

$$\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa)) = \frac{1}{T} \frac{\kappa^2 \psi^2}{\sigma_g^2 + \kappa^2 \left(\psi^2 + \frac{N-1}{T} \right)}. \quad (36)$$

2. *When the covariance matrix Σ is unknown and estimated by $\hat{\Sigma}$ in (1), and $T \geq N + 4$,*

$$\tilde{\mathcal{V}}_0(\hat{w}_{mv}^{\hat{\sigma}}(\kappa)) = \frac{\frac{1}{T-N-1} \sigma_g^2 \psi^2 + \kappa^2 \psi^2 f(T, N, \psi^2)}{\frac{T-2}{T-N-1} \sigma_g^2 + \kappa^2 \frac{T^2(T-2)}{(T-N)(T-N-1)(T-N-3)} \left(\psi^2 + \frac{N-1}{T} \right)}, \quad (37)$$

where

$$f(T, N, \psi^2) = \frac{2T(N+1) + T^2(T-N-3) + 2(T-N)\psi^2}{(T-N)(T-N-1)^2(T-N-3)}. \quad (38)$$

We then consider the optimization problem in (33) and use the combination coefficient

$$\kappa_\lambda^\sigma = \arg \max_{\kappa} \tilde{\mathcal{S}}_0(\hat{w}_{mv}^\sigma(\kappa)) - \lambda \sqrt{\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))} \quad (39)$$

when the covariance matrix Σ is known, and

$$\kappa_\lambda^{\hat{\sigma}} = \arg \max_{\kappa} \tilde{\mathcal{S}}_0(\hat{w}_{mv}^{\hat{\sigma}}(\kappa)) - \lambda \sqrt{\tilde{\mathcal{V}}_0(\hat{w}_{mv}^{\hat{\sigma}}(\kappa))} \quad (40)$$

when using the sample covariance matrix $\hat{\Sigma}$, where $\lambda \geq 0$. In the next proposition, we derive a closed-form solution for κ_λ^σ in (39). Moreover, while $\kappa_\lambda^{\hat{\sigma}}$ does not have an easy closed-form expression, we show that as λ goes from zero to infinity, $\kappa_\lambda^{\hat{\sigma}}$ goes from $\kappa_{sr}^{\hat{\sigma}}$ to zero.

PROPOSITION 7. *The combination coefficient optimizing the tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio satisfies the following:*

1. *When the covariance matrix Σ is known, κ_λ^σ in (39) is*

$$\kappa_\lambda^\sigma = \begin{cases} \kappa_{sr}^\sigma \left(1 - \frac{\lambda}{\psi\sqrt{T}}\right) & \text{if } \lambda \leq \psi\sqrt{T}, \\ 0 & \text{if } \lambda \geq \psi\sqrt{T}, \end{cases} \quad (41)$$

where κ_{sr}^σ in (24) is the combination coefficient maximizing expected out-of-sample Sharpe ratio.

2. *When the covariance matrix Σ is unknown and estimated by $\hat{\Sigma}$ in (1), $\kappa_\lambda^{\hat{\sigma}} = \kappa_{sr}^{\hat{\sigma}}$ if $\lambda = 0$, $\kappa_\lambda^{\hat{\sigma}} = 0$ if $\lambda \rightarrow \infty$, and $0 \leq \kappa_\lambda^{\hat{\sigma}} \leq \kappa_{sr}^{\hat{\sigma}}$ otherwise with strict inequalities if $\psi^2 > 0$, where $\kappa_{sr}^{\hat{\sigma}}$ in (29) is the combination coefficient maximizing expected out-of-sample Sharpe ratio.¹⁷*

3. *Both coefficients $\kappa_\lambda^\sigma, \kappa_\lambda^{\hat{\sigma}} \rightarrow 1/\gamma_{sr}$ as $T \rightarrow \infty$ and thus deliver consistent estimators of w_{sr}^* .*

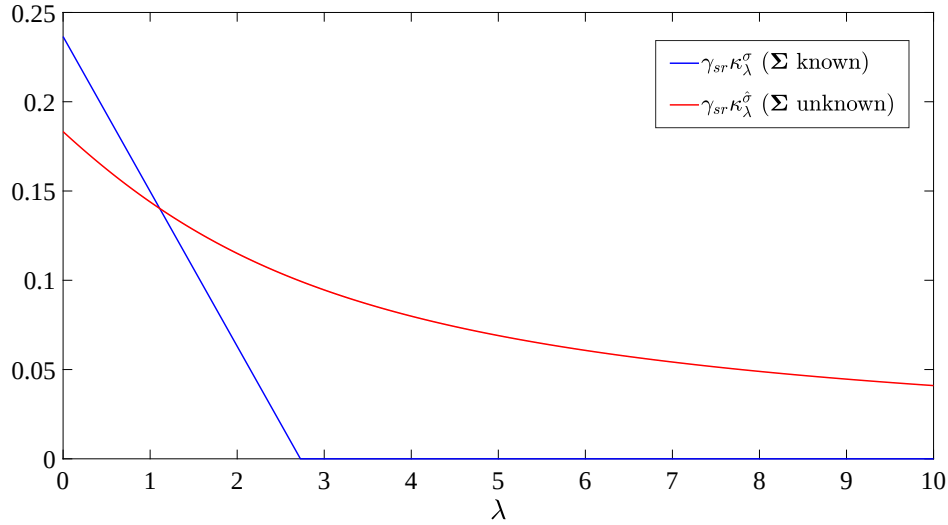
Proposition 7 shows that accounting for the standard deviation of out-of-sample Sharpe ratio acts as an additional source of shrinkage because it amounts to shrinking the portfolio maximizing expected out-of-sample Sharpe ratio toward the GMV portfolio.¹⁸ The impact on the optimal combination coefficient is different if Σ is known or unknown. When Σ is known, Equation (41) shows that the optimal combination coefficient is zero if λ is large enough, and the portfolio corresponds to the GMV portfolio. When using the sample covariance matrix however, this only happens as $\lambda \rightarrow \infty$; otherwise, the combination coefficient is strictly positive (if $\psi^2 > 0$). This difference can be explained because, when Σ is known, the out-of-sample Sharpe ratio variance of the GMV portfolio is zero, and thus it is optimal to fully shrink toward GMV once λ is large enough. When Σ is estimated however, the out-of-sample Sharpe ratio variance of the GMV portfolio is not zero if the sample size is finite, and thus it is not optimal to fully shrink toward GMV. We illustrate this difference in Figure 2 for the dataset considered in Figure 1 and a sample size $T = 120$.

7. Empirical analysis

In this section, we evaluate the out-of-sample performance of the sample mean-variance portfolios maximizing expected out-of-sample Sharpe ratio introduced in this paper. We consider the setting

¹⁷ Given this result, the combination coefficient $\kappa_\lambda^{\hat{\sigma}}$ can be computed using a fine grid of values in the interval $[0, \kappa_{sr}^{\hat{\sigma}}]$ and finding the value maximizing the objective function in (40).

¹⁸ Kan and Wang (2021, Section 3.2) account for estimation errors in their optimal combination coefficients by shrinking more intensively toward the risk-free asset, and they calibrate the optimal intensity via cross-validation taking into account that combination coefficients are estimated. Similarly, we shrink more intensively toward the GMV portfolio, and the intensity depends on the coefficient λ that can also be calibrated in a cross-validated fashion.

Figure 2 Optimal tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio

Note. This figure is constructed with population values of the mean and covariance matrix of asset returns calibrated from a dataset of monthly returns on 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021. It depicts the shrinkage intensities maximizing the tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio for a sample size of $T = 120$ monthly observations. The combination coefficient $\kappa_{\lambda}^{\sigma}$ in (39) assumes the covariance matrix is known, and the coefficient $\kappa_{\lambda}^{\hat{\sigma}}$ in (40) uses the sample covariance matrix. The horizontal axis, λ , controls the tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio. The vertical axis is the intensity with which the global-minimum-variance portfolio is shrunk toward the sample tangent portfolio obtained for a coefficient $\kappa = 1/\gamma_{sr}$ defined in (17).

with fully invested portfolios because, as shown in Section 4, when a risk-free investment is allowed and the combination coefficients are unconstrained, the expected out-of-sample Sharpe ratio is not a discriminant criterion and is maximized when optimizing expected out-of-sample utility as considered in prior literature. The portfolio strategies we compare are listed in Table 1.

7.1. Estimation of optimal combination coefficients

The optimal combination coefficients proposed in Sections 5 and 6 depend on the moments of asset returns via the parameters γ_{sr} in (17) and ψ^2 , σ_g^2 in (4), which we need to estimate. Concerning ψ^2 , we use the adjusted estimator of Kan and Zhou (2007):

$$\hat{\psi}^2 = \frac{(T - N - 1)\hat{\psi}_p^2 - (N - 1)}{T} + \frac{2(\hat{\psi}_p^2)^{\frac{N-1}{2}}(1 + \hat{\psi}^2)^{-\frac{T-2}{2}}}{T \times B\left(\frac{\hat{\psi}_p^2}{1 + \hat{\psi}^2}; \frac{N-1}{2}, \frac{T-N+1}{2}\right)}, \quad (42)$$

where $\hat{\psi}_p^2 = \hat{\mu}' \hat{\mathbf{B}} \hat{\mu}$ is the plug-in estimator of ψ^2 and $B(x; a, b) = \int_0^x y^{a-1}(1-y)^{b-1} dy$ is the incomplete beta function. Note that this estimate is always positive, as desired.

Concerning σ_g^2 , we estimate it via the plug-in estimator. However, to alleviate the impact of estimation errors, we rely on recent developments in covariance matrix estimation and plug in the

nonlinear shrinkage estimator proposed by Ledoit and Wolf (2020), denoted by $\hat{\Sigma}_{nl}$. This estimator is well suited for portfolio optimization because it is designed to minimize variance as in Ledoit and Wolf (2017), and benefits from a fast analytical solution. The estimator of σ_g^2 we use is thus

$$\hat{\sigma}_g^2 = \left(\iota' \hat{\Sigma}_{nl}^{-1} \iota \right)^{-1}. \quad (43)$$

Finally, concerning γ_{sr} , we estimate it by plugging in $\hat{\Sigma}_{nl}$ and $\hat{\mu}$ but, following Kircher and Rosch (2021), we also set a lower bound $\gamma_{\min}$ that γ_{sr} cannot cross:

$$\hat{\gamma}_{sr} = \max\left(\gamma_{\min}, \iota' \hat{\Sigma}_{nl}^{-1} \hat{\mu}\right). \quad (44)$$

Imposing a lower bound $\gamma_{\min}$ is useful because the tangent portfolio is obtained for a weight $1/\gamma_{sr}$ on the zero-cost portfolio in (18). Therefore, as noted by Kirby and Ostdiek (2012), it is overly exposed to sample mean returns and performs very poorly when γ_{sr} is estimated to be close to zero, which also negatively impacts the performance of portfolios that combine the sample GMV and sample tangent portfolios. We have tested $\gamma_{\min} = 3$ as in Kircher and Rosch (2021) and $\gamma_{\min} = 5$ and choose the latter because it systematically delivers a better performance for real data.

7.2. Simulated data

7.2.1. Methodology for simulated data We begin our empirical investigation using simulated multivariate normal asset returns. Specifically, we calibrate the mean μ and covariance matrix Σ from monthly returns of four datasets and then randomly draw $M = 10,000$ times T observations from a normal distribution, $\mathcal{N}(\mu, \Sigma)$. For each of the M simulations, we compute the estimated portfolio $\hat{w}_m$ and evaluate its annualized expected out-of-sample Sharpe ratio as

$$\sqrt{12}\mathcal{S}(\hat{w}) \approx \frac{1}{M} \sum_{m=1}^M \sqrt{12} \frac{\hat{w}_m' \mu}{\sqrt{\hat{w}_m' \Sigma \hat{w}_m}}.$$

We consider a sample size $T \in (60, 120, 240, 480)$, i.e., 5, 10, 20, and 40 years of returns.

We consider four monthly return datasets commonly used in the literature: (i) 10 momentum portfolios from January 1927 to June 2021 (10MOM), (ii) 25 portfolios formed on size and book-to-market from January 1927 to June 2021 (25SBTM), (iii) 25 portfolios formed on operating profitability and investment from July 1963 to June 2021 (25OPINV), and (iv) 30 industry portfolios from January 1927 to June 2021 (30IND).

We consider seven portfolio strategies, which are described in Table 1. First, the mean-variance portfolio maximizing expected out-of-sample Sharpe ratio as in Section 5. Second, the mean-variance portfolio maximizing the tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio as in Section 6 with a coefficient $\lambda = 1$.¹⁹ Third, the mean-variance portfolio of

¹⁹ We find that the results are not very sensitive to the value of the coefficient λ . In particular, choosing $\lambda = 2$ or calibrating λ via cross-validation yields similar conclusions, and thus we omit these results for brevity.

Kan, Wang, and Zhou (2021b) that maximizes expected out-of-sample utility with a risk-aversion coefficient of $\gamma = 5$.²⁰ Fourth, the GMV portfolio. Fifth, the tangent portfolio. Sixth, the equally weighted portfolio. Seventh, the reward-to-risk timing strategy of Kirby and Ostdiek (2012),

$$w_i = \frac{\max(0, \hat{\mu}_i)/\hat{\sigma}_i^2}{\sum_{j=1}^N \max(0, \hat{\mu}_j)/\hat{\sigma}_j^2}, \quad (45)$$

where we estimate $\hat{\sigma}_i^2$ via the sample or shrinkage estimator of the covariance matrix. In the analysis with real data in Section 7.3, we also consider the bootstrap-based portfolio of Kircher and Rosch (2021), but not in this simulated data analysis due to computational constraints.

