Optimal Portfolio Choice with Fat Tails and Parameter Uncertainty

Raymond Kan and Nathan Lassance*

November 28, 2024

Abstract

Existing portfolio combination rules that optimize the out-of-sample performance under parameter uncertainty assume multivariate normally distributed returns. However, we show that this assumption is not innocuous because fat tails in returns lead to poorer out-of-sample performance of the sample mean-variance and sample global minimum-variance portfolios relative to normality. Consequently, when returns are fat-tailed, portfolio combination rules should allocate less to the sample mean-variance and sample global minimum-variance portfolios, and more to the risk-free asset, than the normality assumption prescribes. Empirical evidence shows that accounting for fat tails in the construction of optimal portfolio combination rules significantly improves their out-of-sample performance.

Keywords: portfolio combination, elliptical distribution, estimation risk.

---

*Raymond Kan is affiliated at University of Toronto Rotman School of Management, e-mail: raymond.kan@rotman.utoronto.ca. Nathan Lassance is corresponding author and is affiliated at UCLouvain LFIN/LIDAM, e-mail: nathan.lassance@uclouvain.be. We are grateful to an anonymous referee, George G. Pennacchi (the editor), Xinghua Zheng, and Guofu Zhou for their valuable comments.

I. Introduction

Sample mean-variance portfolios are not optimal out of sample because of the impact of estimation errors in the mean and the covariance matrix; see, e.g. Michaud (1989), Chopra and Ziemba (1993), Kan and Zhou (2007), DeMiguel, Garlappi and Uppal (2009), and Barroso and Saxena (2021). To study the impact of estimation risk on out-of-sample performance and optimal portfolio choice, most papers assume that returns are multivariate normally distributed.[1] However, this assumption is not innocuous because it disregards the impact of fat tails on estimation risk.

In this paper, we study the impact of estimation errors on the out-of-sample performance of sample mean-variance portfolios when asset returns are multivariate elliptically distributed, and thus, are fat-tailed.[2] We first quantify the impact of fat tails on the out-of-sample mean, variance, and utility of sample mean-variance portfolios. Then, we extend the results in Kan and Zhou (2007), who show that with parameter uncertainty it is suboptimal to hold the sample mean-variance portfolio and the risk-free asset with the proportions advised by portfolio theory and that investors benefit from holding a three-fund rule that also includes the sample global minimum-variance (GMV) portfolio, to the case with fat tails in returns. Specifically, we derive the optimal two-fund and three-fund portfolio combination rules in Kan and Zhou (2007) when

[1]See Jobson and Korkie (1980), Jorion (1986), Okhrin and Schmid (2006), Kan and Zhou (2007), Kan and Smith (2008), DeMiguel et al. (2009), Frahm and Memmel (2010), Tu and Zhou (2011), DeMiguel, Martín-Utrera and Nogales (2013, 2015), Ao, Li and Zheng (2019), Kan, Wang and Zhou (2021), Yuan and Zhou (2023), Kan and Wang (2023), Lassance, Martín-Utrera and Simaan (2024a), and Kan, Wang and Zheng (2024).

[2]Not all elliptical distributions have a kurtosis larger than three (Bentler and Berkane, 1986). However, we consider the subclass of elliptical distributions in El Karoui (2010, 2013), which are all fat-tailed.

asset returns are multivariate elliptically distributed instead of normally distributed. We analyze these two problems in finite samples and in a high-dimensional asymptotic setting.

Our theoretical results show that fat tails have a substantial impact on the out-of-sample performance of sample mean-variance portfolios. Specifically, multivariate elliptical returns increase both the expected out-of-sample mean and variance of sample mean-variance portfolios compared with the case of multivariate normal returns. However, the increase in the out-of-sample variance is larger than the increase in the out-of-sample mean, which hurts the risk-adjusted performance. As a result, it is optimal in a two-fund rule that we allocate less weight to the sample mean-variance portfolio, and more weight to the risk-free asset, relative to the normality case. Similarly, in a three-fund rule, it is optimal to allocate less weights to both the sample mean-variance portfolio and the sample global minimum-variance (GMV) portfolio. These results mean that the normality assumption can deliver misleading forecasts of out-of-sample performance and suboptimal portfolio rules.

In our empirical analysis across six datasets of equity returns, the optimal weights on the sample mean-variance portfolio and the sample GMV portfolio calibrated to the multivariate elliptical distribution are on average 24% and 49% smaller, respectively, when the sample size is 120 months compared with the weights calibrated to the multivariate normal distribution in Kan and Zhou (2007). The resulting gain in performance can be large. On average across datasets and across the two-fund and three-fund rules, the annualized net-of-cost utility gain when switching from the normal to the elliptical distribution is 32.9, 12.6, and 9.3 percentage points when the sample size is 60, 120, and 240 months, respectively, and the risk-aversion coefficient is equal to one. These results are robust to using daily data, which, relative to monthly data, have fatter tails but allow a more accurate estimation of asset return covariances.[3] We also find that calibrating the two-fund and three-fund rules to the multivariate elliptical distribution helps reduce kurtosis significantly relative to the multivariate normal calibration.

Instead of obtaining the combination coefficients in the two-fund and three-fund rules based on the multivariate elliptical distribution, one could use a non-parametric approach to estimate them entirely based on the data. To get a taste of how such an approach might fare, we consider estimating the combination coefficients using a five-fold cross-validation method. We find that although cross-validation often outperforms the multivariate normality assumption, it generally delivers a lower out-of-sample utility compared with the multivariate elliptical distribution. This suggests there are benefits of making good distributional assumptions as opposed to a pure non-parametric approach.

As a robustness test of our empirical experiments, we consider the optimal combination of the sample GMV portfolio with the risk-free asset, which can be an appealing rule because it ignores expected returns and delivers less extreme portfolio weights and transaction costs. Although estimation risk is less of a concern when we sideline expected returns, we find that calibrating this portfolio combination to elliptically instead of normally distributed returns continues to deliver superior performance, especially when the sample size is small.

Few papers study the impact of estimation risk on sample mean-variance portfolios under the multivariate elliptical distribution. Mainly, we can cite El Karoui (2010, 2013) who studies the high-dimensional asymptotic out-of-sample mean and variance of sample mean-variance

[3]Fat tails are pronounced for both monthly and daily data in the datasets we consider. Specifically, when we assume returns are multivariate t-distributed, the estimated number of degrees of freedom varies, across datasets, between roughly 4 and 8 for monthly returns and 3 and 6 for daily returns.

portfolios. Lassance, Vanderveken and Vrins (2024b) exploit these asymptotic results to combine the sample mean-variance and 1/N portfolios. In this paper, we provide asymptotic results that complement El Karoui (2010, 2013), but more importantly, we also provide exact finite-sample results. Moreover, we derive optimal two-fund and three-fund rules under elliptical returns in both the finite-sample and asymptotic cases. We find that the asymptotic setting under-estimates the impact of fat tails and over-estimates the optimal weights on the sample mean-variance portfolio and the sample GMV portfolio in the two-fund and three-fund rules. However, the asymptotic setting delivers faster analytical solutions and thus can be a viable approach in practice if the sample size is large enough.

In addition to proposing new optimized portfolio rules that account for fat tails, we also contribute to explaining why existing optimized portfolio rules often disappoint out of sample when they are explicitly designed to address estimation error. Specifically, Barroso and Saxena (2021) find that the three-fund rules in Kan and Zhou (2007) and Tu and Zhou (2011) often fail to outperform simple naive portfolios and substantially under-estimate the out-of-sample portfolio risk. Our results show that at least part of this disappointing performance can be attributed to the ignorance of fat tails. Indeed, we show theoretically that fat tails can substantially increase the out-of-sample variance of estimated mean-variance portfolios, and empirically that the two-fund and three-fund rules calibrated to the multivariate elliptical distribution provide better forecasts of out-of-sample return statistics. Moreover, whereas in our empirical analysis the two-fund and three-fund rules of Kan and Zhou (2007) under normality often do not perform well and deliver negative utilities, our proposed portfolio rules corrected for fat tails offer a consistently better performance.

Finally, our finding that fat tails can have a substantial impact on the out-of-sample performance of mean-variance portfolios contrasts with Tu and Zhou (2004), who find that the performance losses associated with ignoring fat tails are small. Their different conclusion can be explained because estimation risk is more limited in their setting, which also limits how much fat tails can impact parameter uncertainty and portfolio performance. First, they conduct an in-sample analysis with a long time window spanning 1963 to 1997, instead of an out-of-sample analysis.[4] Second, the base case results in Tu and Zhou (2004, Table 7) are based on 12 assets, which is rather small in comparison with their sample size. The robustness test with 23 assets in their Table 8 suggests larger performance losses due to fat tails. Third, they optimize their mean-variance portfolio under weight constraints, which make portfolios less sensitive to estimation risk (Jagannathan and Ma, 2003). For these reasons, one ought to be cautious in generalizing their conclusion that fat tails have a small impact on portfolio performance to a more general setting. Indeed, we reach different conclusions in our setting in which the impact of estimation risk can be far more serious.

The rest of the paper is organized as follows. Section II sets out the assumptions, notation, and objectives. Sections III and IV present our theoretical results about the impact of fat tails on estimation risk. Section V explains how we estimate our optimal two-fund and three-fund rules. Section VI evaluates the out-of-sample performance using simulated and empirical data.

---

[4]Specifically, Tu and Zhou (2004) estimate the mean and covariance matrix of returns using a Bayesian framework in which the investor has a prior that allows the returns to be multivariate t-distributed with different degrees of freedom. Then, over the whole sample of data, they compute the in-sample utility delivered by a mean-variance portfolio constructed with these estimates of the mean and covariance matrix. Finally, they compare this utility to that of a mean-variance portfolio constructed with estimates of the mean and covariance matrix that force the investor to have a prior belief that returns are multivariate normal.

Section VII concludes. In Section IV of the supplementary material, we detail the proofs of all theoretical results.

II. Assumptions, Notations, and Objectives

Let r_t be the vector of excess returns of $N \geq 2$ assets at time t. We assume r_t has a mean of μ and a positive-definite covariance matrix of Σ. In addition, we assume r_t is multivariate elliptically distributed. Following El Karoui (2010, 2013), we assume r_t has the following stochastic representation:

$$r_t = \mu + (\tau_t \Sigma)^{1/2} Y_t, \tag{1}$$

where $Y_t \sim \mathcal{N}(0_N, I_N)$, τ_t is a univariate random variable that satisfies $\mathbb{E}[\tau_t] = 1$, and Y_t is independent of τ_t. Two important cases are the multivariate normal distribution, obtained for $\tau_t = 1$, and the multivariate t-distribution with $\nu > 2$ degrees of freedom, obtained for

$$\tau_t \sim (\nu - 2)/\chi^2_\nu, \tag{2}$$

where χ^2_ν stands for a chi-square distribution with ν degrees of freedom. When the fourth moment of returns exists, we denote

$$\kappa = \mathbb{E}[\tau_t^2] - 1 \tag{3}$$

as the excess kurtosis of the returns divided by three. For example, when r_t follows a multivariate t-distribution in (2) with $\nu > 4$, we have $\kappa = 2/(\nu - 4)$.

As El Karoui (2010, 2013) explain, the stochastic representation in (1) does not encompass all multivariate elliptical distributions, but we focus on this class because it possesses fat tails. The elliptical distribution assumption is less restrictive than normality but still simple enough to allow us to derive detailed asymptotic and finite-sample properties of the considered portfolios.