For each considered portfolio strategy, we distinguish between two cases. First, the case in Section 5.3 where the covariance matrix is assumed known, which we then estimate via the Ledoit and Wolf (2020) nonlinear shrinkage estimator as in Section 7.2.1. For comparison, we also report results using the true, but unfeasible, combination coefficients and covariance matrix. Second, the case in Section 5.4 where we use the sample estimator of the covariance matrix, $\hat{\Sigma}$ in (1). For comparison, we also report results using the true, but unfeasible, combination coefficients.²¹

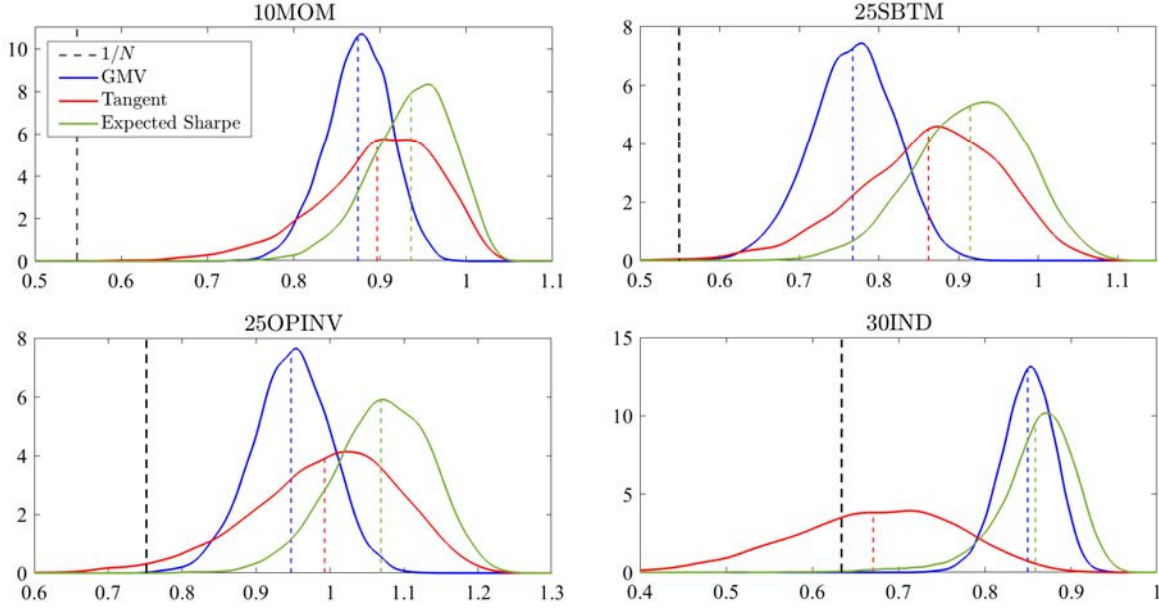
7.2.2. Results for simulated data and Ledoit-Wolf covariance matrix The annualized expected out-of-sample Sharpe ratios for the Ledoit-Wolf covariance matrix are reported in Table 2. A first observation we make is that while the $1/N$ portfolio can sometimes outperform the tangent portfolio when the sample size T is not large ($T = 60$ or 120), it is outperformed by all the other strategies for all sample sizes. Thus, when returns are normally distributed and there are no transaction costs, it is not optimal to fully ignore information in the estimated mean and covariance matrix of asset returns. In particular, the $1/N$ portfolio is outperformed by the simple reward-to-risk timing rule of Kirby and Ostdiek (2012), which is however suboptimal compared to the other portfolios except the tangent portfolio.

Second, a large sample size of $T = 240$ is necessary for the tangent portfolio to outperform the GMV portfolio despite the use of a robust estimator of the covariance matrix. For the 30IND dataset, it is even outperformed for all considered sample sizes.

Third, it is generally useful to combine the GMV and tangent portfolios to improve expected out-of-sample Sharpe ratio. Specifically, for the 10MOM, 25SBTM, and 25OPINV datasets, the proposed mean-variance portfolio designed to optimize expected out-of-sample Sharpe ratio outperforms both the GMV and tangent portfolios for all considered sample sizes. For the 30IND dataset however, the results derived using the true optimal combination coefficients and the true covariance

²⁰ Choosing $\gamma = 5$ is consistent with the lower bound $\gamma_{\min} = 5$ we use to estimate the parameter γ_{sr} in Equation (44). We have also tested $\gamma = 3$ as in Kan, Wang, and Zhou (2021b) but it delivers worse out-of-sample Sharpe ratios.

²¹ Note that the tangent portfolio is different if γ_{sr} in (17) is assumed known or estimated by $\hat{\gamma}_{sr}$ in (44).

Figure 3 Density of out-of-sample Sharpe ratio using the Ledoit-Wolf covariance matrix and $T = 240$ 

Note. This figure depicts the density of out-of-sample Sharpe ratio delivered by the equally weighted portfolio (dotted black vertical line), the GMV portfolio (in blue), the tangent portfolio (in red), and the mean-variance portfolio maximizing expected out-of-sample Sharpe ratio (in green). We consider the theoretical setting in which the covariance matrix is assumed known as in Section 5.3, which is here replaced by the nonlinear shrinkage estimator of Ledoit and Wolf (2020) to ensure that all portfolios are feasible. The densities are obtained by simulating 10,000 datasets of $T = 240$ monthly returns from a multivariate normal distribution whose mean and covariance matrix are calibrated from each of the four datasets listed in Section 7.2.1. The dotted vertical lines depict the mean of each density.

matrix show that the maximum improvement upon the GMV portfolio is limited. Therefore, when considering the feasible case with estimated combination coefficients and the Ledoit-Wolf covariance matrix, it is only for $T = 240$ and 480 that the proposed mean-variance portfolio delivers a larger expected out-of-sample Sharpe ratio than the GMV portfolio.

Fourth, the mean-variance portfolio maximizing expected out-of-sample utility is outperformed by the mean-variance portfolio maximizing expected out-of-sample Sharpe ratio, even though we take a risk-aversion coefficient γ consistent with that implied by the Sharpe ratio criterion (see Footnote 20). This finding shows that our theoretical results directed at the expected out-of-sample Sharpe ratio provide relevant and practically useful complements to prior literature that focuses on portfolio combinations maximizing expected out-of-sample utility.

In Figure 3, we illustrate the previously discussed results by depicting the density of out-of-sample Sharpe ratio, estimated across the 10,000 simulations, for the $1/N$, GMV, tangent, and maximum expected out-of-sample Sharpe ratio portfolio strategies. We depict the densities for a sample size $T = 240$. The figure shows that the out-of-sample Sharpe ratio of the proposed

mean-variance portfolio is not only larger than that of the tangent portfolio on average but is also less volatile because shrinking toward GMV diminishes the exposure to sample mean returns and thus parameter uncertainty. Moreover, the proposed mean-variance portfolio delivers the largest expected out-of-sample Sharpe ratio for *all* datasets.

Finally, the robust mean-variance portfolio that trades off between the expectation and standard deviation of out-of-sample Sharpe ratio as proposed in Section 6 outperforms the mean-variance portfolio maximizing expected out-of-sample Sharpe ratio when estimation risk is large, i.e., when the sample size is $T = 60$. Moreover, when considering the 30IND dataset for which sample mean returns contain less useful information to improve out-of-sample performance, this robust methodology is particularly beneficial and improves performance for *all* considered sample sizes.

7.2.3. Results for simulated data and sample covariance matrix The results obtained when exploiting the sample covariance matrix are reported in Table 3 and are generally consistent with the results exploiting the Ledoit-Wolf covariance matrix. In particular, the proposed portfolio combination maximizing expected out-of-sample Sharpe ratio is overall the best strategy. Still, we observe two main differences with the results under the Ledoit-Wolf covariance matrix.

First, the expected out-of-sample Sharpe ratios are lower when using the sample covariance matrix. Therefore, it is preferable to use the Ledoit-Wolf covariance matrix in combination with the wrong shrinkage intensities that assume the covariance matrix is known, rather than using the sample covariance matrix in combination with the correct shrinkage intensities.

Second, the robust methodology that trades off between the expectation and standard deviation of out-of-sample Sharpe ratio delivers better performance when there is more estimation error in the covariance matrix. Specifically, it outperforms the portfolio designed to maximize expected out-of-sample Sharpe ratio also for a sample size $T = 120$ months. Therefore, shrinking more intensively toward the GMV portfolio is useful to alleviate estimation errors in optimal combination coefficients and improves expected out-of-sample Sharpe ratios for realistic sample sizes.

7.3. Real data

The simulated data used previously comply with the theoretical assumption of iid normal returns. In this section, we turn to real data that generally do not adhere to this assumption, thus providing a more challenging but also practically more relevant environment in which to test our method.

7.3.1. Methodology for real data We consider the same four datasets as for the simulated data in Section 7.2.1. Moreover, we consider eight portfolio strategies, which are the seven ones used in Section 7.2.1 as well as the portfolio strategy proposed by Kircher and Rosch (2021) that maximizes expected out-of-sample Sharpe ratio via bootstrap. This strategy is particularly relevant because it is a data-driven alternative to our approach that maximizes expected out-of-sample

Sharpe ratio under the parametric assumption of iid normal returns. We detail their methodology in Section EC.5 of the e-companion. Finally, we distinguish between the two cases in which the covariance matrix is estimated via the nonlinear shrinkage estimator of Ledoit and Wolf (2020) and via the sample estimator in (1).

For each of the eight considered portfolio strategies, we construct a time series of out-of-sample returns using a classical rebalancing methodology as in DeMiguel, Garlappi, and Uppal (2009b), for instance. Specifically, given a sample size T , we estimate the portfolio on the first T months of the sample, and compute the out-of-sample portfolio return on the next month. Then, we roll forward the estimation window by one month, rebalance the portfolio to obtain the estimated optimal portfolio on the new estimation window, and compute again the out-of-sample return on the next month. We proceed iteratively until the end of the sample is reached. Rebalancing the portfolio every month incurs transaction costs, and we consider both the case with zero transaction costs and the case with proportional transaction costs of 20 basis points as in Kan, Wang, and Zhou (2021b). We consider estimation windows of size $T = 120$ and 240 months.

Finally, we use the time series of out-of-sample returns to compare the portfolios in terms of annualized out-of-sample Sharpe ratio, which is the objective function considered in this paper. We test whether our proposed portfolio strategies that use information about sample mean returns statistically outperform the GMV portfolio that ignores this information. Specifically, we compute two-sided p -values for the difference in out-of-sample Sharpe ratio using the block bootstrap approach of Ledoit and Wolf (2008) with 1000 bootstrap resamples and a block size of five as in DeMiguel, Garlappi, Nogales, and Uppal (2009a). We also report the average monthly turnover induced by rebalancing as in DeMiguel, Garlappi, and Uppal (2009b). Finally, for those portfolio strategies that combine the GMV and tangent portfolios, we report the average weight allocated to the tangent portfolio across all estimation windows, where as explained in Section 5.3 this weight is defined as $\hat{\gamma}_{sr\kappa}$ for a given combination coefficient κ .

7.3.2. Discussion of results for real data Tables 4 and 5 report the out-of-sample portfolio performance for the case with the Ledoit-Wolf and sample covariance matrices, respectively.

First, we corroborate the conclusions in Kirby and Ostdiek (2012) that the reward-to-risk timing strategy outperforms the $1/N$ portfolio, both before and after transaction costs. However, we observe that our two proposed portfolio combinations outperform the reward-to-risk portfolio even after transaction costs, therefore capturing the information in asset returns more effectively.

Second, the out-of-sample performance of the tangent portfolio is not as abysmal as often reported in other papers in the portfolio selection literature. As discussed in Section 7.1, this is due to the lower bound $\gamma_{\min} = 5$ that is imposed on the risk-aversion coefficient implied by the tangent portfolio, γ_{sr} . In particular, there are several cases where the tangent portfolio outperforms GMV.

Third, the two optimal portfolio combinations we propose deliver a better performance than the two portfolios being combined, i.e., the GMV and tangent portfolios. Compared to the tangent portfolio, both shrinkage portfolios outperform in all cases with and without transaction costs, which highlights the necessity of shrinking toward the GMV portfolio to improve out-of-sample performance. Compared to the GMV portfolio, both shrinkage portfolios outperform with and without transaction costs except for two cases, and the outperformance is often statistically significant. The outperformance after transaction costs is noteworthy given that the two shrinkage portfolios have a larger turnover than the GMV portfolio. Moreover, even for the cases in which the GMV portfolio outperforms, the two shrinkage portfolios only yield a small performance loss.

Fourth, the mean-variance portfolio maximizing expected out-of-sample utility delivers smaller out-of-sample Sharpe ratios than the two proposed portfolios designed to maximize it. However, in unreported results, we find that it delivers larger out-of-sample utilities, which highlights the importance of calibrating the combination coefficient with the criterion of interest to the investor.

Fifth, the robust shrinkage portfolio that aims at reducing the standard deviation of out-of-sample Sharpe ratio shrinks more intensively toward the GMV portfolio. Specifically, on average across the 16 considered cases²², the weight allocated to the tangent portfolio is 23% smaller than the weight that maximizes expected out-of-sample Sharpe ratio. This more intensive shrinkage has practical benefits. When using the Ledoit-Wolf covariance matrix, the robust shrinkage portfolio outperforms in 5 out of 8 cases before transaction costs and *in all cases* after transaction costs.²³ The outperformance is more considerable when using the sample covariance matrix that contains larger estimation errors, in which case the robust shrinkage portfolio outperforms in all cases *both* before and after transaction costs. Therefore, we advocate this robust portfolio for the investor who aims to maximize out-of-sample Sharpe ratio, particularly net of transaction costs.

Sixth, we observe that the portfolio of Kircher and Rosch (2021) designed to maximize expected out-of-sample Sharpe ratio via bootstrap successfully achieves this objective. Indeed, it delivers large out-of-sample Sharpe ratios and is the best considered strategy in several cases. The results obtained with this approach are similar to those obtained with our shrinkage portfolio that maximizes expected out-of-sample Sharpe ratio assuming iid Gaussian returns, which is because they share a similar objective. However, the robust shrinkage portfolio that accounts for the standard deviation of out-of-sample Sharpe ratio delivers a better performance in 11 of 16 cases before transaction costs and 13 of 16 cases after transaction costs. Moreover, the bootstrap portfolio systematically generates a larger turnover because it is more heavily exposed to the tangent portfolio.

²² The 16 cases comprise the four datasets, two sample sizes, and two estimators of the covariance matrix.

²³ On average across the four datasets with $T = 120$, the robust shrinkage portfolio delivers an out-of-sample Sharpe ratio of 1.110 and 1.007 before and after transaction costs, respectively. In comparison, the GMV portfolio estimated with the Ledoit-Wolf covariance matrix yields 0.988 and 0.914, which leads to the numbers mentioned in the abstract.