We define the mean return and squared Sharpe ratio of the GMV portfolio, the maximum squared Sharpe ratio, and the difference between the maximum squared Sharpe ratio and that of the GMV portfolio as

$$\mu_g = \frac{\mu^\top \Sigma^{-1} 1_N}{1_N^\top \Sigma^{-1} 1_N}, \tag{4}$$

$$\theta_g^2 = \frac{(\mu^\top \Sigma^{-1} 1_N)^2}{1_N^\top \Sigma^{-1} 1_N}, \tag{5}$$

$$\theta^2 = \mu^\top \Sigma^{-1} \mu, \tag{6}$$

$$\psi^2 = \theta^2 - \theta_g^2, \tag{7}$$

respectively. Given a risk-aversion coefficient γ, the mean-variance utility of a portfolio w is

$$U(w) = w^\top \mu - \frac{\gamma}{2} w^\top \Sigma w. \tag{8}$$

The weights of the portfolio that maximizes $U(w)$ are

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

The mean, variance, and utility of $w^\star$ are

$$\mu_p = \theta^2/\gamma, \tag{10}$$

$$\sigma_p^2 = \theta^2/\gamma^2, \tag{11}$$

$$U(w^\star) = \theta^2/(2\gamma). \tag{12}$$

Given an i.i.d. sample of size $T > N$, $(r_1, \ldots, r_T)$, the sample estimates of μ and Σ are

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{13}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} (r_t - \hat{\mu})(r_t - \hat{\mu})^\top. \tag{14}$$

The weights of the sample mean-variance portfolio that estimates $w^\star$ are then given by

$$\hat{w} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}. \tag{15}$$

We have two main objectives in this paper. Our first objective is to study the out-of-sample mean, variance, and utility of $\hat{w}$:

$$\tilde{\mu}_p = \hat{w}^\top \mu = \frac{1}{\gamma}\hat{\mu}^\top \hat{\Sigma}^{-1}\mu, \tag{16}$$

$$\tilde{\sigma}_p^2 = \hat{w}^\top \Sigma \hat{w} = \frac{1}{\gamma^2}\hat{\mu}^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}, \tag{17}$$

$$U(\hat{w}) = \tilde{\mu}_p - \frac{\gamma}{2}\tilde{\sigma}_p^2. \tag{18}$$

Our second objective is to determine the optimal two-fund and three-fund rules in Kan and Zhou

(2007) but under the multivariate elliptical distribution assumption. That is, we consider the portfolio combination rules

$$\hat{w}_{2f}(c) = \frac{c}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}, \tag{19}$$

$$\hat{w}_{3f}(c_1,c_2) = \frac{c_1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu} + \frac{c_2}{\gamma}\hat{\Sigma}^{-1}1_N, \tag{20}$$

and we determine the combination coefficients c, c_1, and c_2 so as to maximize the expected out-of-sample utility. In the next proposition, we show that the optimal combination coefficients depend on five different constants.

Proposition 1. *Let* $\tilde{\mu}_1 = \mathbb{E}[\hat{\mu}^\top\hat{\Sigma}^{-1}\mu]$*,* $\tilde{\mu}_2 = \mathbb{E}[\mu^\top\hat{\Sigma}^{-1}1_N]$*,* $\tilde{\sigma}_1^2 = \mathbb{E}[\hat{\mu}^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}]$*,* $\tilde{\sigma}_2^2 = \mathbb{E}[1_N^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N]$*, and* $\tilde{\sigma}_{12} = \mathbb{E}[\hat{\mu}^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N]$*. The optimal combination coefficient that maximizes the expected out-of-sample utility of the two-fund rule is*

$$c^\star = \underset{c}{\operatorname{argmax}}\,\mathbb{E}\big[U(\hat{w}_{2f}(c))\big] = \tilde{\mu}_1/\tilde{\sigma}_1^2, \tag{21}$$

and the optimal combination coefficients for the three-fund rule are

$$(c_1^\star, c_2^\star) = \underset{(c_1,c_2)}{\operatorname{argmax}}\,\mathbb{E}\big[U(\hat{w}_{3f}(c_1,c_2))\big] = \left(\frac{\tilde{\mu}_1\tilde{\sigma}_2^2 - \tilde{\mu}_2\tilde{\sigma}_{12}}{\tilde{\sigma}_1^2\tilde{\sigma}_2^2 - \tilde{\sigma}_{12}^2}, \frac{\tilde{\mu}_2\tilde{\sigma}_1^2 - \tilde{\mu}_1\tilde{\sigma}_{12}}{\tilde{\sigma}_1^2\tilde{\sigma}_2^2 - \tilde{\sigma}_{12}^2}\right). \tag{22}$$

Finally, note that we focus on mean-variance portfolios even though the returns are multivariate elliptically distributed, and thus, are fat tailed. This is because the multivariate elliptical distribution is closed under linear transformations, which means that only the portfolio mean return and variance can vary as we change the portfolio weights w, and all efficient portfolios maximizing expected utility are mean-variance portfolios; see Chamberlain (1983) and Schuhmacher, Kohrs and Auer (2021). Therefore, under parameter uncertainty, we only consider portfolio rules built from $\hat{\mu}$ and $\hat{\Sigma}$; namely, the sample mean-variance portfolio, the two-fund rule, and the three-fund rule. The result we uncover and detail in this paper is that even though the sample mean-variance portfolio only depends on $\hat{\mu}$ and $\hat{\Sigma}$, the way it should be optimally combined with the risk-free asset in the two-fund rule, and also with the sample GMV portfolio in the three-fund rule, also depends on the fat tails of returns via the distribution of τ_t. This is because fat tails impact the effect that estimation risk has on the out-of-sample performance of the sample mean-variance portfolio.

III. Asymptotic Results

In this section, we study the impact of fat tails on estimation risk in two different asymptotic regimes. First, the standard regime in which N is fixed and $T \to \infty$. Second, as in Ledoit and Wolf (2017), Ao et al. (2019), and Kan et al. (2024), among others, the high-dimensional regime in which $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$.

A. Fixed N Asymptotic Regime

In the next proposition, we derive the fixed N asymptotic distribution of the sample mean-variance portfolio $\hat{w}$ and its out-of-sample mean, variance, and utility under the assumption that returns follow a multivariate elliptical distribution.

Proposition 2. *Let N be fixed, $T \to \infty$, and the kurtosis parameter κ in* (3) *exists.*

1. *The asymptotic distribution of the sample mean-variance portfolio* $\hat{w}$ *is*

$$\sqrt{T}(\hat{w}-w^{\star}) \xrightarrow{d} N\left(0_N, \frac{1}{\gamma^2}\left[[1+(1+\kappa)\theta^2]\Sigma^{-1}+(1+2\kappa)\Sigma^{-1}\mu\mu^{\top}\Sigma^{-1}\right]\right). \tag{23}$$

2. *The joint asymptotic distribution of the out-of-sample mean and variance of* $\hat{w}$ *is*

$$\sqrt{T}\begin{bmatrix}\tilde{\mu}_p-\mu_p\\ \tilde{\sigma}_p^2-\sigma_p^2\end{bmatrix} \xrightarrow{d} N\left(0_2, \frac{\theta^2[1+(2+3\kappa)\theta^2]}{\gamma^2}\begin{bmatrix}1 & 2/\gamma\\ 2/\gamma & 4/\gamma^2\end{bmatrix}\right). \tag{24}$$

3. *The asymptotic distribution of the out-of-sample utility of* $\hat{w}$ *is*[5]

$$T[U(\hat{w})-U(w^{\star})] \xrightarrow{d} -\frac{1}{2\gamma}\left([1+(1+\kappa)\theta^2]\chi^2_{N-1}+[1+(2+3\kappa)\theta^2]\chi^2_1\right), \tag{25}$$

where the two chi-squared random variables are independent of each other.

4. *The first-order biases of* $\tilde{\mu}_p$, $\tilde{\sigma}_p^2$, *and* $U(\hat{w})$ *are given by*

$$\mathbb{E}[\tilde{\mu}_p]-\mu_p = \frac{(N+2)(1+\kappa)\theta^2}{\gamma T}+O(T^{-2}), \tag{26}$$

$$\mathbb{E}[\tilde{\sigma}_p^2]-\sigma_p^2 = \frac{N+[3(N+2)(1+\kappa)-1]\theta^2}{\gamma^2 T}+O(T^{-2}), \tag{27}$$

$$\mathbb{E}[U(\hat{w})]-U(w^{\star}) = -\frac{N+[(N+2)(1+\kappa)-1]\theta^2}{2\gamma T}+O(T^{-2}). \tag{28}$$

Proposition 2 shows that $\hat{w}$ is asymptotically unbiased when N is fixed. Therefore, for the optimal combination coefficients, we have $\lim_{T\to\infty} c^{\star}=1$ and $\lim_{T\to\infty}(c_1^{\star}, c_2^{\star})=(1,0)$.

[5] $U(\hat{w})$ is T-consistent instead of $\sqrt{T}$-consistent because it is bounded above by $U(w^{\star})$.

Proposition 2 also shows that while $\tilde{\mu}_p$, $\tilde{\sigma}_p^2$, and $U(\hat{w})$ are consistent estimators of μ_p, σ_p^2, and $U(w^\star)$, respectively, the kurtosis parameter κ increases the magnitude of both the first-order bias and the asymptotic variance of $\tilde{\mu}_p$, $\tilde{\sigma}_p^2$, and $U(\hat{w})$. For example, take $N = 5$ and $T = 120$, which is a case where applying the fixed N asymptotic is reasonable. Then, when $\theta = 0.3$ and $\gamma = 1$, the first-order bias of $\tilde{\mu}_p$, $\tilde{\sigma}_p^2$, and $U(\hat{w})$ is 0.0053, 0.057, and -0.023 when $\kappa = 0$, respectively, versus 0.016, 0.088, and -0.028 when $\kappa = 2$, respectively.

The asymptotic regime considered in Proposition 2 is practically relevant when N is small relative to T. Next, we analyze the asymptotic regime in which both N and T are large.

B. Fixed N/T Asymptotic Regime

In the next proposition, we follow El Karoui (2010, 2013) to derive the high-dimensional asymptotic limit of the out-of-sample mean, variance, and utility of the sample mean-variance portfolio $\hat{w}$ as $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$. In this asymptotic regime, a higher ρ is associated with more estimation risk. In this section, we assume that (Assumption-BB) in El Karoui (2013) holds.[6]

Proposition 3. *Let $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$ while θ^2 stays bounded.*

1. *The out-of-sample mean of the sample mean-variance portfolio $\hat{w}$ converges to*

$$\tilde{\mu}_p \xrightarrow{p} \frac{\eta \mu_p}{1-\rho}, \tag{29}$$

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

where the parameter η *is the unique positive solution to*

$$\mathbb{E}\left[(1-\rho+\rho\eta\tau_t)^{-1}\right]=1, \tag{30}$$

and it satisfies $1 \leq \eta \leq \mathbb{E}[1/\tau_t]$.

2. *The out-of-sample variance of the sample mean-variance portfolio* $\hat{w}$ *converges to*

$$\tilde{\sigma}_p^2 \xrightarrow{p} \frac{\varphi\sigma_p^2}{(1-\rho)^3}+\frac{\eta\rho}{\gamma^2(1-\rho)^3}, \tag{31}$$

where the parameter φ *is*

$$\varphi=(1-\rho)\left(\eta^{-2}-\mathbb{E}\left[\frac{\rho\tau_t^2}{(1-\rho+\rho\eta\tau_t)^2}\right]\right)^{-1}, \tag{32}$$

and it satisfies $\eta^2 \leq \varphi \leq (\mathbb{E}[1/\tau_t])^2$.

3. *The out-of-sample utility of the sample mean-variance portfolio* $\hat{w}$ *converges to*

$$U(\hat{w}) \xrightarrow{p} \left(\frac{2\eta}{1-\rho}-\frac{\varphi}{(1-\rho)^3}\right)U(w^\star)-\frac{\eta\rho}{2\gamma(1-\rho)^3}, \tag{33}$$

and this limit is smaller than that under the multivariate normal distribution.

4. *The parameters* η *and* φ *attain their minimum value of one either if asset returns are multivariate normally distributed or if* $\rho \to 0$. *Moreover,* η *and* φ *attain their maximum value of* $\mathbb{E}[1/\tau_t]$ *and* $(\mathbb{E}[1/\tau_t])^2$*, respectively, when* $\rho \to 1$.

Proposition 3 shows that the impact of fat tails on the fixed N/T asymptotic limit of the out-of-sample mean and variance of $\hat{w}$ is controlled by the parameters η and φ. These two parameters increase with the dispersion of the random variable τ_t. Moreover, they determine the increase in the out-of-sample mean and variance of $\hat{w}$ relative to normality, as they are both greater than one and equal to one when returns are multivariate normally distributed. Due to the fact that $\varphi \geq \eta^2 \geq \eta$, we can show that the percentage increase in the asymptotic limit of $\tilde{\sigma}_p^2$ due to fat tails is larger than the corresponding percentage increase in the asymptotic limit of $\tilde{\mu}_p$. Therefore, $U(\hat{w})$ is negatively affected by fat tails, and in the next proposition we show that under the multivariate elliptical distribution it is optimal to allocate less weight to $\hat{w}$ in the two-fund and three-fund rules.

Proposition 4. *Let $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$ while $(\mu_g, \psi^2, \theta^2)$ stay bounded.*

1. *The optimal combination coefficient in the two-fund rule converges to*

$$c^\star \to (1-\rho)^2 \frac{\theta^2}{\frac{\varphi}{\eta}\theta^2 + \rho}. \tag{34}$$

2. *The optimal combination coefficients in the three-fund rule converge to*

$$c_1^\star \to (1-\rho)^2 \frac{\psi^2}{\frac{\varphi}{\eta}\psi^2 + \rho}, \tag{35}$$

$$c_2^\star \to \mu_g (1-\rho)^2 \frac{\frac{\eta}{\varphi}\rho}{\frac{\varphi}{\eta}\psi^2 + \rho}. \tag{36}$$

3. *Given $(\rho, \mu_g, \psi^2, \theta^2)$, the optimal combination coefficients $c^\star$, $c_1^\star$, and $c_2^\star$ attain their maximum value when returns are multivariate normally distributed (i.e., $\varphi = \eta = 1$).*

Proposition 3 shows that fat tails reduce the out-of-sample utility of $\hat{w}$ compared with that under the multivariate normality assumption. It is of interest to see if this continues to be the case for the optimal two-fund and three-fund rules derived in Proposition 4. In the next proposition, we show that when $c^\star$ and $(c_1^\star, c_2^\star)$ are chosen optimally based on the multivariate elliptical distribution, the out-of-sample utility of the two-fund and three-fund rules can actually be larger than that under the normality case under certain conditions.

Proposition 5. *Let $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$ while (ψ^2, θ^2) stay bounded.*

1. *The out-of-sample utility of the optimal two-fund rule converges to*

$$U(\hat{w}_{2f}(c^\star)) \xrightarrow{p} U(w^\star)(1-\rho)\frac{\eta\theta^2}{\frac{\varphi}{\eta}\theta^2+\rho}, \tag{37}$$

   which is larger than that when returns are multivariate normally distributed (i.e., $\varphi = \eta = 1$) if and only if $\theta^2 < \rho\eta(\eta-1)/(\varphi-\eta^2)$.

2. *The out-of-sample utility of the optimal three-fund rule converges to*

$$U(\hat{w}_{3f}(c_1^\star, c_2^\star)) \xrightarrow{p} U(w^\star)(1-\rho)\frac{\eta\psi^2}{\frac{\varphi}{\eta}\psi^2+\rho}\left(1+\frac{\eta\rho\theta_g^2}{\varphi\theta^2\psi^2}\right), \tag{38}$$

   which is larger than that when returns are multivariate normally distributed (i.e., $\varphi = \eta = 1$) if and only if $\eta(\psi^2+\rho)(\theta^2\psi^2+\frac{\eta}{\varphi}\theta_g^2\rho) > (\frac{\varphi}{\eta}\psi^2+\rho)(\theta^2\psi^2+\theta_g^2\rho)$.

Proposition 5 shows that fat tails could potentially help investors obtain better out-of-sample performance. This is because when investors understand that the returns have fat tails, they would reduce their exposure to the sample mean-variance portfolio and the sample GMV portfolio. Although the specific conditions derived in Proposition 5 do not carry much intuition, we show below they are easily satisfied in the case of the multivariate t distribution.

In Section IV, we compare, in the finite-sample setting, the out-of-sample utility of the optimal two-fund and three-rules with the utility obtained when using the combination coefficients that wrongly assume the returns are multivariate normal. Moreover, we do a similar analysis in Section VI using estimated combination coefficients and empirical data.

To illustrate the theoretical results derived in this section, we now assume a specific multivariate elliptical distribution. Specifically, we employ the commonly used multivariate t-distribution, i.e., τ_t is distributed as (2) with $\nu > 2$ degrees of freedom. In the next proposition, we derive the parameters η and φ when returns follow a multivariate t-distribution.

Proposition 6. *Let $N \to \infty$, $T \to \infty$, $N/T \to \rho \in (0,1)$, and asset returns follow a multivariate t-distribution with $\nu > 2$ degrees of freedom.*

1. *The parameter η is the unique positive solution to*

$$ye^y E_{\frac{\nu}{2}}(y) = \rho \quad \textit{with} \quad y = \frac{(\nu-2)\rho\eta}{2(1-\rho)}, \tag{39}$$

   where $E_n(x) = \int_1^\infty t^{-n} e^{-xt} dt$ is the exponential integral, and it satisfies $1 \leq \eta \leq \frac{\nu}{\nu-2}$.

2. *The parameter φ is*

$$\varphi = \frac{2\eta^2(1-\rho)}{\nu - \eta(\nu-2)}, \tag{40}$$

   and it satisfies $\eta^2 \leq \varphi \leq \frac{\nu^2}{(\nu-2)^2}$.

In Figure 1, we depict the parameters η and φ as a function of ρ for $\nu = 4$ and 8. We find that η and φ can be substantially larger than one either when ν is smaller and thus tails are fatter or when ρ is larger and thus there is more estimation risk. Moreover, φ is significantly larger than η and thus fat tails have a larger impact on the out-of-sample variance than on the out-of-sample mean of the sample mean-variance portfolio.

[Insert Figure 1 approximately here]

In Figure 2, we depict how fat tails impact the asymptotic limit of the out-of-sample utility of the sample mean-variance portfolio in (33). We set $\gamma = 1$, $\rho = (0.1, 0.3, 0.5)$, $\theta = (0.3, 0.5)$, and we depict the limit of $U(\hat{w})$ as a function of ν. We observe that fat tails can have a large negative effect on the out-of-sample utility of $\hat{w}$. Thus, the multivariate normality assumption is overly optimistic in assessing the performance of sample mean-variance portfolios.

[Insert Figure 2 approximately here]

In Figure 3, we depict the ratio between the optimal three-fund combination coefficients $(c_1^\star, c_2^\star)$ under the multivariate t-distribution and those under the multivariate normal, i.e.,

$$\frac{c_1^\star}{c_{1,n}^\star} \to \frac{\psi^2 + \rho}{\frac{\varphi}{\eta}\psi^2 + \rho}, \tag{41}$$

$$\frac{c_2^\star}{c_{2,n}^\star} \to \frac{\eta}{\varphi}\left(\frac{\psi^2 + \rho}{\frac{\varphi}{\eta}\psi^2 + \rho}\right). \tag{42}$$

We depict (41)–(42) as a function of ρ for $\nu = (4, 8)$ and $\psi = (0.2, 0.4)$. The figure shows that the higher the ρ and lower the ν, the more the combination coefficients under the multivariate t-distribution depart from those under normality. The difference is especially pronounced for the coefficient $c_2^\star$ on the sample GMV portfolio. This can be explained because while the sample GMV portfolio only depends on $\hat{\Sigma}$, the sample mean-variance portfolio depends on both $\hat{\mu}$ and $\hat{\Sigma}$, and the dependence between $\hat{\mu}$ and $\hat{\Sigma}$ can alleviate the impact of fat tails.[7]

[Insert Figure 3 approximately here]

Finally, we can determine when the out-of-sample utility of the optimal two-fund and three-fund rules is larger under the multivariate t-distribution than under the multivariate normal. Using Propositions 5 and 6, we can show that this is the case for the two-fund rule if

$$\theta^2 < \frac{\rho(1-1/\eta)(\nu-\eta(\nu-2))}{\eta(\nu-2)-\nu+2(1-\rho)}. \tag{43}$$

We find that (43) is typically a large threshold, e.g., $\theta < 0.70$ for $\rho = 0.3$ and $\nu = 8$. Similarly, the condition for the three-fund rule in Proposition 5 can be easily satisfied.

IV. Finite-Sample Results

The high-dimensional asymptotic results in Section III may not deliver a good approximation to the exact finite-sample results when N or T is not large. Therefore, in the next proposition we present the exact expectation of the out-of-sample mean and variance of the sample mean-variance portfolio $\hat{w}$.

Proposition 7. *Let $T > N+4$, $M = I_T - (1_T 1_T^\top)/T$, Y be a $T \times N$ matrix of independent standard normal random variables, $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_T)$ with $\lambda_t = \sqrt{\tau_t}$, and Y and Λ are independent.*

[7] Specifically, the out-of-sample variance of $\hat{\Sigma}^{-1}1_N$ satisfies $1_N^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N \xrightarrow{p} \varphi\theta_g^2/\mu_g^2$ while the out-of-sample variance of $\hat{\Sigma}^{-1}\hat{\mu}$ satisfies $\hat{\mu}^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu} \xrightarrow{p} (\varphi\theta^2+\eta\rho)/(1-\rho)^3$. Given that $\varphi \geq \eta^2 \geq \eta$, the percentage increase in the out-of-sample variance due to fat tails is larger for $\hat{\Sigma}^{-1}1_N$ than $\hat{\Sigma}^{-1}\hat{\mu}$.

The expected out-of-sample mean and variance of the sample mean-variance portfolio $\hat{w}$ are

$$\mathbb{E}[\tilde{\mu}_p] = \frac{T}{T-N-2} k_1 \mu_p, \tag{44}$$

$$\mathbb{E}[\tilde{\sigma}_p^2] = \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \left(k_2 \sigma_p^2 + \frac{k_3}{\gamma^2} \frac{N}{T} \right), \tag{45}$$

where k_1, k_2, and k_3 are functions of N, T, and the distribution of τ_t, but not of μ and Σ:

$$k_1 = \frac{T-N-2}{N} \mathbb{E}\left[\operatorname{tr}\left((Y^\top \Lambda M \Lambda Y)^{-1} \right) \right], \tag{46}$$

$$k_2 = \frac{(T-N-1)(T-N-2)(T-N-4)}{N(T-2)} \mathbb{E}\left[\operatorname{tr}\left((Y^\top \Lambda M \Lambda Y)^{-2} \right) \right], \tag{47}$$

$$k_3 = \frac{(T-N-1)(T-N-2)(T-N-4)}{NT(T-2)} \mathbb{E}\left[1_T^\top \Lambda Y (Y^\top \Lambda M \Lambda Y)^{-2} Y^\top \Lambda 1_T \right], \tag{48}$$

provided the expectations exist. Moreover, k_1, k_2, and k_3 are all equal to one when asset returns are multivariate normally distributed.