8. Conclusion

Maximizing out-of-sample Sharpe ratio is a widespread objective among investors and academics. However, most approaches aiming for that objective optimize it only indirectly, by estimating the inputs more accurately, robustifying the optimization step, or imposing a number of portfolio weight constraints, among other methods. In this paper, we show how to *directly* optimize out-of-sample Sharpe ratio. We contribute to the literature on portfolio choice in the presence of parameter uncertainty introduced by Kan and Zhou (2007) by analytically identifying portfolio combinations maximizing expected out-of-sample Sharpe ratio. We show that when a risk-free investment is allowed and combination coefficients are unconstrained, this metric is maximized by optimizing expected out-of-sample utility. However, the two criteria differ for portfolios fully invested in risky assets, and we show how to adapt the optimal portfolio combination of Kan, Wang, and Zhou (2021b) to maximize expected out-of-sample Sharpe ratio. Empirically, the resulting optimal portfolios generally deliver better out-of-sample Sharpe ratios than the considered benchmarks. Our sound and direct methodology for optimizing out-of-sample Sharpe ratio paves the way for development in other applications where this objective matters.

Tables

Table 1 List of portfolio strategies considered in the empirical analysis

Name	Description
<i>Portfolio strategies introduced in the paper</i>	
Expected Sharpe	Mean-variance portfolio maximizing expected out-of-sample Sharpe ratio estimated with the sample covariance matrix ($\kappa_{sr}^{\hat{\sigma}}$ in (29)) or the Ledoit-Wolf covariance matrix (κ_{sr}^{σ} in (24))
Robust Sharpe	Mean-variance portfolio maximizing the tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio estimated with the sample covariance matrix ($\kappa_{\lambda}^{\sigma}$ in (39)) or the Ledoit-Wolf covariance matrix ($\kappa_{\lambda}^{\hat{\sigma}}$ in (40)) using a coefficient $\lambda = 1$
<i>Benchmark portfolio strategies</i>	
Bootstrap Sharpe	Bootstrap estimate of the mean-variance portfolio maximizing expected out-of-sample Sharpe ratio developed by Kircher and Rosch (2021) estimated with the sample covariance matrix ($\kappa_{boot}^{\hat{\sigma}}$) or the Ledoit-Wolf covariance matrix (κ_{boot}^{σ})
Expected Utility	Mean-variance portfolio maximizing expected out-of-sample utility developed by Kan, Wang, and Zhou (2021b) estimated with the sample covariance matrix ($\kappa_{\gamma}^{\hat{\sigma}}$ in (27)) or the Ledoit-Wolf covariance matrix (κ_{γ}^{σ} in (22)) using a risk-aversion coefficient $\gamma = 5$
GMV	Global-minimum-variance portfolio in (3) estimated with the sample covariance matrix or the Ledoit-Wolf covariance matrix
Tangent	Tangent portfolio ($\kappa = 1/\gamma_{sr}$ in (17)) estimated with the sample covariance matrix or the Ledoit-Wolf covariance matrix
1/N	Equally weighted portfolio
RTR Timing	Reward-to-risk timing strategy of Kirby and Ostdiek (2012) in (45) estimated with the sample covariance matrix or the Ledoit-Wolf covariance matrix

The table describes the portfolio strategies considered in the empirical analysis of Section 7. The portfolio strategies named Expected Sharpe, Robust Sharpe, Expected Utility, and Tangent depend on a combination coefficient κ whose estimation is described in Section 7.1. The portfolio strategy named Bootstrap Sharpe is described in Section EC.5 of the e-companion. The Ledoit-Wolf covariance matrix is the nonlinear shrinkage estimator proposed by Ledoit and Wolf (2020).

Table 2 Annualized expected out-of-sample Sharpe ratio with simulated data and Ledoit-Wolf covariance matrix

	Estimated (feasible)				Optimal (unfeasible)			
	$T = 60$	$T = 120$	$T = 240$	$T = 480$	$T = 60$	$T = 120$	$T = 240$	$T = 480$
10MOM ($\theta = 1.055$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.844	0.896	0.937	0.975	0.919	0.937	0.963	0.991
Robust Sharpe	0.854	0.896	0.930	0.966	0.902	0.924	0.953	0.984
<i>Benchmark portfolios</i>								
Expected Utility	0.828	0.889	0.936	0.974	0.918	0.937	0.963	0.991
GMV	0.837	0.862	0.874	0.883	0.892	0.892	0.892	0.892
Tangent	0.701	0.806	0.896	0.962	0.668	0.796	0.897	0.965
$1/N$	0.549	0.549	0.549	0.549	0.549	0.549	0.549	0.549
RTR Timing	0.671	0.673	0.672	0.670	0.671	0.675	0.672	0.670
25SBTM ($\theta = 1.180$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.734	0.828	0.914	0.999	0.866	0.909	0.967	1.030
Robust Sharpe	0.738	0.821	0.902	0.989	0.850	0.896	0.958	1.026
<i>Benchmark portfolios</i>								
Expected Utility	0.716	0.825	0.916	1.002	0.866	0.909	0.967	1.030
GMV	0.693	0.737	0.767	0.785	0.806	0.806	0.806	0.806
Tangent	0.579	0.722	0.862	0.980	0.564	0.720	0.870	0.990
$1/N$	0.592	0.592	0.592	0.592	0.592	0.592	0.592	0.592
RTR Timing	0.616	0.619	0.620	0.619	0.616	0.619	0.620	0.619
25OPINV ($\theta = 1.325$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.883	0.981	1.067	1.152	1.045	1.083	1.135	1.191
Robust Sharpe	0.896	0.981	1.060	1.145	1.032	1.073	1.127	1.186
<i>Benchmark portfolios</i>								
Expected Utility	0.811	0.940	1.041	1.123	1.027	1.055	1.100	1.153
GMV	0.872	0.915	0.947	0.968	0.993	0.993	0.993	0.993
Tangent	0.677	0.835	0.992	1.122	0.691	0.866	1.026	1.146
$1/N$	0.752	0.752	0.752	0.752	0.752	0.752	0.752	0.752
RTR Timing	0.778	0.778	0.778	0.777	0.778	0.779	0.778	0.777
30IND ($\theta = 1.030$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.742	0.812	0.859	0.896	0.881	0.888	0.901	0.921
Robust Sharpe	0.773	0.828	0.865	0.896	0.875	0.883	0.896	0.918
<i>Benchmark portfolios</i>								
Expected Utility	0.658	0.765	0.832	0.879	0.879	0.885	0.896	0.913
GMV	0.813	0.834	0.850	0.860	0.872	0.872	0.872	0.872
Tangent	0.448	0.552	0.670	0.787	0.411	0.541	0.674	0.797
$1/N$	0.633	0.633	0.633	0.633	0.633	0.633	0.633	0.633
RTR Timing	0.672	0.674	0.673	0.671	0.673	0.674	0.672	0.671

The table reports the annualized expected out-of-sample Sharpe ratio for the simulated data analysis in Section 7.2.1. Four monthly return datasets are used and the covariance matrix is estimated via the nonlinear shrinkage estimator of Ledoit and Wolf (2020). The seven portfolio strategies are described in Table 1. The expected out-of-sample Sharpe ratios are obtained as an average across 10,000 simulated datasets simulated from a multivariate normal distribution whose mean and covariance matrix are calibrated from monthly returns of each dataset. The sample size T used to estimate the portfolios varies between 60 and 480 months. The quantity θ reported besides the dataset name is the maximum Sharpe ratio attainable without parameter uncertainty. The four columns named 'Estimated (feasible)' use the estimated combination coefficients κ as in Section 7.1, and the four columns named 'Optimal (unfeasible)' consider the optimal combination coefficients κ and the true covariance matrix. The best feasible strategy for each dataset and sample size is depicted in bold.

Table 3 Annualized expected out-of-sample Sharpe ratio with simulated data and sample covariance matrix

	Estimated (feasible)				Optimal (unfeasible)			
	$T = 60$	$T = 120$	$T = 240$	$T = 480$	$T = 60$	$T = 120$	$T = 240$	$T = 480$
10MOM ($\theta = 1.055$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.822	0.886	0.933	0.972	0.846	0.901	0.944	0.982
Robust Sharpe	0.834	0.891	0.932	0.967	0.843	0.897	0.939	0.976
<i>Benchmark portfolios</i>								
Expected Utility	0.817	0.883	0.932	0.972	0.845	0.901	0.944	0.982
GMV	0.822	0.857	0.875	0.883	0.821	0.857	0.874	0.883
Tangent	0.629	0.774	0.884	0.958	0.575	0.747	0.874	0.955
$1/N$	0.549	0.549	0.549	0.549	0.549	0.549	0.549	0.549
RTR Timing	0.671	0.674	0.672	0.670	0.671	0.674	0.672	0.670
25SBTM ($\theta = 1.180$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.645	0.796	0.901	0.993	0.668	0.813	0.917	1.005
Robust Sharpe	0.652	0.797	0.896	0.986	0.665	0.809	0.912	1.001
<i>Benchmark portfolios</i>								
Expected Utility	0.651	0.795	0.903	0.996	0.665	0.810	0.916	1.004
GMV	0.621	0.721	0.766	0.786	0.621	0.721	0.765	0.786
Tangent	0.419	0.643	0.829	0.966	0.373	0.605	0.808	0.958
$1/N$	0.592	0.592	0.592	0.592	0.592	0.592	0.592	0.592
RTR Timing	0.616	0.619	0.620	0.619	0.616	0.619	0.620	0.619
25OPINV ($\theta = 1.325$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.773	0.945	1.057	1.149	0.807	0.969	1.076	1.162
Robust Sharpe	0.784	0.950	1.056	1.144	0.805	0.966	1.072	1.158
<i>Benchmark portfolios</i>								
Exp Utility	0.767	0.927	1.040	1.127	0.806	0.962	1.057	1.132
GMV	0.766	0.889	0.942	0.968	0.768	0.888	0.942	0.969
Tangent	0.486	0.749	0.956	1.110	0.440	0.724	0.949	1.110
$1/N$	0.752	0.752	0.752	0.752	0.752	0.752	0.752	0.752
RTR Timing	0.778	0.779	0.778	0.777	0.778	0.779	0.778	0.777
30IND ($\theta = 1.030$)								
<i>Proposed portfolios</i>								
Expected Sharpe	0.580	0.745	0.830	0.884	0.630	0.773	0.845	0.893
Robust Sharpe	0.594	0.755	0.836	0.886	0.630	0.773	0.844	0.892
<i>Benchmark portfolios</i>								
Expected Utility	0.590	0.729	0.816	0.872	0.629	0.773	0.844	0.888
GMV	0.624	0.759	0.818	0.845	0.624	0.759	0.818	0.845
Tangent	0.244	0.445	0.618	0.766	0.203	0.406	0.600	0.759
$1/N$	0.633	0.633	0.633	0.633	0.633	0.633	0.633	0.633
RTR Timing	0.673	0.674	0.672	0.671	0.673	0.674	0.672	0.671

The table reports the annualized expected out-of-sample Sharpe ratio for the simulated data analysis in Section 7.2.1. Four monthly return datasets are used and the covariance matrix is estimated via the sample estimator in (1). The seven portfolio strategies are described in Table 1. The expected out-of-sample Sharpe ratios are obtained as an average across 10,000 simulated datasets simulated from a multivariate normal distribution whose mean and covariance matrix are calibrated from monthly returns of each dataset. The sample size T used to estimate the portfolios varies between 60 and 480 months. The quantity θ reported besides the dataset name is the maximum Sharpe ratio attainable without parameter uncertainty. The four columns named 'Estimated (feasible)' use the estimated combination coefficients κ as in Section 7.1, and the four columns named 'Optimal (unfeasible)' consider the optimal combination coefficients κ and the true covariance matrix. The best feasible strategy for each dataset and sample size is depicted in bold.

Table 4 Results for real data with Ledoit-Wolf covariance matrix

	$T = 120$				$T = 240$			
	SR	Net SR	Turnover	$\hat{\gamma}_{sr\kappa}$	SR	Net SR	Turnover	$\hat{\gamma}_{sr\kappa}$
10MOM								
<i>Proposed portfolios</i>								
Expected Sharpe	1.069**	0.991*	0.652	0.473	1.194*	1.137	0.491	0.590
Robust Sharpe	1.068***	1.000**	0.517	0.324	1.204**	1.153*	0.399	0.446
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	1.071	0.994	0.651	0.494	1.199	1.141	0.496	0.618
Expected Utility	1.058	0.962	1.099	0.875	1.133	1.061	0.787	0.951
GMV	0.862	0.824	0.242	0	0.986	0.960	0.151	0
Tangent	1.007	0.903	1.117	1	1.162	1.091	0.715	1
1/N	0.667	0.664	0.017	/	0.744	0.741	0.017	/
RTR Timing	0.761	0.757	0.025	/	0.859	0.857	0.013	/
25SBTM								
<i>Proposed portfolios</i>								
Expected Sharpe	1.141**	0.998	0.909	0.352	1.333**	1.228	0.621	0.461
Robust Sharpe	1.138**	1.008*	0.782	0.261	1.331**	1.235*	0.542	0.364
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	1.143	0.999	0.871	0.336	1.343	1.240	0.593	0.435
Expected Utility	1.117	0.911	2.148	1.159	1.159	1.018	1.494	1.270
GMV	1.007	0.912	0.523	0	1.182	1.116	0.340	0
Tangent	0.950	0.735	2.129	1	1.221	1.073	1.162	1
1/N	0.712	0.710	0.017	/	0.806	0.804	0.017	/
RTR Timing	0.750	0.747	0.021	/	0.850	0.848	0.014	/
25OPINV								
<i>Proposed portfolios</i>								
Expected Sharpe	1.337*	1.173	0.930	0.348	1.169	1.062	0.651	0.412
Robust Sharpe	1.345**	1.197*	0.801	0.259	1.193	1.094	0.577	0.319
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	1.335	1.181	0.863	0.329	1.200	1.096	0.618	0.419
Expected Utility	1.107	0.865	2.403	1.161	0.747	0.599	1.734	1.154
GMV	1.206	1.098	0.549	0	1.216	1.146	0.363	0
Tangent	1.010	0.764	2.226	1	1.019	0.876	1.156	1
1/N	0.792	0.790	0.016	/	0.737	0.735	0.016	/
RTR Timing	0.827	0.824	0.019	/	0.774	0.772	0.012	/
30IND								
<i>Proposed portfolios</i>								
Expected Sharpe	0.868	0.794	0.384	0.187	0.925	0.882	0.209	0.147
Robust Sharpe	0.889	0.822	0.339	0.119	0.936	0.895	0.195	0.085
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	0.831	0.759	0.379	0.193	0.906	0.863	0.210	0.160
Expected Utility	0.709	0.602	0.734	0.546	0.820	0.760	0.316	0.350
GMV	0.878	0.821	0.278	0	0.950	0.912	0.181	0
Tangent	0.435	0.305	1.350	1	0.624	0.538	0.650	1
1/N	0.742	0.739	0.027	/	0.809	0.805	0.027	/
RTR Timing	0.777	0.772	0.032	/	0.844	0.841	0.021	/

The table reports the annualized out-of-sample Sharpe ratio for the real data analysis in Section 7.3. Four monthly return datasets are used and the covariance matrix is estimated via the nonlinear shrinkage estimator of Ledoit and Wolf (2020). The eight portfolio strategies are described in Table 1. The sample size T used to estimate the portfolios is 120 or 240 months. The net Sharpe ratio (net SR) is adjusted using proportional transaction costs of 20 basis points. The turnover is computed on a monthly basis. The quantity $\hat{\gamma}_{sr\kappa}$ is the average weight allocated to the tangent portfolio across all estimation windows. The stars *, **, *** in superscript of the SR and net SR of the two proposed strategies mean that the two-sided p -value for the difference in performance relative to the GMV portfolio is smaller than 10%, 5%, and 1%. The p -values are computed from Ledoit and Wolf (2008). The best portfolio for each dataset and sample size is depicted in bold.