Proposition 7 shows that the increase in the out-of-sample mean and variance caused by fat tails is driven by the parameters k_1, k_2, and k_3. It is difficult to find general analytical expressions for these parameters, but they can readily be evaluated using Monte Carlo simulations given the distribution of τ_t. In the fixed N/T asymptotic regime where $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$, we have from Proposition 3 that

$$k_1 \to \eta, \quad k_2 \to \varphi, \quad k_3 \to \eta. \tag{49}$$

In the next proposition, we present the exact optimal two-fund and three-fund rules when returns follow a multivariate elliptical distribution.

Proposition 8. *Let $T > N+4$ and define k_1, k_2, and k_3 as in* (46)–(48).

1. *The optimal combination coefficient in the two-fund rule is*

$$c^{\star} = \frac{(T-N-1)(T-N-4)}{T(T-2)}\left(\frac{k_1\theta^2}{k_2\theta^2 + k_3\frac{N}{T}}\right). \tag{50}$$

2. *The optimal combination coefficients in the three-fund rule are*

$$c_1^{\star} = \frac{(T-N-1)(T-N-4)}{T(T-2)}\left(\frac{k_1\psi^2}{k_2\psi^2 + k_3\frac{N}{T}}\right), \tag{51}$$

$$c_2^{\star} = \mu_g\frac{(T-N-1)(T-N-4)}{T(T-2)}\frac{k_3}{k_2}\left(\frac{k_1\frac{N}{T}}{k_2\psi^2 + k_3\frac{N}{T}}\right). \tag{52}$$

In Figure 4, we compare the exact finite-sample value of (k_1,k_2,k_3) with their asymptotic approximation (η,φ,η), assuming the returns follow a multivariate t-distribution. We set a sample size $T = 100$, a number of degrees of freedom $\nu = (4,8)$, and we vary the number of assets N from 2 to 60. The figure shows that the asymptotic approximation under-estimates the exact value of the parameters (k_1,k_2,k_3), particularly for k_2, and this effect is more pronounced as N and ν get smaller. This means that the asymptotic approximation under-estimates the impact of fat tails on the out-of-sample performance. However, the asymptotic approximation is overall reasonably accurate. This observation, combined with the fact that we can compute (η,φ) in a fast way while computing (k_1,k_2,k_3) requires simulations, makes the asymptotic approximation an attractive alternative.[8]

[8] Whether the asymptotic combination coefficients in Proposition 4 are good approximations for the exact finite-sample ones in Proposition 8 is not only affected by the accuracy of approximating (k_1,k_2,k_3) by (η,φ,η), but also by the accuracy of approximating $(T-N-1)(T-N-4)/(T(T-2))$ by $(1-N/T)^2$.

[Insert Figure 4 approximately here]

In Figure 5, we depict the expected out-of-sample utility loss incurred when using combination coefficients in the two-fund and three-fund rules that assume returns are multivariate normal. Specifically, we calibrate (k_1, k_2, k_3) to the multivariate t-distribution and we depict $\mathbb{E}[U(\hat{w}_{2f}(c_n^\star))] - \mathbb{E}[U(\hat{w}_{2f}(c^\star))]$ and $\mathbb{E}[U(\hat{w}_{3f}(c_{1,n}^\star, c_{2,n}^\star))] - \mathbb{E}[U(\hat{w}_{3f}(c_1^\star, c_2^\star))]$, where $(c_n^\star, c_{1,n}^\star, c_{2,n}^\star)$ are obtained from Proposition 8 with $k_1 = k_2 = k_3 = 1$. To conduct this analysis, we set $\gamma = 1$, $(\psi, \theta) = (0.2, 0.3)$ or $(0.4, 0.5)$, $T = 100$, $N = (10, 25, 50)$ and we vary ν between 3 and 20. The first observation in Figure 5 is that the utility loss is generally small for the two-fund rule, but can nonetheless get important as ν decreases, (ψ, θ) increase, and N increases. The second observation is that the utility loss is much more substantial for the three-fund rule. This finding is consistent with Figure 3 that indicates that the sample GMV portfolio is more affected by fat tails than the sample mean-variance portfolio because $c_2^\star / c_{2,n}^\star$ is significantly smaller than $c_1^\star / c_{1,n}^\star$ in the three-fund rule.

[Insert Figure 5 approximately here]

The analysis in Figure 5 uses known combination coefficients. In practice, the utility loss is also affected by estimation errors in these coefficients. When allowing for such errors, we find in the simulation analysis of Section VI.A that the utility loss incurred when assuming normality becomes much more substantial both in the two-fund rule and the three-fund rule.

To summarize, we now have three different ways of determining the optimal combination coefficients in the two-fund rule and the three-fund rule. First, we can assume returns are multivariate normally distributed and use the exact finite-sample formula, as in Kan and Zhou (2007), which corresponds to setting $k_1 = k_2 = k_3 = 1$ in Proposition 8. Second, we can assume returns are multivariate elliptically distributed and use the fixed N/T asymptotic formula, which corresponds to Proposition 4. Third, we can assume returns are multivariate elliptically distributed and use the exact finite-sample formula, which corresponds to Proposition 8. In Section V, we explain how we estimate the optimal combination coefficients in each case. In Section VI, we compare the out-of-sample performance of the two-fund and three-fund rules using simulated and empirical data. Using simulated data allows us to comply with the multivariate elliptical distribution and to compare the case in which the optimal combination coefficients are known with the case in which they are estimated. This, in turn, allows us to disentangle the impact of fat tails and the impact of estimation errors in the combination coefficients. Using empirical data allows us to assess the out-of-sample performance of the two-fund and three-fund rules when the data is not exactly multivariate elliptically distributed and when transaction costs play a role.

V. Estimation of the Two-Fund and Three-Fund Rules

In this section, we explain how we estimate the parameters needed to determine the optimal two-fund and three-fund combination coefficients. There are three types of parameters. First, the parameters μ_g, θ^2, and ψ^2 that depend on μ and Σ. Second, the parameters η and φ in the fixed N/T asymptotic regime that depend on the distribution of τ_t but not on μ and Σ. Third, the parameters k_1, k_2, and k_3 in the finite-sample case that also depend on the distribution of τ_t but not on μ and Σ. In Table 1, we provide formulas for the estimated two-fund and three-fund combination coefficients under different calibration methods.

[Insert Table 1 approximately here]

Because of estimation risk in the combination coefficients, it is unclear how much of the theoretical expected out-of-sample utility of the optimal two-fund and three-fund rules can be realized in practice. To address this point, in the simulation analysis of Section VI.A we compare the cases with known and estimated combination coefficients.

A. Estimation of μ_g, θ^2, and ψ^2

We estimate μ_g, θ^2, and ψ^2 as in Kan and Zhou (2007). Specifically, we use the plug-in estimator of μ_g and the adjusted estimators of θ^2 and ψ^2, which are defined as

$$\text{(53)} \qquad \hat{\mu}_g = \frac{1_N^\top \hat{\Sigma}^{-1}\hat{\mu}}{1_N^\top \hat{\Sigma}^{-1} 1_N},$$

$$\text{(54)} \qquad \hat{\theta}_a^2 = \frac{(T-N-2)\hat{\theta}^2 - N}{T} + \frac{2(\hat{\theta}^2)^{N/2}(1+\hat{\theta}^2)^{-(T-2)/2}}{T \times B_{\hat{\theta}^2/(1+\hat{\theta}^2)}(N/2,(T-N)/2)},$$

$$\text{(55)} \qquad \hat{\psi}_a^2 = \frac{(T-N-1)\hat{\psi}^2 - (N-1)}{T} + \frac{2(\hat{\psi}^2)^{(N-1)/2}(1+\hat{\psi}^2)^{-(T-2)/2}}{T \times B_{\hat{\psi}^2/(1+\hat{\psi}^2)}((N-1)/2,(T-N+1)/2)},$$

where $\hat{\theta}^2$ and $\hat{\psi}^2$ and the plug-in estimators of θ^2 and ψ^2, respectively, and $B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t$ is the incomplete beta function. The rationale behind the adjusted estimators in (54)–(55) is to adjust the unbiased estimator, i.e., the first term, so as to prevent the estimator from becoming negative. In Section II of the supplementary material, we show that even though the adjusted estimators $\hat{\theta}_a^2$ and $\hat{\psi}_a^2$ are derived under the multivariate normal distribution, their root mean squared error does not increase much with the fat tails of returns.

B. Estimation of η and φ in the Asymptotic Setting

We consider two different methods to estimate η and φ. The first method is to assume a specific multivariate elliptical distribution, and we employ the multivariate t-distribution, i.e., τ_t is distributed as (2) with $\nu > 2$ degrees of freedom. We estimate ν by the maximum likelihood method, giving us $\hat{\nu}$, whose estimation accuracy we analyze in Section I of the supplementary material. This allows us to obtain estimates of η and φ using Proposition 6, which we denote $\hat{\eta}$ and $\hat{\varphi}$. Specifically, $\hat{\eta}$ is the unique positive solution to

$$ye^y E_{\frac{\hat{\nu}}{2}}(y) = \frac{N}{T} \quad \text{with} \quad y = \frac{(\hat{\nu}-2)N\hat{\eta}}{2(T-N)}, \tag{56}$$

and the estimate of φ is

$$\hat{\varphi} = \frac{2\hat{\eta}^2(1-N/T)}{\hat{\nu}-\hat{\eta}(\hat{\nu}-2)}. \tag{57}$$

Calibrating η and φ to the multivariate t-distribution implies a larger risk of model misspecification than assuming the more general class of multivariate elliptical distributions. Therefore, the second method we consider is to estimate η and φ directly from the data without specifying which particular multivariate elliptical distribution the returns follow. The principle is to rely on a sample distribution of τ_t. Specifically, suppose we have T sample observations $(\hat{\tau}_1, \ldots, \hat{\tau}_T)$. Then, $\tilde{\eta}$ is the unique positive solution to

$$\sum_{t=1}^{T} \frac{1}{T-N+N\tilde{\eta}\hat{\tau}_t} = 1, \tag{58}$$

and the estimate of φ is

$$\tilde{\varphi} = \left(1-\frac{N}{T}\right)\left(\tilde{\eta}^{-2} - \sum_{t=1}^{T} \frac{N\hat{\tau}_t^2}{(T-N+N\hat{\tau}_t\tilde{\eta})^2}\right)^{-1}. \tag{59}$$

El Karoui (2010, 2013) propose using the following sample distribution of τ_t:

$$\hat{\tau}_t = \frac{(r_t - \hat{\mu})^\top (r_t - \hat{\mu})}{\frac{1}{T}\sum_{i=1}^{T}(r_i - \hat{\mu})^\top (r_i - \hat{\mu})}, \quad t = 1, \dots, T, \tag{60}$$

which he shows is a consistent estimator of the true distribution of τ_t as $N \to \infty$. In Section I of the supplementary material, we illustrate how the distribution of $\hat{\tau}_t$ approaches that of τ_t as N increases.