Table 5 Results for real data with sample covariance matrix

	$T = 120$				$T = 240$			
	SR	Net SR	Turnover	$\hat{\gamma}_{sr}\kappa$	SR	Net SR	Turnover	$\hat{\gamma}_{sr}\kappa$
10MOM								
<i>Proposed portfolios</i>								
Expected Sharpe	1.057**	0.973*	0.691	0.426	1.183*	1.123	0.499	0.561
Robust Sharpe	1.060***	0.985**	0.567	0.330	1.194**	1.141*	0.415	0.453
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	1.053	0.968	0.724	0.473	1.180	1.120	0.517	0.605
Expected Utility	1.053	0.955	1.048	0.727	1.133	1.061	0.754	0.869
GMV	0.865	0.821	0.277	0	0.985	0.957	0.161	0
Tangent	0.969	0.852	1.323	1	1.139	1.063	0.772	1
1/N	0.667	0.664	0.017	/	0.744	0.741	0.017	/
RTR Timing	0.761	0.757	0.025	/	0.859	0.857	0.013	/
25SBTM								
<i>Proposed portfolios</i>								
Expected Sharpe	1.077	0.897	1.169	0.273	1.323**	1.206	0.697	0.409
Robust Sharpe	1.088**	0.918	1.062	0.231	1.328**	1.218*	0.632	0.348
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	1.072	0.885	1.203	0.288	1.324	1.205	0.701	0.413
Expected Utility	1.099	0.871	2.065	0.716	1.189	1.039	1.425	1.013
GMV	0.980	0.846	0.771	0	1.185	1.107	0.410	0
Tangent	0.777	0.493	3.386	1	1.174	1.002	1.438	1
1/N	0.712	0.710	0.017	/	0.806	0.804	0.017	/
RTR Timing	0.750	0.747	0.021	/	0.850	0.848	0.014	/
25OPINV								
<i>Proposed portfolios</i>								
Expected Sharpe	1.330	1.122	1.249	0.270	1.160	1.039	0.755	0.366
Robust Sharpe	1.334*	1.138	1.143	0.230	1.176	1.061	0.694	0.314
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	1.335	1.133	1.224	0.278	1.184	1.063	0.744	0.392
Expected Utility	1.154	0.887	2.414	0.717	0.806	0.649	1.675	0.921
GMV	1.212	1.059	0.823	0	1.192	1.107	0.447	0
Tangent	0.942	0.626	3.706	1	0.978	0.810	1.477	1
1/N	0.792	0.790	0.016	/	0.737	0.735	0.016	/
RTR Timing	0.827	0.825	0.019	/	0.774	0.772	0.012	/
30IND								
<i>Proposed portfolios</i>								
Expected Sharpe	0.796	0.692	0.564	0.137	0.887	0.834	0.260	0.128
Robust Sharpe	0.802	0.703	0.531	0.115	0.891	0.840	0.252	0.107
<i>Benchmark portfolios</i>								
Bootstrap Sharpe	0.748	0.642	0.586	0.156	0.867	0.814	0.265	0.156
Expected Utility	0.738	0.612	0.788	0.302	0.819	0.753	0.345	0.266
GMV	0.779	0.693	0.447	0	0.902	0.855	0.230	0
Tangent	0.335	0.147	2.447	1	0.574	0.470	0.853	1
1/N	0.742	0.739	0.027	/	0.809	0.805	0.027	/
RTR Timing	0.779	0.773	0.032	/	0.844	0.841	0.021	/

The table reports the annualized out-of-sample Sharpe ratio for the real data analysis in Section 7.3. Four monthly return datasets are used and the covariance matrix is estimated via the sample estimator in (1). The eight portfolio strategies are described in Table 1. The sample size T used to estimate the portfolios is 120 or 240 months. The net Sharpe ratio (net SR) is adjusted using proportional transaction costs of 20 basis points. The turnover is computed on a monthly basis. The quantity $\hat{\gamma}_{sr}\kappa$ is the average weight allocated to the tangent portfolio across all estimation windows. The stars *, **, *** in superscript of the SR and net SR of the two proposed strategies mean that the two-sided p -value for the difference in performance relative to the GMV portfolio is smaller than 10%, 5%, and 1%. The p -values are computed from Ledoit and Wolf (2008). The best portfolio for each dataset and sample size is depicted in bold.

References

- Ao M, Li Y, Zheng X (2019) Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies* 32(7):2890–2919.
- Barroso P, Saxena K (2021) Lest we forget: Using out-of-sample forecast errors in portfolio optimization. *The Review of Financial Studies* (forthcoming).
- Bodnar T, Okhrin Y, Parolya N (2021) Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business and Economic Statistics* (forthcoming).
- Branger N, Lučivjanská K, Weissensteiner A (2019) Optimal granularity for portfolio choice. *Journal of Empirical Finance* 50:125–146.
- Candelon B, Hurlin C, Tokpavi S (2012) Sampling error and double shrinkage estimation of minimum variance portfolios. *Journal of Empirical Finance* 19(4):511–527.
- Chakrabarti D (2021) Parameter-free robust optimization for the maximum-sharpe portfolio problem. *European Journal of Operational Research* 293(1):388–399.
- DeMiguel V, Garlappi L, Nogales F, 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, 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, Martin-Utrera A, Nogales F (2013) Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking and Finance* 37(8):3018–3034.
- DeMiguel V, Martin-Utrera A, Nogales F (2015) Parameter uncertainty in multiperiod portfolio optimization with transaction costs. *Journal of Financial and Quantitative Analysis* 50(6):1443–1471.
- Frahm G, Memmel C (2010) Dominating estimators for minimum-variance portfolios. *Journal of Econometrics* 159(2):289–302.
- Garlappi L, Uppal R, Wang T (2007) Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies* 20(1):41–81.
- Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58(4):1651–1684.
- Kan R, Smith D (2008) The distribution of the sample minimum-variance frontier. *Management Science* 54(7):1364–1380.
- Kan R, Wang X (2021) Optimal portfolio choice with benchmark. *Rotman School of Management Working Paper No. 3760640*.
- Kan R, Wang X, Zheng X (2021a) In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. *Available at SSRN 3454628*.

- Kan R, Wang X, Zhou G (2021b) Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science* (forthcoming).
- Kan R, Zhou G (2007) Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis* 42(3):621–656.
- Kelly B, Malamud S, Zhou K (2021) The virtue of complexity in machine learning portfolios. *Swiss Finance Institute Research Paper* 21-90.
- Kirby C, Ostdiek B (2012) It's all in the timing: Simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis* 47(2):437–467.
- Kircher F, Rosch D (2021) A shrinkage approach for sharpe ratio optimal portfolios with estimation risks. *Journal of Banking and Finance* 133.
- Lassance N, DeMiguel V, Vrins F (2021a) Optimal portfolio diversification via independent component analysis. *Operations Research* (forthcoming).
- Lassance N, Martin-Utrera A, Simaan M (2021b) A robust approach to optimal portfolio choice with parameter uncertainty. *Available at SSRN 3855546*.
- Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2):365–411.
- Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance* 15(5):850–859.
- Ledoit O, 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, Wolf M (2020) Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics* 48(5):3043–3065.
- Martellini L, Ziemann V (2010) Improved estimates of higher-order comoments and implications for portfolio selection. *The Review of Financial Studies* 23(4):1467–1502.
- Paulsen D, Söhl J (2020) Noise fit, estimation error and a sharpe information criterion. *Quantitative Finance* 20(6):1027–1043.
- Rapponi V, Uppal R, Zaffaroni P (2021) Robust portfolio choice. *Available at SSRN 3933063*.
- Sharpe W (1966) Mutual fund performance. *The Journal of Business* 39(1):119–138.
- Sharpe W (1994) The sharpe ratio. *The Journal of Portfolio Management* 21(1):49–58.
- Tu J, Zhou G (2011) Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99(1):204–215.
- Yuan M, Zhou G (2021) Why naïve 1/N diversification is not so naive, and how to beat it? *Available at SSRN 3991279*.
- Zhou G (2008) On the fundamental law of active portfolio management: What happens if our estimates are wrong? *The Journal of Portfolio Management* 34(4):26–33.

Electronic companion to “Maximizing the Out-of-Sample Sharpe Ratio”

This electronic companion to the main paper contains seven sections. Section EC.1 provides an exact semi-analytical expression for the mean and variance out-of-sample Sharpe ratio in the special case where portfolio weights are normally distributed. Section EC.2 derives closed-form expressions for the two approximations to the expected out-of-sample Sharpe ratio we consider in the paper. Section EC.3 applies our theoretical results to the portfolio combination considered by Tu and Zhou (2011). Section EC.4 studies two approximations to the variance of out-of-sample Sharpe ratio. Section EC.5 describes the bootstrap portfolio strategy of Kircher and Rosch (2021). Section EC.6 reports the out-of-sample performance obtained with improved estimators of the expectation and variance of out-of-sample Sharpe ratio. Finally, Section EC.7 reports the proofs of all results.

EC.1. Exact expressions for moments of out-of-sample Sharpe ratio

In this section, we explain how to exactly evaluate the expected out-of-sample Sharpe ratio, $\mathcal{S}(\hat{w})$ in (5), and the variance of out-of-sample Sharpe ratio, $\mathcal{V}(\hat{w})$ in (31), for the special case where $\hat{w}$ is normally distributed, $\hat{w} \sim \mathcal{N}(\mathbb{E}[\hat{w}], \mathbb{V}[\hat{w}])$.²⁴

EC.1.1. Expected out-of-sample Sharpe ratio

Denote by r the rank of $\mathbb{V}[\hat{w}]$, where $r \leq N$. Let $\mathbb{V}[\hat{w}]$ be decomposed into $\mathbb{V}[\hat{w}] = \mathbf{C}\mathbf{C}'$, where $\mathbf{C}$ is of size $N \times r$. Let $U = \hat{w}'\mu$ and $V = \hat{w}'\Sigma\hat{w}$ denote the out-of-sample mean return and variance. Then, the joint moment generating function of U and V is (Mathai and Provost 1992, Theorem 3.2c.4)

$$M_{U,V}(s, t) = |\mathbf{I}_r - 2t\mathbf{C}'\Sigma\mathbf{C}|^{-1/2} \exp\left(s\mathbb{E}[\hat{w}']\mu + t\mathbb{E}[\hat{w}']\Sigma\mathbb{E}[\hat{w}] + 0.5\beta'(\mathbf{I}_r - 2t\mathbf{C}'\Sigma\mathbf{C})^{-1}\beta\right), \quad (\text{EC.1})$$

where

$$\beta = s\mathbf{C}'\mu + 2t\mathbf{C}'\Sigma\mathbb{E}[\hat{w}]. \quad (\text{EC.2})$$

Let $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$ be the eigenvalue decomposition of the matrix $\mathbf{C}'\Sigma\mathbf{C}$. Define further the vectors $u = \mathbf{Q}'\mathbf{C}'\mu$ and $b = -2t\mathbf{Q}'\mathbf{C}'\Sigma\mathbb{E}[\hat{w}]$. Then, applying Meng (2005, Lemma 1), we finally obtain the expected out-of-sample Sharpe ratio of the portfolio $\hat{w}$ as

$$\begin{aligned} \mathcal{S}(\hat{w}) &= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{\partial M_{U,V}(s, -t)}{\partial s} \Big|_{s=0} t^{-1/2} dt \\ &= \frac{1}{\sqrt{\pi}} \int_0^\infty |\mathbf{I}_r + 2t\mathbf{\Lambda}|^{-1/2} \exp\left(-t\mathbb{E}[\hat{w}']\Sigma\mathbb{E}[\hat{w}] + 0.5b'(\mathbf{I}_r + 2t\mathbf{\Lambda})^{-1}b\right) [\mathbb{E}[\hat{w}']\mu + u'(\mathbf{I}_r + 2t\mathbf{\Lambda})^{-1}b] t^{-1/2} dt. \end{aligned} \quad (\text{EC.3})$$

²⁴ We are thankful to Raymond Kan for the material presented in this section.