Figure 6 compares the estimators $(\hat{\eta}, \hat{\varphi})$ and $(\tilde{\eta}, \tilde{\varphi})$. We set $N = 25$, $T = (60, 120, 240)$, and we simulate 10,000 return vectors of size T from a multivariate t-distribution with $\nu = (4, 6, 8)$.[9] This allows us to generate 10,000 estimates $(\hat{\eta}, \hat{\varphi})$ and $(\tilde{\eta}, \tilde{\varphi})$, whose distribution we summarize using boxplots. We find that the estimates are more accurate as ν and T increase. Moreover, even though $(\hat{\eta}, \hat{\varphi})$ assume the right distribution, they display a similar level of bias and variance as $(\tilde{\eta}, \tilde{\varphi})$. This result, combined with the fact that $(\tilde{\eta}, \tilde{\varphi})$ can adapt to other elliptical distributions than multivariate t, suggests that using $(\tilde{\eta}, \tilde{\varphi})$ is preferable in practice. Our empirical analysis confirms this finding. In particular, when using empirical data in Section VI.B, we find that $(\tilde{\eta}, \tilde{\varphi})$ are generally larger than $(\hat{\eta}, \hat{\varphi})$, which suggests that fat tails in empirical data have a larger impact on the out-of-sample mean and variance than the multivariate t-distribution suggests. Therefore, using $(\tilde{\eta}, \tilde{\varphi})$ instead of $(\hat{\eta}, \hat{\varphi})$ allows us to improve the performance of the two-fund and three-fund rules in practice.

[Insert Figure 6 approximately here]

[9] We set $(\mu, \Sigma) = (0_N, I_N)$ without loss of generality because η and φ do not depend on (μ, Σ).

C. Estimation of k_1, k_2, and k_3 in the Finite-Sample Setting

We estimate (k_1, k_2, k_3) using two similar methods to those used for estimating η and φ. First, we calibrate them to the multivariate t-distribution whose degrees of freedom ν is estimated by maximum likelihood. To achieve this, we simulate 1,000 matrices Y_m and Λ_m independently, where $m = 1, \ldots, 1{,}000$. Each Y_m is a $T \times N$ matrix of independent standard normal variables and each $\Lambda_m = \text{Diag}(\sqrt{\tau_1}, \ldots, \sqrt{\tau_T})$, where the τ_t's are simulated from (2) with $\nu = \hat{\nu}$. Then, we estimate (k_1, k_2, k_3) using the sample counterparts of (46)–(48), i.e.,

(61)
$$\hat{k}_1 = \frac{T-N-2}{N}\left[\frac{1}{1{,}000}\sum_{m=1}^{1{,}000}\text{tr}\Big((Y_m^\top \Lambda_m M \Lambda_m Y_m)^{-1}\Big)\right],$$

(62)
$$\hat{k}_2 = \frac{(T-N-1)(T-N-2)(T-N-4)}{N(T-2)}\left[\frac{1}{1{,}000}\sum_{m=1}^{1{,}000}\text{tr}\Big((Y_m^\top \Lambda_m M \Lambda_m Y_m)^{-2}\Big)\right],$$

(63)
$$\hat{k}_3 = \frac{(T-N-1)(T-N-2)(T-N-4)}{NT(T-2)}\left[\frac{1}{1{,}000}\sum_{m=1}^{1{,}000} 1_T^\top \Lambda_m Y_m (Y_m^\top \Lambda_m M \Lambda_m Y_m)^{-2} Y_m^\top \Lambda_m 1_T\right].$$

Second, we estimate (k_1, k_2, k_3) without specifying a particular elliptical distribution by using the sample distribution of τ_t by El Karoui (2010, 2013), $\hat{\tau}_t$ in (60). Specifically, using the fact that Y and Λ are independent, we first estimate Λ by $\hat{\Lambda} = \text{Diag}(\sqrt{\hat{\tau}_1}, \ldots, \sqrt{\hat{\tau}_T})$. Then, we simulate 1,000 matrices Y_m, $m = 1, \ldots, 1{,}000$, and estimate (k_1, k_2, k_3) as

$$\tilde{k}_1 = \frac{T-N-2}{N}\left[\frac{1}{1{,}000}\sum_{m=1}^{1{,}000}\text{tr}\Big((Y_m^\top \hat{\Lambda} M \hat{\Lambda} Y_m)^{-1}\Big)\right], \quad (64)$$

$$(65)\quad \tilde{k}_2 = \frac{(T-N-1)(T-N-2)(T-N-4)}{N(T-2)}\left[\frac{1}{1{,}000}\sum_{m=1}^{1{,}000}\mathrm{tr}\Big((Y_m^\top \hat{\Lambda} M \hat{\Lambda} Y_m)^{-2}\Big)\right],$$

$$(66)\quad \tilde{k}_3 = \frac{(T-N-1)(T-N-2)(T-N-4)}{NT(T-2)}\left[\frac{1}{1{,}000}\sum_{m=1}^{1{,}000} 1_T^\top \hat{\Lambda} Y_m (Y_m^\top \hat{\Lambda} M \hat{\Lambda} Y_m)^{-2} Y_m^\top \hat{\Lambda} 1_T\right].$$

In unreported results, we compare the estimates of (k_1, k_2, k_3) according to these two different methods and reach similar conclusions to those drawn for the comparison of the estimates of (η, φ) in Figure 6. Therefore, in our empirical analysis we also find that the approach that does not require specifying a particular elliptical distribution to estimate (k_1, k_2, k_3) instead of specifying a multivariate t-distribution improves the performance of the two-fund and three-fund rules.

VI. Performance Analysis

We now evaluate the out-of-sample performance of the two-fund and three-fund rules that are calibrated to fat-tailed elliptical returns using our asymptotic and finite-sample theory presented in Sections III and IV, respectively, and the estimation procedure laid out in Section V.

A. Simulated Data

In this section, we simulate i.i.d. excess returns from a multivariate t-distribution with $\nu = 4$, 6, or 8 degrees of freedom. We calibrate (μ, Σ) to a dataset of $N = 25$ portfolios of firms sorted on size and book-to-market spanning July 1926 to July 2023. When asset returns are multivariate elliptically distributed, we can show that the out-of-sample utility of the two-fund and three-fund rules only depends on (μ, Σ) via (μ_g, ψ, θ), and for this dataset we have

$(\mu_g, \psi, \theta) = (0.0075, 0.250, 0.302)$ in monthly terms. We use a non-large value of $N = 25$ to be able to identify non-negligible differences between portfolios derived under the fixed N/T asymptotic regime in Section III.B and the finite-sample setting in Section IV. We can nonetheless consider high degrees of estimation risk by varying the sample size T, which is why we consider both a small sample size of $T = 60$ months and larger values of $T = 120$ and 240 months.

We report the expected out-of-sample utility delivered by two-fund and three-fund rules with combination coefficients calibrated in six different ways. First, the combination coefficients are calibrated to the multivariate normal distribution as in Kan and Zhou (2007). Second, the coefficients are calibrated to the multivariate t-distribution using the asymptotic formula and a number degrees of freedom ν estimated by maximum likelihood. Third, the coefficients are calibrated to the multivariate t-distribution using the exact finite-sample formula and ν estimated in the same way. Fourth, the coefficients are calibrated to the multivariate elliptical distribution using the asymptotic formula and the sample distribution of τ_t by El Karoui (2010, 2013). Fifth, the coefficients are calibrated to the multivariate elliptical distribution using the exact finite-sample formula and the sample distribution of τ_t. Sixth, the coefficients are fully calibrated from the data by using a five-fold cross-validation.[10] Formulas for the combination coefficients are

[10]The five-fold cross-validation method for the two-fund rule works as follows; the method is similar for the three-fund rule but over two coefficients instead of one. First, set $c = 0$. Then, divide the estimation window in five intervals of equal length. Select one interval, remove it from the estimation window, and use the remaining data to compute the two-fund rule. Use these portfolio weights to compute out-of-sample returns on the removed interval. After repeating this procedure for all five intervals, compute the utility on all the out-of-sample returns. Increase c by 0.01 and repeat the process above. Finally, when $c = 1$ is reached, select the value of c delivering the maximum out-of-sample utility.

available in Table 1.

We compute the expected out-of-sample utility of a given portfolio strategy $\hat{w}$ by simulating $M = 10{,}000$ multivariate t-distributed returns and obtaining an estimated portfolio $\hat{w}_m$ for each simulation, where $m = 1, \ldots, M$. Then, given the true μ and Σ, we estimate the expected out-of-sample utility of $\hat{w}$ as

$$\mathbb{E}[U(\hat{w})] \approx \frac{1}{M}\left[\sum_{m=1}^{M} \hat{w}_m^\top \mu - \frac{\gamma}{2}\hat{w}_m^\top \Sigma \hat{w}_m\right]. \tag{67}$$

We consider a risk-aversion coefficient $\gamma = 1$. Given that the utility of each portfolio rule we consider is proportional to $1/\gamma$, the choice of γ does not affect their rankings.[11]

In Table 2, we report the annualized expected out-of-sample utility of the different two-fund and three-fund rules as well as the mean value of the combination coefficients.[12] To disentangle the impact of fat tails and the impact of estimation errors in the combination coefficients, we report results for the case with known and estimated combination coefficients.

[Insert Table 2 approximately here]

Our first observation is that when the combination coefficients are known, the out-of-sample utility loss incurred when calibrating the combination coefficients to the multivariate normal distribution instead of the multivariate t-distribution is small for the two-fund rule but can be substantial for the three-fund rule. This finding is consistent with the conclusion drawn from Figure 5. For example, when $\nu = 4$ and we use the exact finite-sample combination

[11]The expected out-of-sample utility for other values of γ are simply obtained by dividing the values we obtain with $\gamma = 1$ by the desired value of γ.

[12]We annualize the monthly utilities by multiplying them by 12.

coefficients, the annualized out-of-sample utility loss for the three-fund rule is 4.4, 2.8, and 1.5 percentage points for a sample size $T = 60$, 120, and 240 months, respectively. When $\nu = 6$, this loss decreases to 1.1, 0.7, and 0.4 percentage points, respectively.

However, the corresponding out-of-sample utility loss is substantially larger when the combination coefficients are estimated. For example, when $\nu = 6$ and we use the exact finite-sample combination coefficients, the annualized out-of-sample utility loss when calibrating the two-fund rule to the multivariate normal distribution instead of the multivariate elliptical distribution is 12.2, 4.6, and 1.7 percentage points for a sample size $T = 60$, 120, and 240 months, respectively. For the three-fund rule, the utility loss increases to 21.3, 7.7, and 2.4 percentage points, respectively.

The reason why the out-of-sample utility loss due to the multivariate normality assumption is larger when the combination coefficients are estimated instead of known is that the combination coefficients are larger when they are calibrated to the multivariate normal distribution. This makes the two-fund and three-fund rules more exposed to estimation errors in the combination coefficients when these coefficients are calibrated to the multivariate normal distribution than to a multivariate elliptical distribution. To alleviate the impact of estimation errors in the combination coefficients, Kan and Wang (2023) show that shrinking them helps, and calibrating the combination coefficients to the multivariate elliptical distribution has this shrinkage effect relative to normality. Moreover, this shrinkage of the combination coefficients when switching from multivariate normal to multivariate elliptical is more substantial the lower the sample size and the fatter the tails, which is exactly when estimation errors in the combination coefficients are important.[13] This is why, in Table 2, the out-of-sample utility loss due to the

[13]The accuracy of the estimates of the parameters $(\mu_g, \psi^2, \theta^2)$ is negatively impacted by fat tails. For example,

normality assumption is larger as T and ν decrease.[14]

In summary, both the multivariate normality assumption and the importance of shrinking the combination coefficients to alleviate their estimation risk explain the utility loss incurred by calibrating the two-fund and three-fund rules to the multivariate normal distribution instead of the multivariate t or elliptical distribution, but it is the second factor that plays a more prominent role.

Our second observation is that when comparing portfolio rules that use combination coefficients obtained from the finite-sample formula versus the asymptotic formula, we find that the former consistently delivers higher expected out-of-sample utility than the latter. The gain in performance is particularly important when the sample size is small, i.e., $T = 60$. For $T = 120$, and particularly $T = 240$, the utility gain is more limited, and in such cases one could opt to use the asymptotic formula that admits a fast analytical solution.

Our third observation is that the estimated combination coefficients are smaller when they are calibrated to the multivariate t and elliptical distributions than to the multivariate normal distribution, consistent with Proposition 4. For example, when $\nu = 6$ and the coefficients are found with the exact finite-sample formula, the mean value of $\hat{c}_1$ in the three-fund rule when the

El Karoui (2010) shows that in the fixed N/T asymptotic regime, the plug-in estimator of θ^2 satisfies $\hat{\theta}^2 \xrightarrow{p} (\eta\theta^2 + \rho)/(1-\rho)$. That is, the bias of $\hat{\theta}^2$ increases with fat tails via η. In Section II of the supplementary material, we evaluate the root mean square error of the adjusted estimators $\hat{\theta}_a^2$ and $\hat{\psi}_a^2$ and find that it increases as T and ν decrease.

[14]A similar reasoning explains why calibrating the combination coefficients to the general multivariate elliptical distribution instead of the multivariate t-distribution improves the expected out-of-sample utility even if the asset returns are multivariate t-distributed. This is because the combination coefficients are smaller under the multivariate elliptical distribution, which reduces the impact of estimation errors in these coefficients on the performance of the two-fund and three-fund rules.

sample size $T = 60$, 120, and 240 months is (0.079, 0.179, 0.303) for multivariate normal, (0.072, 0.167, 0.287) for multivariate t, and (0.054, 0.140, 0.253) for multivariate elliptical. As observed in Figure 3, the difference is more substantial for the coefficient $\hat{c}_2/\hat{\mu}_g$. Its mean value when the sample size $T = 60$, 120, and 240 months is (0.224, 0.425, 0.488) for multivariate normal, (0.162, 0.331, 0.411) for multivariate t, and (0.068, 0.190, 0.285) for multivariate elliptical. As explained above, this shrinkage of combination coefficients helps alleviate the impact of both fat tails and estimation errors in the combination coefficients.

Our fourth observation is that the calibration of the two-fund and three-fund rules to the multivariate elliptical distribution outperforms that to the multivariate t-distribution, and they both outperform the calibration to the multivariate normal distribution. The difference is particularly substantial for a small sample size of $T = 60$, in which case the multivariate normal and multivariate t-distributions deliver negative utilities while the multivariate elliptical always delivers positive utilities. For example, when $\nu = 6$, $T = 60$, and we use the combination coefficients found with the exact finite-sample formula, the two-fund and three-fund rules deliver utilities of $(-0.087, -0.159)$ for multivariate normal, $(-0.039, -0.055)$ for multivariate t, but (0.035, 0.054) for multivariate elliptical. Moreover, the improvement is important for $T = 120$ too with utilities of (0.071, 0.073) for multivariate normal, (0.091, 0.116) for multivariate t, and (0.117, 0.150) for multivariate elliptical. That the calibration of the two-fund and three-fund rules to the multivariate elliptical distribution outperforms that to the multivariate t-distribution seems surprising given that the true distribution is multivariate t. This can be explained because the combination coefficients are larger under the multivariate t-distribution and, as explained above, this makes the two-fund and three-fund rules more exposed to estimation errors in the combination coefficients.

Our fifth observation is that the cross-validation method outperforms the calibration to the multivariate normal and t distributions if the sample size is small enough and the data is fat-tailed enough. Specifically, it outperforms the calibration to the multivariate normal distribution when $T = 60$ for all degrees of freedom ν, when $T = 120$ for $\nu = 4$ and 6, and when $T = 240$ for $\nu = 4$. Cross-validation also outperforms the calibration to the multivariate t-distribution when $T = 60$. In contrast, cross-validation systematically underperforms the calibration to the multivariate elliptical distribution. This suggests that there are benefits from relying on parametric assumptions instead of a fully data-driven approach, but as long as the parametric assumptions are not too restrictive.

Our final observation is that the three-fund rule consistently outperforms the two-fund rule when the sample size $T = 120$ and 240, but can underperform when the sample size $T = 60$. We can explain this result because there is an additional combination coefficient to estimate in the three-fund rule, and when the sample size is small, the estimation risk induced by this additional coefficient can be large enough to render the three-fund rule less attractive. Interestingly, when $T = 60$, the three-fund rule underperforms the two-fund rule when the combination coefficients are calibrated to the multivariate normal and multivariate t-distributions, but it outperforms the two-fund rule when the combination coefficients are calibrated to the multivariate elliptical distribution.

B. Empirical Data

We now evaluate the out-of-sample performance of the two-fund and three-fund rules considered in Section VI.A using empirical data. We also report the out-of-sample performance of the equally weighted portfolio $w_{ew} = 1_N/N$,[15] the optimal combination of the equally weighted portfolio with the risk-free asset (see DeMiguel et al. (2009, Appendix B)),

$$\hat{w}_{ewrf} = \frac{1_N^\top \hat{\mu}}{\gamma 1_N^\top \hat{\Sigma} 1_N} 1_N, \tag{68}$$

and the two funds composing the two-fund and three-fund rules, i.e., the sample GMV portfolio,

$$\hat{w}_g = \frac{\hat{\Sigma}^{-1} 1_N}{1_N^\top \hat{\Sigma}^{-1} 1_N}, \tag{69}$$

and the sample mean-variance portfolio in (15). The risk-free asset is also a fund in the two-fund and three-fund rules we consider, and it delivers a utility of zero by definition. The sample mean-variance portfolio is by far the worst-performing strategy, and thus, we do not discuss it further below. We also consider additional empirical tests in Section VI.C.

We consider six different datasets of monthly returns that we obtain from Kenneth French's and Robert Novy-Marx's websites. First, a dataset of 10 portfolios of firms sorted on momentum spanning January 1927 to July 2023 (10MOM). Second, a dataset composed of the long and short legs of eight low-turnover anomalies in Novy-Marx and Velikov (2016) spanning July 1963 to December 2013 (16ANOM). Third, a dataset of 25 portfolios of firms sorted on size and market beta spanning July 1963 to July 2023 (25SBETA). Fourth, a dataset of 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to July 2023 25SBTM. Fifth, a dataset of 25 portfolios of firms sorted on operating profitability and investment spanning July 1963 to July 2023 (25OPINV). Sixth, a dataset of 30 industry portfolios spanning July 1926 to

[15] The performance of the equally weighted portfolio is close to that of the value-weighted market portfolio.

July 2023 (30IND).

In Table 3, we report for each dataset the number of degrees of freedom estimated by maximum likelihood, $\hat{\nu}$, which we show in Section I of the supplementary material is an accurate estimator when the sample size is large as it is the case here over the whole sample. We observe that fat tails are pronounced as $\hat{\nu}$ varies between roughly 4 and 8 across datasets. Given that tails get fatter as the return frequency increases, we also use daily data in Section VI.C.3 for which $\hat{\nu}$ varies between roughly 3 and 6.

[Insert Table 3 approximately here]

[Insert Table 4 approximately here]

We evaluate the out-of-sample utility of the considered portfolio strategies using a standard rebalancing procedure. For a given month, we estimate the portfolio weights based on T months of historical data and evaluate the out-of-sample portfolio return on the next month. We then roll the estimation window forward by one month and proceed iteratively until we reach the end of the sample.[16] We finally compute the out-of-sample utility from the time series of out-of-sample portfolio returns. Although transaction costs do not enter the optimization of our portfolio rules, these costs matter in practice. Therefore, we also evaluate the net-of-cost out-of-sample utility by using the standard academic practice of adjusting the out-of-sample returns to transaction costs proportional to turnover, as in DeMiguel et al. (2009) and Kan et al. (2021), for instance. As in Ao et al. (2019), we use proportional transaction costs of 10 basis points, in line with Engle, Ferstenberg and Russell (2012) who find an average cost of 8.8 basis

[16]We use rolling windows of a fixed size instead of expanding windows for consistency with the theory in Section IV and to be able to test the effect of using both small and large fixed sample sizes.

points for the NYSE stocks.[17] As in Section VI.A, we consider a risk-aversion coefficient $\gamma = 1$ and a sample size $T = 60$, 120, and 240 months.

In Table 3, we report the gross and net-of-cost annualized out-of-sample utility of the two-fund and three-fund rules estimated in six different ways.[18] We also report the mean value of the combination coefficients. All the portfolios are constructed using the sample mean $\hat{\mu}$ and the sample covariance matrix $\hat{\Sigma}$ for consistency with the theory. This should be considered a starting point as the portfolios can then be combined with shrinkage estimators of the inputs as in Kan et al. (2021) to further improve performance and reduce portfolio turnover. The main conclusions drawn from simulated data in Section VI.A hold as well for empirical data. First, the exact finite-sample combination coefficients outperform the asymptotic ones, and substantially so when $T = 60$.

Second, the combination coefficients are smaller when they are calibrated to the multivariate elliptical distribution than to the multivariate t-distribution, and even smaller than those calibrated to the multivariate normal distribution. On average across the six datasets, the mean value of the elliptical two-fund combination coefficient $\hat{c}$ is 29.7%, 23.8%, and 18.2% smaller than the corresponding ones under the multivariate normal assumption for $T = 60$, 120, and 240, respectively. For the three-fund combination coefficient $\hat{c}_2$, the decrease is even larger

[17]Transaction costs are larger in the older part of the sample, but our objective is to adjust returns to transaction costs of a magnitude that investors encounter nowadays. Moreover, our computation of transaction costs should be recognized as a first-order approximation as it ignores turnover inside the basis portfolios themselves and heterogeneity in transaction costs across stocks.

[18]Given that we recognize the fat tails of returns, in Section III of the supplementary material we also report the skewness and excess kurtosis of the two-fund and three-fund rules calibrated either to the multivariate normal or elliptical distribution. We find that whereas the trend for skewness is unclear, calibrating the portfolio rules to the multivariate elliptical distribution helps reduce kurtosis significantly.

with a reduction of 61.7%, 49.4%, and 38.3% for $T = 60$, 120, and 240, respectively.

To understand why the combination coefficients are smaller when they are calibrated to the multivariate elliptical distribution than to the multivariate t and multivariate normal distributions, in Figure 7 we depict boxplots of the estimates of the parameters (η, φ) and (k_1, k_2, k_3) for the 25SBTM dataset. These parameters control the increase in the out-of-sample mean and variance relative to the normality case in the asymptotic and finite-sample setting, respectively. We get two main findings from Figure 7. Firstly, these parameters can be significantly larger than one, and this is particularly the case for k_2. Given that k_2 increases the out-of-sample variance, it is optimal to lower the combination coefficients when accounting for fat tails. Secondly, whereas calibrating these parameters to the multivariate t or the multivariate elliptical distribution leads to similar values when asset returns are multivariate t-distributed in Figure 6, under empirical data (η, φ) and (k_1, k_2, k_3) can be significantly larger when they are calibrated to the multivariate elliptical distribution. This suggests that fat tails in empirical data have a larger impact on the out-of-sample mean and variance than the multivariate t-distribution suggests. As a result, the combination coefficients in the two-fund and three-fund rules are lower under the multivariate elliptical distribution compared with multivariate t, and as we explain below, this helps improve performance.

[Insert Figure 7 approximately here]

Third, the calibration of the two-fund and three-fund rules to the multivariate elliptical distribution outperforms that to the multivariate t-distribution, which in turn outperforms that to the multivariate normal distribution in Kan and Zhou (2007). On average across datasets, the increase in the gross annualized out-of-sample utility when switching from multivariate normal to multivariate elliptical is 28.7, 10.7, and 8.65 percentage points for $T = 60$, 120, and 240, respectively. In Figure 8, we depict the net-of-cost annualized out-of-sample utility delivered by the two-fund and three-fund rules calibrated to the multivariate normal, t, and elliptical distributions. We find that the multivariate elliptical distribution delivers the best performance and that the multivariate normal and t-distributions deliver negative utilities when $T = 60$. Moreover, when $T = 120$ and 240, the three-fund rule calibrated to the multivariate elliptical distribution always delivers a positive net utility, which is not always true when it is calibrated to the multivariate normal and t-distributions.