EC.1.2. Variance of out-of-sample Sharpe ratio

The variance of out-of-sample Sharpe ratio is

$$\mathcal{V}(\hat{w}) = \mathbb{E} \left[\frac{(\hat{w}'\mu)^2}{\hat{w}'\Sigma\hat{w}} \right] - (\mathcal{S}(\hat{w}))^2,$$

where $\mathcal{S}(\hat{w})$ can be evaluated exactly from (EC.3). Moreover, applying Meng (2005, Lemma 1), the first term in $\mathcal{V}(\hat{w})$ is

$$\begin{aligned} \mathbb{E} \left[\frac{(\hat{w}'\mu)^2}{\hat{w}'\Sigma\hat{w}} \right] &= \int_0^\infty \frac{\partial^2 M_{U,V}(s, -t)}{\partial s^2} \Big|_{s=0} dt \\ &= \int_0^\infty |\mathbf{I}_r + 2t\mathbf{\Lambda}|^{-1/2} \exp(-t\mathbb{E}[\hat{w}]'\Sigma\mathbb{E}[\hat{w}] + 0.5b'(\mathbf{I}_r + 2t\mathbf{\Lambda})^{-1}b) \\ &\quad \times [u'(\mathbf{I}_r + 2t\mathbf{\Lambda})^{-1}u] [\mathbb{E}[\hat{w}]'\mu + u'(\mathbf{I}_r + 2t\mathbf{\Lambda})^{-1}b]^2 dt. \end{aligned} \quad (\text{EC.4})$$

EC.2. Approximations to expected out-of-sample Sharpe ratio

In this section, we derive analytical expressions for the refined approximation to the expected out-of-sample Sharpe ratio proposed in Section 3, $\tilde{\mathcal{S}}_1$, and compare it to $\tilde{\mathcal{S}}_0$ that we use throughout the paper. We focus on fully invested sample mean-variance portfolios as in Section 5 because this is the setting we consider in the empirical analysis. We distinguish the cases where the covariance matrix is assumed known or estimated via the sample estimator in (1).

EC.2.1. Known covariance matrix

In the next proposition, we consider the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ defined in (20) and derive the approximation $\tilde{\mathcal{S}}_1$ to its expected out-of-sample Sharpe ratio. The approximation $\tilde{\mathcal{S}}_0$ is given by Equation (23).

PROPOSITION EC.1. *When the covariance matrix Σ is known, the approximation $\tilde{\mathcal{S}}_1$ in (7) to the expected out-of-sample Sharpe ratio of the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ in (20) is*

$$\begin{aligned} \tilde{\mathcal{S}}_1(\hat{w}_{mv}^\sigma(\kappa)) &= \frac{\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\mu]}{\sqrt{\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)]}} + \frac{3\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\mu]\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)]}{8(\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)])^{5/2}} - \frac{\text{Cov}[\hat{w}_{mv}^\sigma(\kappa)'\mu, \hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)]}{2(\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)])^{3/2}}, \end{aligned} \quad (\text{EC.5})$$

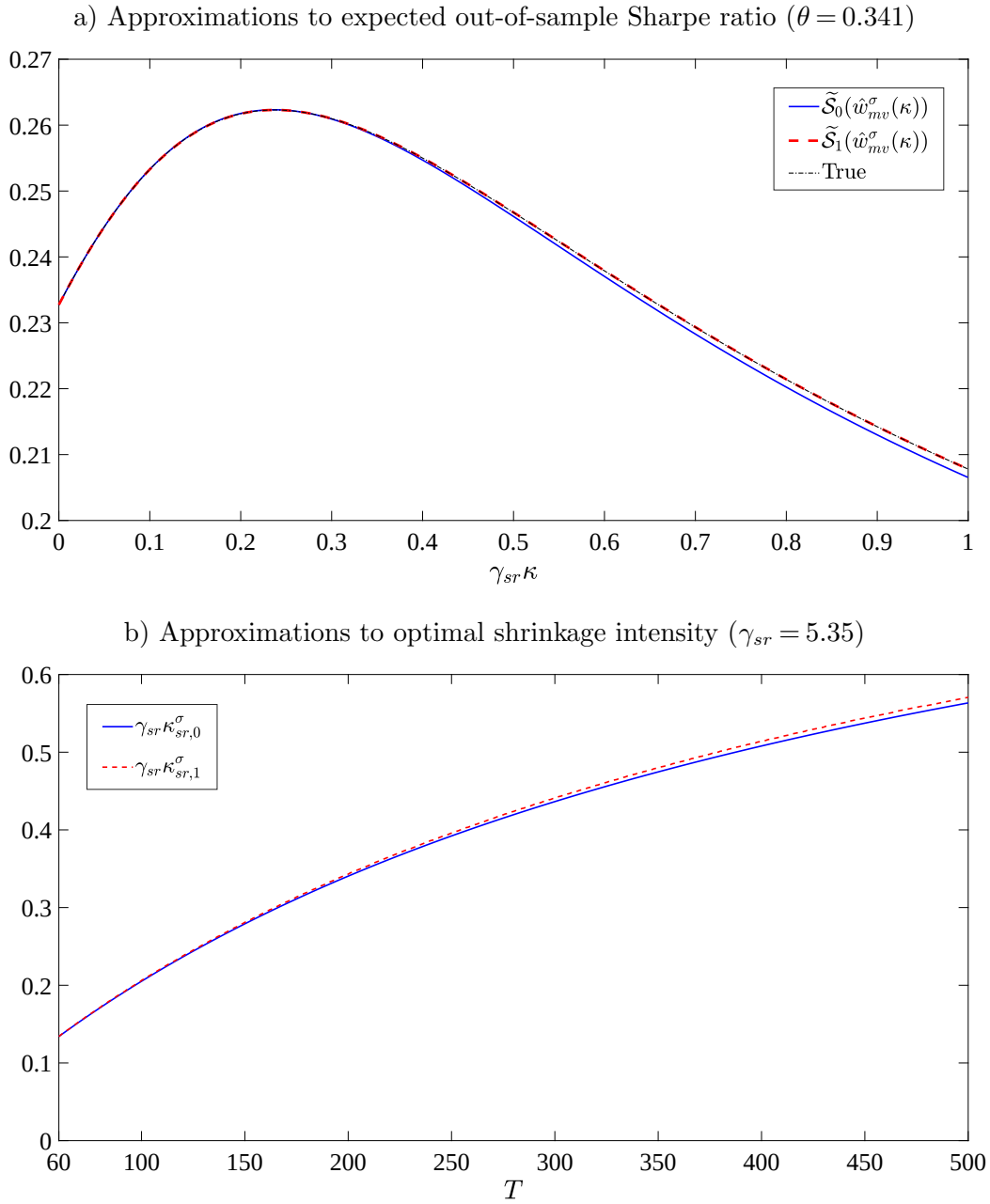
where

$$\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\mu] = \mu_g + \kappa\psi^2, \quad (\text{EC.6})$$

$$\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)] = \sigma_g^2 + \kappa^2 \left(\psi^2 + \frac{N-1}{T} \right), \quad (\text{EC.7})$$

$$\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)] = \frac{2\kappa^4}{T} \left(2\psi^2 + \frac{N-1}{T} \right), \quad (\text{EC.8})$$

$$\text{Cov}[\hat{w}_{mv}^\sigma(\kappa)'\mu, \hat{w}_{mv}^\sigma(\kappa)'\Sigma\hat{w}_{mv}^\sigma(\kappa)] = 2\kappa^3\psi^2/T. \quad (\text{EC.9})$$

Figure EC.1 Approximations to expected out-of-sample Sharpe ratio when the covariance matrix is known

Notes. Both panels are constructed with population values of the mean and covariance matrix of asset returns calibrated from a dataset of monthly returns on 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021. The maximum Sharpe ratio without parameter uncertainty is $\theta = 0.341$. Panel a) depicts the expected out-of-sample Sharpe ratio, according to the two approximations in Section 3, of the fully invested sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ in (20). The true expected out-of-sample Sharpe ratio found analytically with the method presented in Section EC.1 is also depicted. The covariance matrix is assumed known and the portfolios are estimated with a sample size of $T = 120$ monthly observations. The horizontal axis is the intensity with which the global-minimum-variance portfolio is shrunk toward the sample tangent portfolio obtained for a coefficient $\kappa = 1/\gamma_{sr}$ defined in (17). Panel b) depicts the optimal shrinkage intensity maximizing expected out-of-sample Sharpe ratio according to the two approximations, as a function of the sample size T between 60 and 500.

Proposition EC.1 derives a more accurate approximation to the expected out-of-sample Sharpe ratio than $\tilde{\mathcal{S}}_0$ in Proposition 3. However, this approximation no longer deliver a closed-form solution for the optimal combination coefficient κ , which hinders analytical insights.

In Figure EC.1, we use the closed-form expression provided in Proposition EC.1 to compare the two approximations as well as the resulting optimal combination coefficients, which we denote by $\kappa_{sr,0}^\sigma$ and $\kappa_{sr,1}^\sigma$ ($\kappa_{sr,0}^\sigma$ corresponds to κ_{sr}^σ in (23)). We use the same dataset as in Figure 1; we find in unreported results that the conclusions are similar when using the three other datasets considered in the paper. Panel a) considers a sample size of $T = 120$ and depicts the two approximations to the expected out-of-sample Sharpe ratio as a function of the shrinkage intensity, $\gamma_{sr}\kappa$. We compare these approximations to the true expected out-of-sample Sharpe ratio that we find analytically with the method presented in Section EC.1, which can be used because the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ is normally distributed. We observe that the approximation $\tilde{\mathcal{S}}_0$ and $\tilde{\mathcal{S}}_1$ differ only very slightly. Moreover, the approximation $\tilde{\mathcal{S}}_1$ is indistinguishable from the true expected out-of-sample Sharpe ratio. Then, Panel b) depicts the two approximations to the optimal shrinkage intensity as a function of the sample size T and shows that the two intensities agree very closely.

EC.2.2. Unknown covariance matrix

In the next proposition, we consider the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ defined in (26) and derive the approximation $\tilde{\mathcal{S}}_1$ to its expected out-of-sample Sharpe ratio. The approximation $\tilde{\mathcal{S}}_0$ is given by Equation (28).

PROPOSITION EC.2. *Let $T \geq N + 8$. When the covariance matrix Σ is estimated by $\hat{\Sigma}$ in (1), the approximation $\tilde{\mathcal{S}}_1$ in (7) to the expected out-of-sample Sharpe ratio of the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ in (26) is*

$$\begin{aligned} \tilde{\mathcal{S}}_1(\hat{w}_{mv}^\sigma(\kappa)) = & \frac{\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \mu]}{\sqrt{\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)]}} + \frac{3\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \mu] \mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)]}{8(\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)])^{5/2}} - \frac{\text{Cov}[\hat{w}_{mv}^\sigma(\kappa)' \mu, \hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)]}{2(\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)])^{3/2}}, \end{aligned} \quad (\text{EC.10})$$

where

$$\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \mu] = \mu_g + \kappa \frac{T}{T - N - 1} \psi^2, \quad (\text{EC.11})$$

$$\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)] = \frac{T-2}{T-N-1} \sigma_g^2 + \kappa^2 \frac{T^2(T-2)}{(T-N)(T-N-1)(T-N-3)} \left(\psi^2 + \frac{N-1}{T} \right), \quad (\text{EC.12})$$

$$\begin{aligned} \mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)] = & 2\sigma_g^4 \frac{(N-1)(T-2)}{(T-N-1)^2(T-N-3)} \\ & + 4\kappa^2 \sigma_g^2 \frac{T(T-2)(T+N-3)(T\psi^2 + N-1)}{(T-N)(T-N-1)^2(T-N-3)(T-N-5)} \end{aligned}$$

$$+ 2\kappa^4 \frac{T^2(T-2)C(T, N, \psi^2)}{(T-N)^2(T-N-1)^2(T-N-2)(T-N-3)^2(T-N-5)(T-N-7)}, \quad (\text{EC.13})$$

$$\begin{aligned} \text{Cov}[\hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \mu, \hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \Sigma \hat{w}_{mv}^{\hat{\sigma}}(\kappa)] &= 2\kappa \sigma_g^2 \psi^2 \frac{T(T-2)}{(T-N-1)^2(T-N-3)} \\ &+ 2\kappa^3 \psi^2 \frac{T^2(T-2)(T+N-3+2T\psi^2)}{(T-N)(T-N-1)^2(T-N-3)(T-N-5)}, \end{aligned} \quad (\text{EC.14})$$

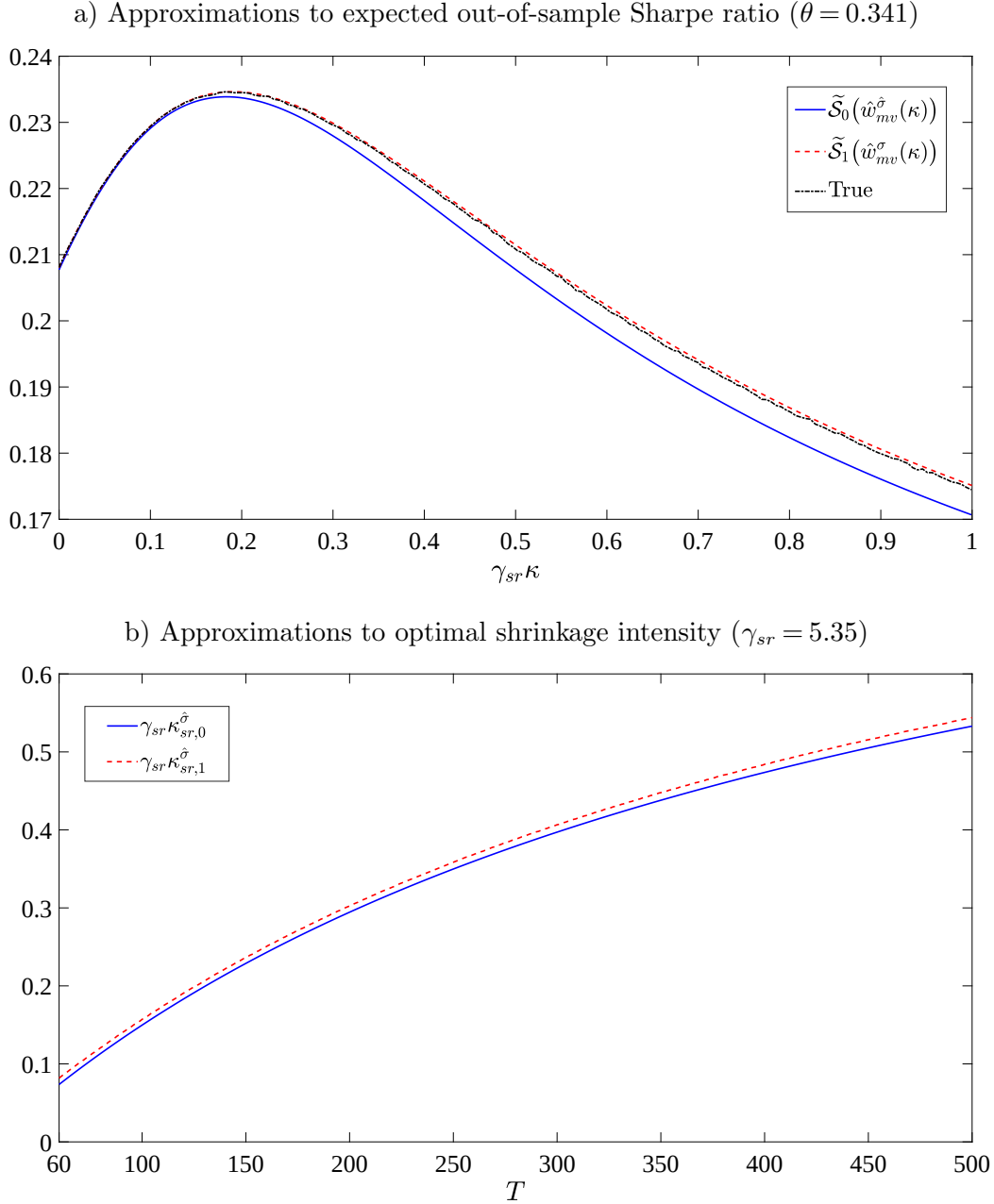
with

$$\begin{aligned} C(T, N, \psi^2) &= (2T\psi^2 + N - 1)(N^4 + N^3T - 3N^3 - 4N^2T^2 + 22N^2T - 31N^2 + NT^3 \\ &\quad - 7NT^2 + 13NT - 5N + T^4 - 12T^3 + 53T^2 - 100T + 70) + T^2\psi^4(N^3 \\ &\quad + 2N^2T - 6N^2 - 7NT^2 + 40NT - 53N + 4T^3 - 34T^2 + 88T - 70). \end{aligned}$$

Similar to the case in which the covariance is known, Proposition EC.2 derives a more accurate approximation to the expected out-of-sample Sharpe ratio than $\tilde{\mathcal{S}}_0$ in Proposition 4, but this approximation no longer delivers a closed-form solution for the optimal combination coefficient.