[Insert Figure 8 approximately here]

Fourth, there are a few cases in which five-fold cross-validation provides the best performance, and it is often superior to the two-fund and three-fund rules calibrated to the multivariate normal distribution. However, in most cases it underperforms the two-fund and three-fund rules calibrated to the multivariate elliptical distribution. This suggests, as in the simulated data, that not overly restrictive parametric assumptions are valuable over a fully data-driven approach.[19]

Fifth, the three-fund rule outperforms the two-fund rule when the sample size $T = 120$ and 240. However, when the sample size decreases to $T = 60$, it is only when we use the multivariate elliptical distribution or cross-validation that adding the sample GMV portfolio in the three-fund

[19] In unreported results, we find that the combination coefficients calibrated with five-fold cross-validation are much more unstable over time than those calibrated with the multivariate elliptical distribution. This instability hurts out-of-sample performance and explains why the two-fund and three-fund rules calibrated with cross-validation often incur larger performance losses due to transaction costs even though they often take a higher exposure to the risk-free asset than the portfolio rules calibrated with the elliptical distribution.

rule outperforms the two-fund rule. This result highlights again the importance of using parametric assumptions that are not too restrictive. On average across datasets, the annualized gross utility gain when switching from the two-fund rule to the three-fund rule, when we use combination coefficients calibrated to the multivariate elliptical distribution and the exact finite-sample formula, is 2.75, 4.25, and 6.11 percentage points for $T = 60$, 120, and 240, respectively.

Finally, another benefit of calibrating the two-fund and three-fund rules to the multivariate elliptical distribution rather than to the multivariate normal distribution is that it delivers lower transaction costs because smaller combination coefficients induce less extreme positions in the risky assets. Specifically, on average across datasets, the annualized out-of-sample utility loss due to transaction costs suffered by the two-fund and three-fund rules calibrated to the multivariate normal distribution is 14.5, 8.27, and 4.87 percentage points for $T = 60$, 120, and 240, respectively. For the two-fund and three-fund rules calibrated to the multivariate elliptical distribution using the exact finite-sample formula, this utility loss decreases to 10.2, 6.33, and 4.21 percentage points for $T = 60$, 120, and 240, respectively. This reduction in transaction costs makes that the two-fund and three-fund rules calibrated to the multivariate elliptical distribution substantially outperform the equally weighted portfolio (combined with the risk-free asset or not) in Table 4 even after transaction costs. In contrast, there are many cases where the two-fund and three-fund rules calibrated to the multivariate normal distribution underperform the equally weighted portfolios.

C. Additional Empirical Results

We consider three additional empirical results in this section. First, we compare the in-sample versus out-of-sample performance of the two-fund and three-fund rules. Second, we study the optimal combination of the sample GMV portfolio with the risk-free asset. Third, we use daily instead of monthly data. The corresponding tables are in Section III of the supplementary material.

1. In-Sample versus Out-of-Sample Performance

We run a similar exercise to that in Barroso and Saxena (2021) by comparing the in-sample versus out-of-sample mean return, volatility, and utility of the two-fund and three-fund rules. Barroso and Saxena (2021) observe that the in-sample risk of the three-fund rule of Kan and Zhou (2007), calibrated to the multivariate normal distribution, substantially under-estimates the out-of-sample portfolio risk. Therefore, our interest is to evaluate whether part of this result can be attributed to ignoring the impact of fat tails on the out-of-sample portfolio performance. We evaluate this by testing whether our two-fund and three-fund rules calibrated to the multivariate elliptical distribution provide a better prediction of the out-of-sample realized mean return, volatility, and utility than the portfolio rules calibrated to the multivariate normal distribution.

For each of the two-fund and three-fund rules calibrated to the multivariate normal, t, or elliptical distribution using the exact finite-sample formula, we compute the difference between 1) the average in-sample annualized mean return, volatility, and utility over all estimation windows of size $T = 60$, 120, and 240 months versus 2) the corresponding out-of-sample realized statistic. The difference for the volatility is always negative and we report it positively. Thus, the lower the differences, the better. We report the results in Table 2 of the supplementary material. We observe that although there is a substantial difference between in-sample versus out-of-sample return statistics, as is typical for optimized mean-variance portfolios (Kan et al., 2024), the mismatch is also smaller when the two-fund and three-fund rules are calibrated to the multivariate elliptical distribution, and often substantially so. Thus, incorporating fat tails in the calibration of portfolio rules by increasing their exposure to the risk-free asset is helpful both for out-of-sample performance and prediction.

2. Combining the Sample GMV Portfolio with the Risk-Free Asset

We consider an alternative portfolio combination rule, the scaled GMV portfolio, that combines the sample GMV portfolio with the risk-free asset:

$$\hat{w}_g(c) = \frac{c}{\gamma}\hat{\Sigma}^{-1}1_N. \tag{70}$$

The scaled GMV portfolio ignores the sample mean return $\hat{\mu}$, which may improve the out-of-sample performance compared with the two-fund and three-fund rules because of the large estimation risk in $\hat{\mu}$. Moreover, such a portfolio rule may appeal to practitioners because ignoring $\hat{\mu}$ typically implies less extreme portfolio weights and lower transaction costs.

Under i.i.d. multivariate normality, the optimal combination coefficient is

$$c_n^{\star} = \mu_g \frac{(T-N-1)(T-N-4)}{T(T-2)}. \tag{71}$$

In the next proposition, we derive the optimal combination coefficient under the multivariate elliptical distribution in the fixed N/T asymptotic regime and in the finite-sample case.

Proposition 9. *The optimal combination coefficient $c^\star$ that maximizes the expected out-of-sample utility of the scaled GMV portfolio $\hat{w}_g(c)$ in* (70) *is:*

1. *When $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0,1)$,*

$$c^\star \to \mu_g (1-\rho)^2 \frac{\eta}{\varphi}. \tag{72}$$

2. *When N and T are finite,*

$$c^\star = \mu_g \frac{(T-N-1)(T-N-4)}{T(T-2)} \frac{k_1}{k_2}. \tag{73}$$

In Table 3 of the supplementary material, we use our six empirical datasets to evaluate the out-of-sample utility of the optimal scaled GMV portfolio with the combination coefficient calibrated in six different ways as in Section VI.B. In Table 4, we also report the out-of-sample utility of the sample GMV portfolio without a risk-free asset. The conclusions drawn in Section VI.B are overall robust to using this alternative portfolio combination rule. In particular, the combination coefficient calibrated to the multivariate elliptical distribution using the finite-sample formula generally delivers the best performance. This is particularly true when the sample size $T = 60$, in which case calibrating the combination coefficient to the multivariate normal delivers a negative net-of-cost utility while the multivariate elliptical distribution delivers a positive net-of-cost utility.

Interestingly, when $T = 60$ and the combination coefficients are calibrated to the multivariate normal distribution, the scaled GMV portfolio considered in this section outperforms the three-fund rule in Table 3. The average gain in annualized out-of-sample utility across datasets is 13.5 percentage points before transaction costs and 18.1 percentage points after transaction costs. This is because the combination coefficient on the sample mean-variance portfolio in the three-fund rule, which exploits expected returns, is too large when we ignore the impact of fat tails. In contrast, when the combination coefficients are calibrated to the multivariate elliptical distribution, the three-fund rule generally outperforms the scaled GMV portfolio even for $T = 60$ and net of transaction costs. Similarly, we observe that the scaled GMV portfolio calibrated to the multivariate elliptical distribution outperforms the plain sample GMV portfolio in Table 4 in the vast majority of cases for all sample sizes, unlike when it is calibrated to the multivariate normal distribution.

However, irrespective of the calibration method, the three-fund rule outperforms the scaled GMV portfolio when the sample size increases to $T = 120$ and 240. For example, when $T = 120$ and we calibrate the combination coefficients to the multivariate elliptical distribution using the exact finite-sample formula, the three-fund rule increases the annualized gross out-of-sample utility compared with the scaled GMV portfolio by 11.1 percentage points, and the net-of-cost out-of-sample utility by 7.72 percentage points. However, a practical drawback of the three-fund rule is that because it exploits expected returns, it suffers from larger performance losses due to transaction costs. Specifically, when we calibrate combination coefficients to the multivariate elliptical distribution using the exact finite-sample formula, the average loss in out-of-sample utility across datasets suffered by the three-fund rule is 10.5, 6.30, and 4.08 percentage points for $T = 60$, 120, and 240, respectively. For the scaled GMV portfolio, the average loss due to transaction costs decreases to 5.64, 2.91, and 2.24 percentage points for $T = 60$, 120, and 240, respectively.

3. Daily Data

We now test the robustness of our results to using daily instead of monthly data, which are available for the 10MOM, 25SBTM, 25OPINV, and 30IND datasets. Using higher-frequency data has two contrasting impacts on estimation risk. On the one hand, the sample covariance matrix is a more accurate estimate, which reduces estimation risk. On the other hand, fat tails are more pronounced (Martellini and Ziemann, 2010), which increases estimation risk.

Although we use daily data, we reevaluate portfolio weights every month to avoid excessive trading. We use a sample size of either 5, 10, or 20 years, i.e., $T = 1260$, 2520, or 5040 days. For the calibration to the multivariate t-distribution and the multivariate elliptical distribution, we find that because the sample size is large, computing the combination coefficients with the exact finite-sample formula or the asymptotic formula delivers similar results. Therefore, we focus on the asymptotic coefficients that are faster to compute.

We report the results in Table 4 of the supplementary material for the scaled GMV portfolio, the two-fund rule, and the three-fund rule. Overall, the conclusions drawn using monthly data are robust to using daily data. In particular, it is remarkable that calibrating the portfolio rules to the multivariate elliptical distribution systematically delivers the best out-of-sample performance.

VII. Conclusion

Portfolio combination rules that optimize the out-of-sample performance under estimation risk are a popular method in portfolio selection. In their derivations, researchers typically assume that asset returns are multivariate normally distributed; see, e.g., Kan and Zhou (2007), Tu and Zhou (2011), DeMiguel et al. (2015), Kan et al. (2021), Lassance et al. (2024a), and Yuan and Zhou (2023). In this paper, we investigate the impact that fat-tailed multivariate elliptical returns have on the out-of-sample mean and variance of the sample mean-variance portfolio, and we adapt the optimal two-fund rule and three-fund rule of Kan and Zhou (2007) accordingly. Our theoretical results are applicable to any multivariate elliptical distribution. Empirically, we do not need to specify which multivariate elliptical distribution the returns follow, and our two-fund and three-fund rules can be calibrated directly from the data.

We show that fat tails in asset returns lead to poorer out-of-sample performance of the sample mean-variance and sample global minimum-variance portfolios relative to normality. Therefore, in the two-fund and three-fund rules it is optimal to allocate less weights to the sample mean-variance portfolio and the sample GMV portfolio, and more weight to the risk-free asset, when returns are fat-tailed than when returns are multivariate normally distributed. These differences can be significant, and integrating the fat-tailed return distribution in the calibration of portfolio combination rules substantially improves their out-of-sample performance in practice.