In Figure EC.2, we use the closed-form expression provided in Proposition EC.2 to compare the two approximations as well as the resulting optimal combination coefficients, which we denote by $\kappa_{sr,0}^{\hat{\sigma}}$ and $\kappa_{sr,1}^{\hat{\sigma}}$ ($\kappa_{sr,0}^{\hat{\sigma}}$ corresponds to $\kappa_{sr}^{\hat{\sigma}}$ in (28)). We use the same dataset as in Figure 1; we find in unreported results that the conclusions are similar when using the three other datasets considered in the paper. Panel a) considers a sample size of $T = 120$ and depicts the two approximations to the expected out-of-sample Sharpe ratio as a function of the shrinkage intensity, $\gamma_{sr}\kappa$. We compare these approximations to the true expected out-of-sample Sharpe ratio that we find numerically using 250,000 Monte Carlo simulations. We observe that the two approximations are not as close to one another compared to the case in which the covariance matrix is known in Figure EC.1. Still, the difference between the simple approximation $\tilde{\mathcal{S}}_0$ and the most accurate one we consider, $\tilde{\mathcal{S}}_1$, is small even when the weight $\gamma_{sr}\kappa$ on the tangent portfolio is large. For example, when considering the tangent portfolio ($\gamma_{sr}\kappa = 1$), the expected out-of-sample Sharpe ratio is 0.171 according to $\tilde{\mathcal{S}}_0$ and 0.176 according to $\tilde{\mathcal{S}}_1$. Also, the approximation $\tilde{\mathcal{S}}_1$ is very close to the true expected out-of-sample Sharpe ratio. Then, Panel b) depicts the two approximations to the optimal shrinkage intensity as a function of the sample size T and shows that the approximation $\tilde{\mathcal{S}}_0$ delivers an optimal shrinkage intensity that is smaller but very close to that obtained with the refined approximation, $\tilde{\mathcal{S}}_1$.²⁵

²⁵ Both approximations $\tilde{\mathcal{S}}_0$ and $\tilde{\mathcal{S}}_1$ converge to the true expected out-of-sample Sharpe ratio as the sample size $T \rightarrow \infty$. Thus, we find in unreported results that they become more similar to one another and more accurate as the sample size increases, e.g., for $T = 240$ that we use in the empirical analysis.

Figure EC.2 Approximations to expected out-of-sample Sharpe ratio when the covariance matrix is unknown

Note. Both panels are constructed with population values of the mean and covariance matrix of asset returns calibrated from a dataset of monthly returns on 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021. The maximum Sharpe ratio without parameter uncertainty is $\theta = 0.341$. Panel a) depicts the expected out-of-sample Sharpe ratio, according to the two approximations in Section 3, of the fully invested sample mean-variance portfolio $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ in (26). The true expected out-of-sample Sharpe ratio found numerically using 250,000 Monte Carlo simulations is also depicted. The covariance matrix is estimated with the sample estimator in (1) and the portfolios are estimated with a sample size of $T = 120$ monthly observations. The horizontal axis is the intensity with which the global-minimum-variance portfolio is shrunk toward the sample tangent portfolio obtained for a coefficient $\kappa = 1/\gamma_{sr}$ defined in (17). Panel b) depicts the optimal shrinkage intensity maximizing expected out-of-sample Sharpe ratio according to the two approximations, as a function of the sample size T between 60 and 500.

EC.3. Portfolio combination of Tu and Zhou (2011)

In this section, we study the portfolio combination considered by Tu and Zhou (2011), which combines the risk-free asset, the tangent portfolio, and the equally weighted ($1/N$) portfolio. First, we show that when the two combination coefficients are constrained to sum to one as imposed by Tu and Zhou (2011), the result in Proposition 1 of the paper does not apply: maximizing expected out-of-sample utility does not maximize expected out-of-sample Sharpe ratio. Then, we extend the analysis by Tu and Zhou to allow for unconstrained combination coefficients, and show that in this case the result in Proposition 1 does apply. For simplicity of exposition, we assume the covariance matrix Σ is known; the same insights hold when it is estimated via the sample estimator in (1).

EC.3.1. Constrained combination coefficients

The three-fund portfolio considered by Tu and Zhou (2011) is

$$\widehat{w}(\kappa) = \kappa_1 \widehat{w}_1 + \kappa_2 \widehat{w}_2, \quad \text{with} \quad \kappa' \iota = \kappa_1 + \kappa_2 = 1,$$

where the two portfolios of risky assets are

$$\widehat{w}_1 = \frac{1}{\gamma} \Sigma^{-1} \widehat{\mu} \quad \text{and} \quad \widehat{w}_2 = w_{ew} = \iota/N.$$

We introduce the following useful notation: $\mu_{ew} = w'_{ew} \mu$, $\sigma_{ew}^2 = w'_{ew} \Sigma w_{ew}$, $\theta_{ew} = \mu_{ew} / \sigma_{ew}$, $\psi_{ew}^2 = \theta^2 - \theta_{ew}^2$, and $\delta_{ew} = \theta^2 / (2\gamma) - (\mu_{ew} - \frac{\gamma}{2} \sigma_{ew}^2)$, where the latter is the difference between the maximum utility without parameter uncertainty and the utility of the $1/N$ portfolio.

The vector η in (9) and the matrix $\mathbf{S}$ in (10) are

$$\eta = \begin{pmatrix} \theta^2 / \gamma \\ \mu_{ew} \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \frac{1}{\gamma^2} (\theta^2 + \frac{N}{T}) & \mu_{ew} / \gamma \\ \mu_{ew} / \gamma & \sigma_{ew}^2 \end{pmatrix}.$$

As a result, the constrained combination coefficients maximizing expected out-of-sample Sharpe ratio in (6) are obtained as in (15):

$$\kappa_{sr} = \arg \max_{\kappa: \kappa' \iota = 1} \widetilde{\mathcal{S}}_0(\widehat{w}(\kappa)) = \frac{\mathbf{S}^{-1} \eta}{\iota' \mathbf{S}^{-1} \eta} = \left(\frac{\psi_{ew}^2}{\psi_{ew}^2 + \frac{1}{\gamma} \frac{\mu_{ew}}{\sigma_{ew}^2} \frac{N}{T}}, \frac{\frac{1}{\gamma} \frac{\mu_{ew}}{\sigma_{ew}^2} \frac{N}{T}}{\psi_{ew}^2 + \frac{1}{\gamma} \frac{\mu_{ew}}{\sigma_{ew}^2} \frac{N}{T}} \right)$$

assuming

$$\iota' \mathbf{S}^{-1} \eta = \frac{\gamma \psi_{ew}^2 + \frac{\mu_{ew}}{\sigma_{ew}^2} \frac{N}{T}}{\psi_{ew}^2 + \frac{N}{T}} > 0,$$

which is fulfilled under the sufficient condition $\mu_{ew} > 0$. The solution κ_{sr} is unique due to the constraint $\kappa' \iota = 1$. In comparison, maximizing expected out-of-sample utility in (11) does not yield the same combination coefficients, and thus does not deliver maximal expected out-of-sample Sharpe ratio. Specifically, these combination coefficients are

$$\kappa_\gamma = \arg \max_{\kappa: \kappa' \iota = 1} \mathcal{U}(\widehat{w}(\kappa)) = \frac{\mathbf{S}^{-1} \iota}{\iota' \mathbf{S}^{-1} \iota} + \frac{1}{\gamma} \Sigma^{-1} \left(\eta - \frac{\iota' \mathbf{S}^{-1} \eta}{\iota' \mathbf{S}^{-1} \iota} \iota \right) = \left(\frac{\delta_{ew}}{\delta_{ew} + \frac{N/T}{2\gamma}}, \frac{\frac{N/T}{2\gamma}}{\delta_{ew} + \frac{N/T}{2\gamma}} \right),$$

and they are not equal to κ_{sr} except if the risk-aversion coefficient is $\gamma = \theta^2 / \mu_{ew}$.

EC.3.2. Unconstrained combination coefficients

We now consider the case where the combination coefficients, κ_1 and κ_2 , are unconstrained, in which case we recover the result in Proposition 1 that maximizing expected out-of-sample utility optimizes expected out-of-sample Sharpe ratio as well. Specifically, the unconstrained combination coefficients maximizing expected out-of-sample utility in (11) are²⁶

$$\kappa_\gamma = \frac{1}{\gamma} \mathbf{S}^{-1} \eta = \left(\frac{\psi_{ew}^2}{\psi_{ew}^2 + N/T}, \frac{1}{\gamma} \frac{\mu_{ew}}{\sigma_{ew}^2} \frac{N/T}{\psi_{ew}^2 + N/T} \right).$$

Moreover, for any risk-aversion coefficient $\gamma > 0$, these two combination coefficients deliver the maximum expected-out-of-sample Sharpe ratio:

$$\max_{\kappa} \tilde{\mathcal{S}}_0(\hat{w}(\kappa)) = \tilde{\mathcal{S}}_0(\hat{w}(\kappa_\gamma)) = \sqrt{\eta' \mathbf{S}^{-1} \eta} = \sqrt{\theta_{ew}^2 + (\theta^2 - \theta_{ew}^2) \frac{\psi_{ew}^2}{\psi_{ew}^2 + N/T}}.$$

That is, as established in Proposition 1, the optimal portfolio combining the risk-free asset, the tangent portfolio, and the $1/N$ portfolio optimizes both expected out-of-sample utility and Sharpe ratio when combination coefficients are unconstrained.

EC.4. Approximation to variance of out-of-sample Sharpe ratio

In Section 6 of the paper, we introduce the following simple approximation to the variance of out-of-sample Sharpe ratio:

$$\tilde{\mathcal{V}}_0(\hat{w}) = \frac{\mathbb{V}[\hat{w}' \mu]}{\mathbb{E}[\hat{w}' \Sigma \hat{w}]} \quad (\text{EC.15})$$

We now propose a refined approximation of which we derive closed-form expressions for fully invested sample mean-variance portfolios. Then, we assess the accuracy of the two approximations.

The true variance of out-of-sample Sharpe ratio is

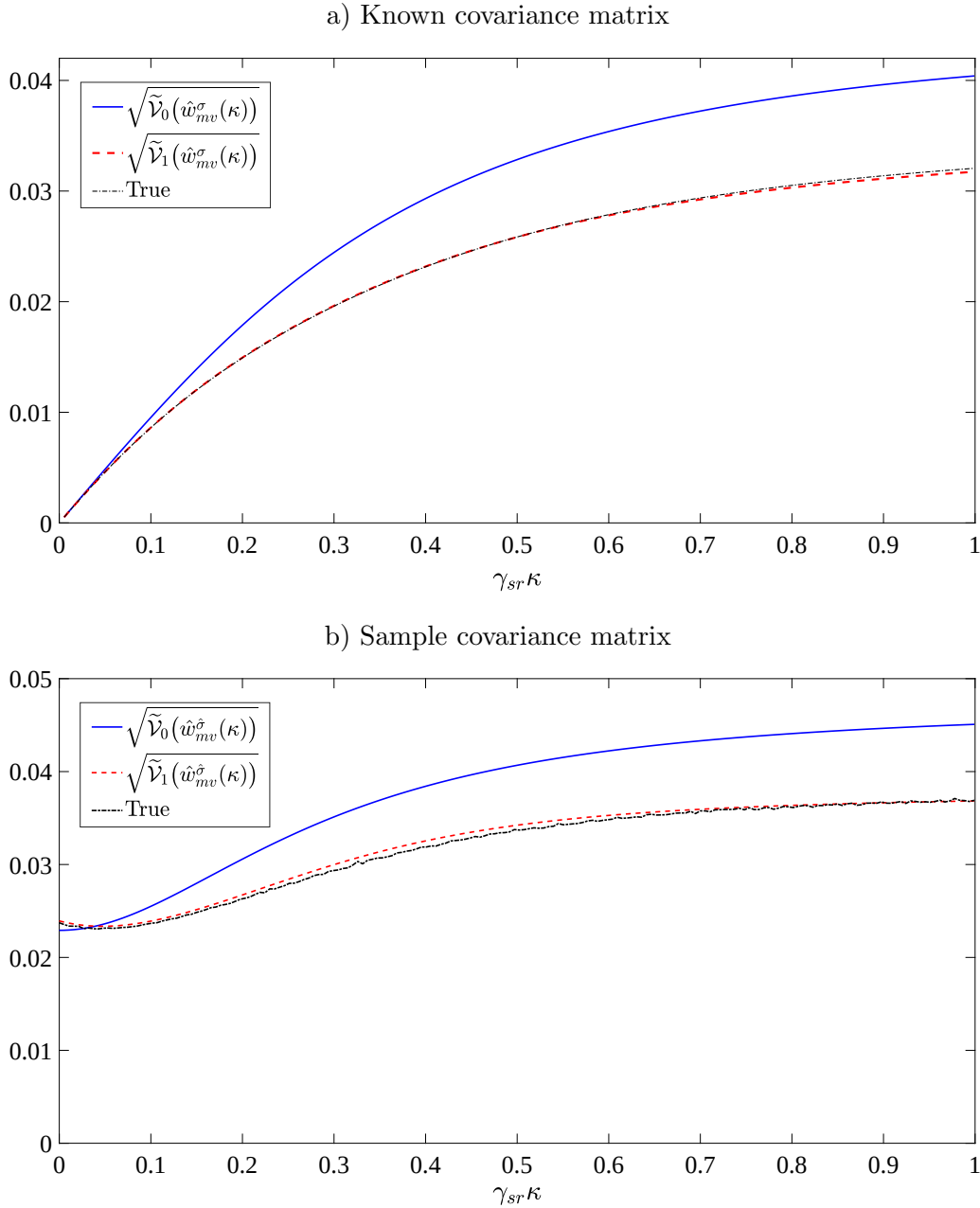
$$\mathcal{V}(\hat{w}) = \mathbb{E} \left[\frac{(\hat{w}' \mu)^2}{\hat{w}' \Sigma \hat{w}} \right] - \mathcal{S}(\hat{w})^2.$$

We obtain the refined approximation, denoted by $\tilde{\mathcal{V}}_1(\hat{w})$, by replacing $\mathcal{S}(\hat{w})$ by its first-order Taylor-series expansion, $\tilde{\mathcal{S}}_1$ in (7), and applying a first-order Taylor-series expansion to $\mathbb{E}[(\hat{w}' \mu)^2 / \hat{w}' \Sigma \hat{w}]$. Applying Footnote 7, this yields

$$\tilde{\mathcal{V}}_1(\hat{w}) = \tilde{\mathcal{V}}_0(\hat{w}) + \frac{(\mathbb{E}[\hat{w}' \mu])^2 \mathbb{V}[\hat{w}' \Sigma \hat{w}]}{(\mathbb{E}[\hat{w}' \Sigma \hat{w}])^3} - \frac{2\mathbb{E}[\hat{w}' \mu] \text{Cov}[\hat{w}' \mu, \hat{w}' \Sigma \hat{w}]}{(\mathbb{E}[\hat{w}' \Sigma \hat{w}])^2} + (\tilde{\mathcal{S}}_0(\hat{w})^2 - \tilde{\mathcal{S}}_1(\hat{w})^2). \quad (\text{EC.16})$$

Just like for the approximation $\tilde{\mathcal{S}}_1$ to the expected out-of-sample Sharpe ratio, the approximation $\tilde{\mathcal{V}}_1$ to the variance of out-of-sample Sharpe ratio depends on five quantities: $\mathbb{E}[\hat{w}' \mu]$, $\mathbb{E}[\hat{w}' \Sigma \hat{w}]$,

²⁶ When the sample mean-variance portfolio is estimated with the sample covariance matrix, i.e., $\hat{w}_1 = \frac{1}{\gamma} \hat{\Sigma} \hat{\mu}$, the two optimal combination coefficients maximizing expected out-of-sample utility are provided by Kan and Wang (2021, Section 4.2).

Figure EC.3 Approximation to standard deviation of out-of-sample Sharpe ratio

Note. The figure is constructed with population values of the mean and covariance matrix of asset returns calibrated from a dataset of monthly returns on 25 portfolios of stocks sorted on size and book-to-market spanning January 1927 to June 2021. The figure depicts the standard deviation of out-of-sample Sharpe ratio, according to the two approximations $\tilde{\mathcal{V}}_0$ in (32) and $\tilde{\mathcal{V}}_1$ in (32), of fully invested sample mean-variance portfolios, which are $\hat{w}_{mv}^\sigma(\kappa)$ in (20) when the covariance matrix is assumed known (Panel a) and $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ in (26) when using the sample covariance matrix (Panel b). The true standard deviation of out-of-sample Sharpe ratio is also depicted, which we find analytically with the method presented in Section EC.1 for $\hat{w}_{mv}^\sigma(\kappa)$, and numerically via 250,000 Monte Carlo simulations for $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$. The portfolios are estimated with a sample size of $T = 120$ monthly observations. The horizontal axis, $\gamma_{sr}\kappa$, is the intensity with which the global-minimum-variance portfolio is shrunk toward the tangent portfolio obtained for a coefficient $\kappa = 1/\gamma_{sr}$ defined in (17).

$\mathbb{V}[\hat{w}'\mu]$, $\mathbb{V}[\hat{w}'\Sigma\hat{w}]$, and $\text{Cov}[\hat{w}'\mu, \hat{w}'\Sigma\hat{w}]$. Therefore, we can derive closed-form expressions of $\tilde{\mathcal{V}}_1$ in the case of fully invested sample mean-variance portfolios, i.e., $\hat{w}_{mv}^\sigma(\kappa)$ in (20) when the covariance matrix Σ is assumed known, and $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ in (26) when using the sample covariance matrix in (1). Indeed, all five needed quantities are derived in Proposition EC.1 for $\hat{w}_{mv}^\sigma(\kappa)$ and Proposition EC.2 for $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$. In Figure EC.3, we use these closed-form expressions to compare the two approximations $\sqrt{\tilde{\mathcal{V}}_0}$ and $\sqrt{\tilde{\mathcal{V}}_1}$ to the standard deviation of out-of-sample Sharpe ratio as a function of the shrinkage intensity, $\gamma_{sr}\kappa$. We compare these approximations to the true standard deviation of out-of-sample Sharpe ratio, which we find analytically with the method presented in Section EC.1 for the normally distributed portfolio $\hat{w}_{mv}^\sigma(\kappa)$, and via 250,000 Monte Carlo simulations for $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$. The figure shows that $\tilde{\mathcal{V}}_0$ overestimates the true standard deviation of out-of-sample Sharpe ratio, which is because it ignores the correlation between out-of-sample mean return and variance that reduces standard deviation due to diversification. Moreover, it is not a too accurate approximation when the shrinkage intensity, $\gamma_{sr}\kappa$, is not small. In contrast, the refined approximation we propose, $\tilde{\mathcal{V}}_1$, provides a very accurate approximation for all shrinkage intensities, but consequently it is also a more complex function of the combination coefficient κ .

EC.5. Portfolio strategy of Kircher and Rosch (2021)

In this section, we explain the methodology proposed by Kircher and Rosch (2021) to maximize expected out-of-sample Sharpe ratio via bootstrap. In our notation and in the case where the sample covariance matrix in (1) is used, this strategy consists of bootstrapping with replacement $B = 1000$ times T returns from the sample of observations and finding the optimal combination coefficient κ that maximizes expected out-of-sample Sharpe ratio as²⁷

$$\kappa_{boot}^{\hat{\sigma}} = \arg \max_{\kappa} \frac{1}{B} \sum_{b=1}^B \frac{\hat{w}_b(\kappa)' \hat{\mu}_a}{\sqrt{\hat{w}_b(\kappa)' \hat{\Sigma} \hat{w}_b(\kappa)}}. \quad (\text{EC.17})$$

In Equation (EC.17), $\hat{w}_b(\kappa)$ is the sample mean-variance portfolio in (26) for the b th bootstrap sample and the adjusted mean-return estimator, $\hat{\mu}_a$, is defined as

$$\hat{\mu}_a = \hat{\mu}_{sh} + \max \left(0, \frac{\hat{\gamma}_a - \iota' \hat{\Sigma}^{-1} \hat{\mu}_{sh}}{\iota' \hat{\Sigma}^{-1} \iota} \right) \iota$$

where

$$\hat{\gamma}_a = \max \left(\gamma_{\min}, \frac{T - N - 2}{T} \iota' \hat{\Sigma}^{-1} \hat{\mu} \right), \quad \hat{\mu}_{sh} = (1 - \hat{\alpha}^*) \hat{\mu} + \hat{\alpha}^* \bar{\mu}, \quad \bar{\mu} = \iota' \hat{\mu} / N, \quad \hat{\alpha}^* = 1 - \sqrt{\tilde{X} / X},$$

²⁷ Our optimal combination coefficients κ derived in Sections 5 and 6 always belong to the interval $[0, 1/\hat{\gamma}_{sr}]$, i.e., the optimal portfolio is always between the GMV and tangent portfolios. Therefore, to ensure that the bootstrap approach is compared with our portfolios on a fair basis, we search κ in this interval when solving (EC.17). We find that this improves performance compared to allowing κ to be larger than $1/\hat{\gamma}_{sr}$.

with X and $\tilde{X}$ being

$$X = \frac{T(\hat{\mu} - \bar{\mu})'(\hat{\mu} - \bar{\mu})}{\text{Trace}(\hat{\Sigma})/N - \iota'\hat{\Sigma}\iota/N^2}, \quad \tilde{X} = X - N + \frac{2X^{N/2}e^{-X/2}}{\int_0^X t^{N/2-1}e^{-t/2}dt}.$$

Kircher and Rosch (2021) use $\gamma_{\min} = 3$ in their results. However, we use $\gamma_{\min} = 5$ to be consistent with the value used in Equation (44) and also because we find that it delivers better performance. Finally, when relying on the nonlinear shrinkage estimator of the covariance matrix of Ledoit and Wolf (2020), $\hat{\Sigma}_{nl}$, instead of the sample estimator, $\hat{\Sigma}$, we obtain the combination coefficient κ_{boot}^σ as above after replacing $\hat{\Sigma}$ by $\hat{\Sigma}_{nl}$ in the formulas, except for $\hat{\gamma}_a$ that remains unchanged.

EC.6. Out-of-sample performance with improved estimators of the expectation and variance of out-of-sample Sharpe ratio

In Tables 4 and 5 of the paper, we report the out-of-sample performance for real data of our two proposed portfolio strategies using the simple approximations $\tilde{\mathcal{S}}_0$ in (6) and $\tilde{\mathcal{V}}_0$ in (32) to the expectation and variance of out-of-sample Sharpe ratio, respectively. In this section, we replicate the out-of-sample performance of the two portfolio strategies using the improved estimators, $\tilde{\mathcal{S}}_1$ in (7) and $\tilde{\mathcal{V}}_1$ in (EC.16), for which we derive closed-form expressions in Sections EC.2 and EC.4 of this e-companion. We restrict to the mean-variance portfolio combination $\hat{w}_{mv}^\sigma(\kappa)$ in (20) that assumes the covariance matrix Σ is known, which is then replaced by the nonlinear shrinkage estimator of Ledoit and Wolf (2020), for the sake of brevity and because it yields a better performance than using the sample covariance matrix as we observed in Tables 4 and 5. Specifically, we rely on the combination coefficient solving

$$\max_{\kappa} \tilde{\mathcal{S}}_1(\hat{w}_{mv}^\sigma(\kappa)) - \lambda \sqrt{\tilde{\mathcal{V}}_1(\hat{w}_{mv}^\sigma(\kappa))}. \quad (\text{EC.18})$$

As in the paper, we consider $\lambda = 0$, which is the ‘Expected Sharpe’ strategy, and $\lambda = 1$, which is the ‘Robust Sharpe’ strategy.

The results are reported in Table EC.1. Regarding the ‘Expected Sharpe’ strategy, we observe that the two approximations deliver essentially the same performance, which is consistent with the analysis in Section EC.2 where we find that the two approximations are both very accurate and deliver similar optimal combination coefficients. Regarding the ‘Robust Sharpe’ strategy, the results are again quite similar for the two approximations. We also observe that using the approximation $\tilde{\mathcal{V}}_1$ for the variance of out-of-sample Sharpe ratio instead of $\tilde{\mathcal{V}}_0$ amounts to shrinking less intensively toward the GMV portfolio, which is because $\tilde{\mathcal{V}}_0$ overestimates the true variance as we observed in Figure EC.3.

Table EC.1 Results for real data with Ledoit-Wolf covariance matrix and improved estimators of the expectation and variance of out-of-sample Sharpe ratio

	$T = 120$				$T = 240$			
10MOM	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$
Expected Sharpe ($\tilde{S}_0$)	1.069	0.991	0.652	0.473	1.194	1.137	0.491	0.590
Expected Sharpe ($\tilde{S}_1$)	1.066	0.986	0.679	0.487	1.190	1.132	0.509	0.606
Robust Sharpe ($\tilde{S}_0, \tilde{V}_0$)	1.068	1.000	0.517	0.324	1.204	1.153	0.399	0.446
Robust Sharpe ($\tilde{S}_1, \tilde{V}_1$)	1.066	0.989	0.642	0.418	1.190	1.133	0.493	0.555
25SBTM	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$
Expected Sharpe ($\tilde{S}_0$)	1.141	0.998	0.909	0.352	1.333	1.228	0.621	0.461
Expected Sharpe ($\tilde{S}_1$)	1.139	0.995	0.918	0.355	1.331	1.226	0.626	0.464
Robust Sharpe ($\tilde{S}_0, \tilde{V}_0$)	1.138	1.008	0.782	0.261	1.331	1.235	0.542	0.364
Robust Sharpe ($\tilde{S}_1, \tilde{V}_1$)	1.138	1.001	0.855	0.302	1.329	1.228	0.596	0.420
25OPINV	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$
Expected Sharpe ($\tilde{S}_0$)	1.337	1.173	0.930	0.348	1.169	1.062	0.651	0.412
Expected Sharpe ($\tilde{S}_1$)	1.337	1.172	0.937	0.351	1.168	1.061	0.656	0.415
Robust Sharpe ($\tilde{S}_0, \tilde{V}_0$)	1.345	1.197	0.801	0.259	1.193	1.094	0.577	0.319
Robust Sharpe ($\tilde{S}_1, \tilde{V}_1$)	1.349	1.191	0.869	0.299	1.173	1.069	0.628	0.367
30IND	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$	SR	Net SR	Turnover	$\hat{\gamma}_{sr,\kappa}$
Expected Sharpe ($\tilde{S}_0$)	0.868	0.794	0.384	0.187	0.925	0.882	0.209	0.147
Expected Sharpe ($\tilde{S}_1$)	0.868	0.794	0.386	0.188	0.925	0.882	0.209	0.148
Robust Sharpe ($\tilde{S}_0, \tilde{V}_0$)	0.889	0.822	0.339	0.119	0.936	0.895	0.195	0.085
Robust Sharpe ($\tilde{S}_1, \tilde{V}_1$)	0.886	0.816	0.356	0.134	0.933	0.891	0.199	0.095

The table reports the annualized out-of-sample Sharpe ratio for the analysis in Section EC.6 of this e-companion. Four monthly return datasets are used and the covariance matrix is estimated via the nonlinear shrinkage estimator of Ledoit and Wolf (2020). The portfolio strategy ‘Expected Sharpe’ combines the GMV and tangent portfolios to maximize expected out-of-sample Sharpe ratio, which is estimated either via $\tilde{S}_0$ in (6) or the more accurate $\tilde{S}_1$ in (7). The portfolio strategy ‘Expected Sharpe’ combines the GMV and tangent portfolios to maximize the tradeoff between expectation and standard deviation of out-of-sample Sharpe ratio with a coefficient $\lambda = 1$, using the two approximations above for the expectation and $\tilde{V}_0$ in (32) or the more accurate $\tilde{V}_1$ in (EC.16) for the variance. The sample size T used to estimate the portfolios is 120 or 240 months. The net Sharpe ratio (net SR) is adjusted using proportional transaction costs of 20 basis points. The turnover is computed on a monthly basis. The quantity $\hat{\gamma}_{sr,\kappa}$ is the average weight allocated to the tangent portfolio across all estimation windows.

EC.7. Proofs of all results

In this section, we provide proofs for all the results contained in the paper and this e-companion.

Proof of Proposition 1

The proof of this proposition is contained in Section 4.1.

Proof of Proposition 2

The proof of this proposition is contained in Section 5.1.

Proof of Proposition 3

Part 1. The expected out-of-sample Sharpe ratio is obtained by plugging Equations (EC.6) and (EC.7) into the approximation $\tilde{S}_0$ in (6).

Part 2. Taking the derivative of (23) with respect to κ and setting it to zero yields the unique optimum $\kappa = \kappa_{sr}^\sigma$ in (24). It is easy to see that $\kappa = \kappa_{sr}^\sigma \rightarrow 1/\gamma_{sr}$ as $T \rightarrow \infty$. Moreover, as $T \rightarrow \infty$, the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ is no longer random and equals $w_g + \kappa \mathbf{B}\mu$ because $\hat{\mu}$ is a consistent estimator. Therefore, as $T \rightarrow \infty$, the portfolio $\hat{w}_{mv}^\sigma(\kappa_{sr}^\sigma) \rightarrow w_g + \frac{1}{\gamma_{sr}} \mathbf{B}\mu = w_{sr}^*$, and thus it is a consistent estimator.

Part 3. The optimal value of the expected out-of-sample Sharpe ratio is obtained by plugging (24) into (23) and following some algebraic manipulations.

Proof of Proposition 4

Part 1. The expected out-of-sample Sharpe ratio is obtained by plugging Equations (EC.11) and (EC.12) into the approximation $\tilde{S}_0$ in (6).

Part 2. Taking the derivative of (28) with respect to κ and setting it to zero yields the unique optimum $\kappa = \kappa_{sr}^{\hat{\sigma}}$ in (29). It is strictly smaller than the optimum in (24) because $\frac{(T-N)(T-N-3)}{T(T-N-1)} < 1$ for $T \geq N + 4$. It is then easy to see that $\kappa = \kappa_{sr}^{\hat{\sigma}} \rightarrow 1/\gamma_{sr}$ as $T \rightarrow \infty$. Moreover, as $T \rightarrow \infty$, the sample mean-variance portfolio $\hat{w}_{mv}^{\hat{\sigma}}(\kappa)$ is no longer random and equals $w_g + \kappa \mathbf{B}\mu$ because $\hat{\mu}$ and $\hat{\Sigma}$ are consistent estimators. Therefore, as $T \rightarrow \infty$, the portfolio $\hat{w}_{mv}^{\hat{\sigma}}(\kappa_{sr}^{\hat{\sigma}}) \rightarrow w_g + \frac{1}{\gamma_{sr}} \mathbf{B}\mu = w_{sr}^*$, and thus it is a consistent estimator.

Part 3. The optimal value of the expected out-of-sample Sharpe ratio is obtained by plugging (29) into (28) and following some algebraic manipulations. To see that it is smaller than the optimal value in (25), notice that the coefficient in front of θ_g^2 is $\frac{T-N-1}{T-2} < 1$, and the coefficient in front of $(\theta^2 - \theta_g^2) \frac{\psi^2}{\psi^2 + (N-1)/T}$ is $\frac{(T-N)(T-N-3)}{(T-2)(T-N-1)} < 1$, which proves the inequality.

Proof of Proposition 5

As shown in (33), the optimization problem we need to solve is

$$\kappa_\lambda = \arg \max_{\kappa: \kappa' \iota = 1} \frac{\kappa' \eta - \lambda \sqrt{\kappa' \mathbf{V} \kappa}}{\sqrt{\kappa' \mathbf{S} \kappa}}. \quad (\text{EC.19})$$

Given that the objective function is homogeneous of degree zero, we can first solve it without the constraint $\kappa' \iota = 1$ and then normalize the solution. From the first-order conditions, any optimal unconstrained κ is of the form

$$\kappa_{unc} = \left(\frac{1}{s^2} (n - \lambda v) \mathbf{S} + \frac{\lambda}{v} \mathbf{V} \right)^{-1} \eta, \quad (\text{EC.20})$$

which after normalization yields the endogenous solution

$$\kappa_\lambda = \frac{\kappa_{unc}}{\iota' \kappa_{unc}} = \frac{((1 - \delta) \mathbf{S} + \delta \mathbf{V})^{-1} \eta}{\iota' ((1 - \delta) \mathbf{S} + \delta \mathbf{V})^{-1} \eta}, \quad (\text{EC.21})$$

where δ is given by (35).

Proof of Proposition 6

Part 1. Sample mean returns are distributed as $\hat{\mu} \sim \mathcal{N}(\mu, \mathbf{\Sigma}/T)$. Using the property $\mathbf{B} \mathbf{\Sigma} \mathbf{B} = \mathbf{B}$ (Simaan and Simaan 2019, Lemma 2.1), this means that the covariance matrix of the sample mean-variance portfolio $\hat{w}_{mv}^\sigma(\kappa)$ is

$$\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)] = \frac{\kappa^2}{T} \mathbf{B}.$$

Therefore, the numerator of $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ in (36) is $\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)' \mu] = \mu' \mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)] \mu = \kappa^2 \psi^2 / T$, because $\mu' \mathbf{B} \mu = \psi^2$. Moreover, the denominator of $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ is given by (EC.7).

Part 2. The numerator of $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ in (37) is $\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)' \mu]$. As shown by Lassance, Martin-Utrera, and Simaan (2021b, Proposition 3), this variance is equal to

$$\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)' \mu] = \frac{\sigma_g^2 \psi^2}{T - N - 1} + \kappa^2 \psi^2 \frac{2T(N + 1) + T^2(T - N - 3 + 2\psi^2(T - N))}{(T - N)(T - N - 1)^2(T - N - 3)}. \quad (\text{EC.22})$$

Moreover, the denominator of $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ is given by (EC.12).

Proof of Proposition 7

Part 1. Using Equations (23) and (36), we find that maximizing the objective function in (39) amounts to solving

$$\arg \max_{\kappa} \frac{\mu_g + \kappa \psi^2 - \lambda |\kappa| \psi / \sqrt{T}}{\sqrt{\sigma_g^2 + \kappa^2 (\psi^2 + \frac{N-1}{T})}}. \quad (\text{EC.23})$$

First, observe that the optimal κ is necessarily positive. This is because the variance of out-of-sample Sharpe ratio in (36) is an even function of κ but the expected out-of-sample Sharpe ratio

in (23) is larger for $\kappa = 0$ than for any $\kappa < 0$. Therefore, we can get rid of the absolute value in (EC.23). Then, taking the derivative of (EC.23) with respect to κ and setting it to zero yields the solution

$$\kappa = \kappa_{sr}^\sigma \left(1 - \frac{\lambda}{\psi\sqrt{T}} \right), \quad (\text{EC.24})$$

where κ_{sr}^σ is given by (24). If the coefficient $\lambda \leq \psi\sqrt{T}$, the optimal κ in (EC.24) is positive and thus corresponds to the optimum. Otherwise, the optimal κ is zero as discussed above.

Part 2. To prove the result, we first show that the approximation to the variance of out-of-sample Sharpe ratio, $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ in (37), is minimized for $\kappa = 0$. This is the case because, after taking the derivative of $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ with respect to κ and setting it equal to zero, we obtain the following optimality condition:

$$\kappa \frac{T-2}{T-N-1} f(T, N, \psi^2) = \kappa \frac{T^2(T-2)}{(T-N)(T-N-1)^2(T-N-3)} \left(\psi^2 + \frac{N-1}{T} \right). \quad (\text{EC.25})$$

This condition holds if and only if $\kappa = 0$, and thus $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ is minimized by $\kappa = 0$. This result means that, when $\lambda \rightarrow \infty$, the combination coefficient κ_λ^σ in (40) amounts to minimizing $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ and thus equals zero. Moreover, when $\lambda = 0$, $\kappa_\lambda^\sigma = \kappa_{sr}^\sigma$ in (29) by definition. Finally, when $0 < \lambda < \infty$, we necessarily have $0 \leq \kappa_\lambda^\sigma \leq \kappa_{sr}^\sigma$. This is because $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ is increasing in κ for $\kappa \geq 0$, and $\tilde{\mathcal{S}}_0(\hat{w}_{mv}^\sigma(\kappa))$ is increasing in κ for $\kappa \in [0, \kappa_{sr}^\sigma]$ and decreasing in κ for $\kappa \geq \kappa_{sr}^\sigma$, and thus any $\kappa > \kappa_{sr}^\sigma$ necessarily delivers a smaller value of the objective function in (40) than choosing $\kappa \leq \kappa_{sr}^\sigma$. To see that $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ is increasing in κ for $\kappa \geq 0$, observe that the derivative of $\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))$ with respect to κ is positive for $\kappa \geq 0$ if the left-hand side of (EC.25) is greater than its right-hand side, which is true for any $\psi^2 \geq 0$. Finally, the inequalities are strict, $0 < \kappa_\lambda^\sigma < \kappa_{sr}^\sigma$, if $\psi^2 > 0$. Indeed, the derivative of the objective function evaluated at $\kappa = 0$ is strictly positive if $\psi^2 > 0$:

$$\left. \frac{\partial}{\partial \kappa} \left[\tilde{\mathcal{S}}_0(\hat{w}_{mv}^\sigma(\kappa)) - \lambda \sqrt{\tilde{\mathcal{V}}_0(\hat{w}_{mv}^\sigma(\kappa))} \right] \right|_{\kappa=0} = \frac{T}{\sqrt{(T-2)(T-N-1)}} \frac{\psi^2}{\sigma_g}.$$

Part 3. When the sample size $T \rightarrow \infty$, the approximations to the variance of out-of-sample Sharpe ratio in Proposition 6 go to zero, and thus the combination coefficients κ_λ^σ and $\kappa_\lambda^{\hat{\sigma}}$ maximize expected out-of-sample Sharpe ratio alone. The result follows because the combination coefficients maximizing expected out-of-sample Sharpe ratio yield consistent estimators of the tangent portfolio, w_{sr}^* , as shown in Propositions 3 and 4.

Proof of Proposition EC.1

To characterize $\tilde{\mathcal{S}}_1(\hat{w}_{mv}^\sigma(\kappa))$ in (EC.5), we need closed-form expressions for $\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \mu]$, $\mathbb{E}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)]$, $\mathbb{V}[\hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)]$, and $\text{Cov}[\hat{w}_{mv}^\sigma(\kappa)' \mu, \hat{w}_{mv}^\sigma(\kappa)' \Sigma \hat{w}_{mv}^\sigma(\kappa)]$. These quantities can be obtained from Lassance, Martin-Utrera, and Simaan (2021b, Proposition IA.1) and correspond to (EC.6)–(EC.9).

Proof of Proposition EC.2

To characterize $\tilde{S}_1(\hat{w}_{mv}^{\hat{\sigma}}(\kappa))$ in (EC.10), we need closed-form expressions for $\mathbb{E}[\hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \mu]$, $\mathbb{E}[\hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \Sigma \hat{w}_{mv}^{\hat{\sigma}}(\kappa)]$, $\mathbb{V}[\hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \Sigma \hat{w}_{mv}^{\hat{\sigma}}(\kappa)]$, and $\text{Cov}[\hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \mu, \hat{w}_{mv}^{\hat{\sigma}}(\kappa)' \Sigma \hat{w}_{mv}^{\hat{\sigma}}(\kappa)]$. The first two can be obtained from Kan, Wang, and Zhou (2021b, Lemma 1), and the last two from Lassance, Martin-Utrera, and Simaan (2021b, Proposition 3), and correspond to (EC.11)–(EC.14).

References in electronic companion

- Kan R, Wang X (2021) Optimal portfolio choice with benchmark. *Rotman School of Management Working Paper No. 3760640*.
- Kan R, Wang X, Zhou G (2021) Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science* (forthcoming).
- Kircher F, Rosch D (2021) A shrinkage approach for sharpe ratio optimal portfolios with estimation risks. *Journal of Banking and Finance* 133.
- Lassance N, Martin-Utrera A, Simaan M (2021) A robust approach to optimal portfolio choice with parameter uncertainty. *Available at SSRN 3855546*.
- Ledoit O, Wolf M (2020) Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics* 48(5):3043–3065.
- Mathai A, Provost S (1992) *Quadratic Forms in Random Variables* (Marcel Dekker, New York).
- Meng XL (2005) From unit root to stein’s estimator to fisher’s k statistics: If you have a moment, i can tell you more. *Statistical Science* 20(2):141–162.
- Simaan M, Simaan Y (2019) Rational explanation for rule-of-thumb practices in asset allocation. *Quantitative Finance* 19(12):2095–2109.
- Tu J, Zhou G (2011) Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99(1):204–215.