Finally, estimation risk is an important concern in other areas of finance in which sample estimators are used to guide forecasts and optimize future outcomes. The multivariate normality assumption is often used in these situations for theoretical tractability, and our results can help researchers employ more realistic distributional assumptions that deliver improved performance. For example, forecast combination is an established method for predicting the equity premium (Rapach, Strauss and Zhou, 2010; Denk and Löffler, 2024). Given the correspondence between forecast combination and portfolio selection (Beck, Kozbur and Wolf, 2023), estimation risk affects both methods in a similar manner, and our results can thus help incorporate the impact of fat tails in the equity premium prediction. Another example is asset pricing, where the stochastic discount factor (SDF) that prices all test assets and spans the efficient frontier depends on the mean and covariance matrix of the test assets, and thus, suffers from parameter uncertainty out of sample (Bryzgalova, Pelger and Zhu, 2023). Our results can help build better SDF models that account for both the impact of fat tails and parameter uncertainty in spanning the out-of-sample efficient frontier.

Figure 1: Parameters η and φ under the Multivariate t-Distribution.

This figure depicts the parameters η and φ when asset returns follow a multivariate t-distribution as a function of ρ for two different degrees of freedom $\nu = (4, 8)$. The two parameters are obtained by using Proposition 6.

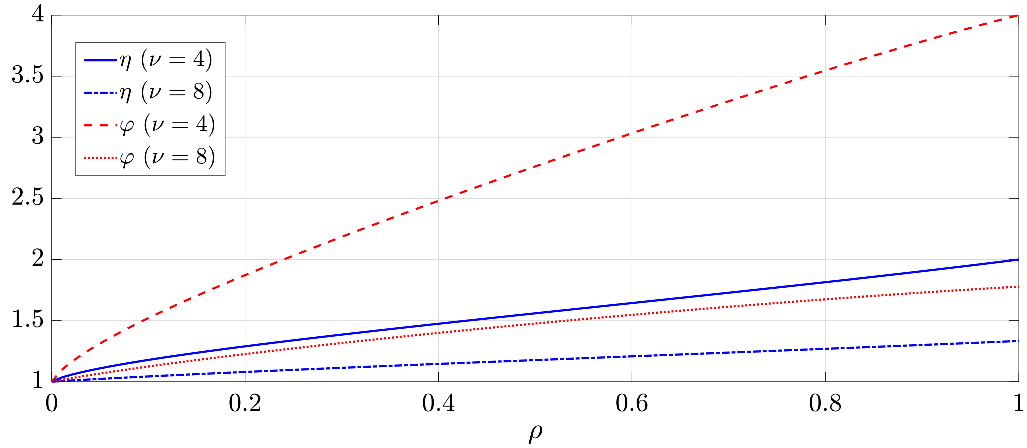


Figure 2: Limit of the Out-of-Sample Utility of the Sample Mean-Variance Portfolio under the Multivariate t-Distribution.

This figure depicts the limit of the out-of-sample utility of the sample mean-variance portfolio, $U(\hat{w})$ in (33), in the fixed N/T asymptotic setting. We set $\gamma = 1$, $\rho = (0.1, 0.3, 0.5)$ $\theta = (0.3, 0.5)$, and we depict the limit of $U(\hat{w})$ as a function of the number of degrees of freedom ν of the multivariate t-distribution.

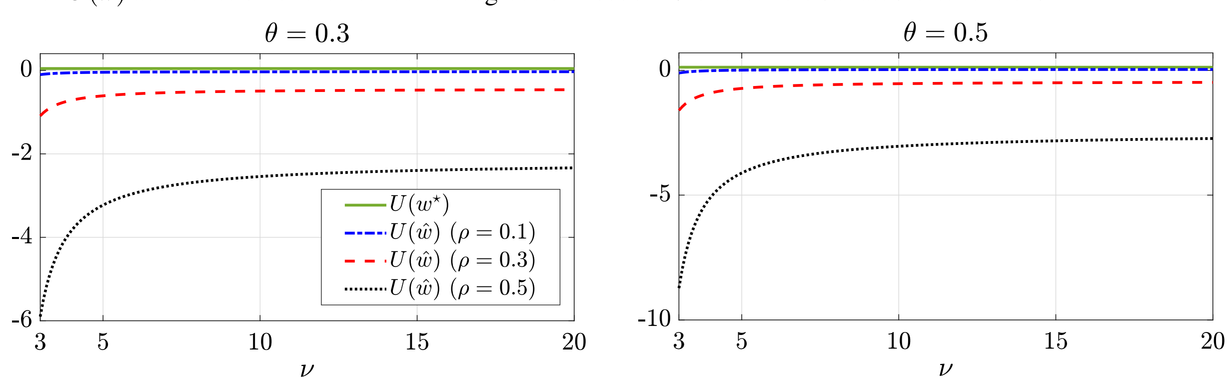


Figure 3: Optimal Three-Fund Combination Coefficients under the Multivariate t-Distribution versus the Normal Distribution.

This figure depicts the ratio between the optimal three-fund combination coefficients $(c_1^\star, c_2^\star)$ under the multivariate t-distribution and those under the multivariate normal distribution. We depict the ratio as a function of ρ for a number of degrees of freedom $\nu = (4, 8)$ and $\psi = (0.2, 0.4)$. The ratios are obtained by using (41)–(42).

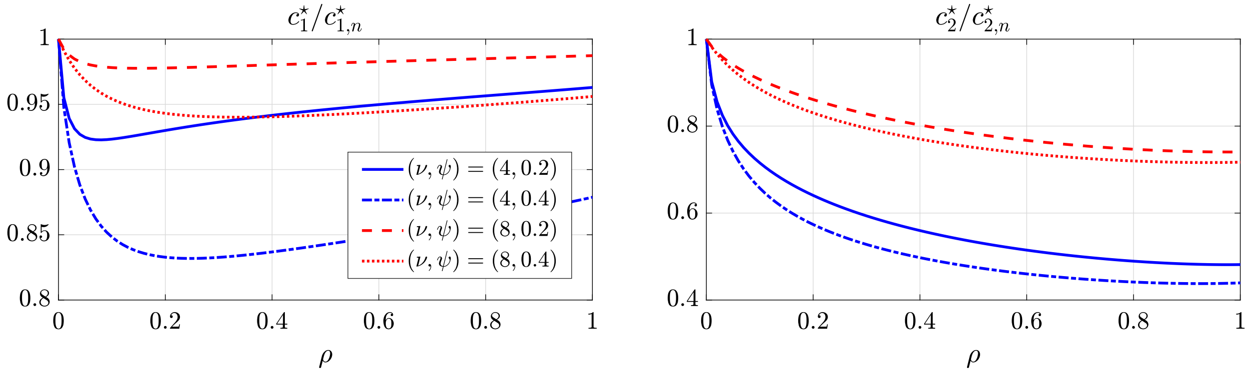


Figure 4: Exact versus Asymptotic (k_1, k_2, k_3) under the Multivariate t-Distribution.

This figure compares the exact finite-sample value of (k_1, k_2, k_3) in Proposition 7 to their asymptotic approximation (η, φ, η). We calibrate these parameters by assuming the returns follow a multivariate t-distribution. We set a sample size $T = 100$, a number of degrees of freedom $\nu = (4, 8)$, and we vary the number of assets N from 2 to 60.

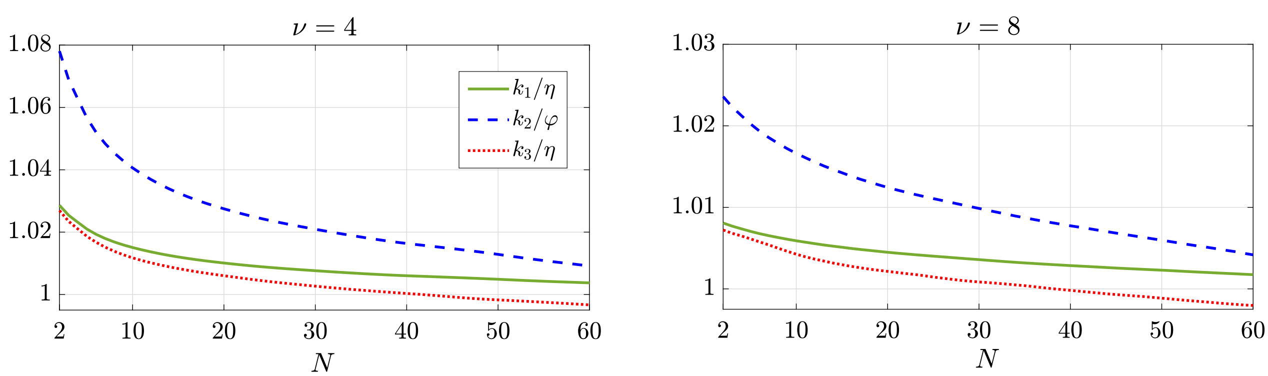


Figure 5: Expected Out-of-Sample Utility Loss Due to the Multivariate Normality Assumption.

This figure depicts the difference between the annualized expected out-of-sample utility of the two-fund and three-fund rules that are wrongly calibrated to the multivariate normal distribution with the annualized utility of the optimal two-fund and three-fund rules. We assume the returns are multivariate t-distributed, $\gamma = 1$, $(\psi, \theta) = (0.2, 0.3)$ or $(0.4, 0.5)$, $T = 100$, $N = (10, 25, 50)$, and the number of degrees of freedom ν is between 3 and 10.

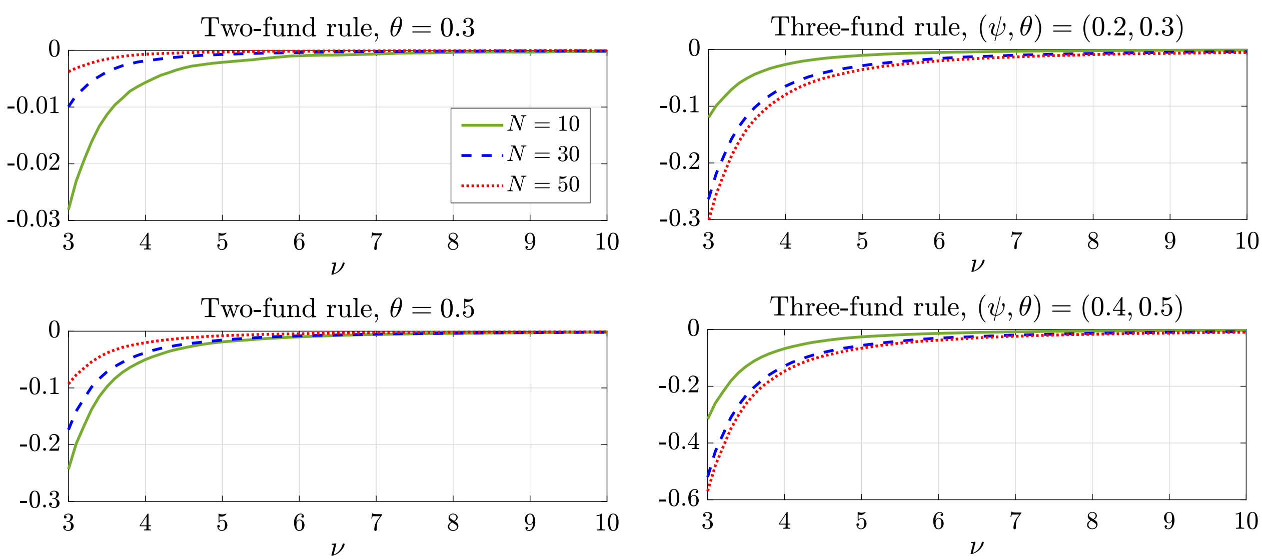


Figure 6: Boxplots of Estimates of the Parameters η and φ.

This figure depicts boxplots of estimates of the parameters η and φ defined in Proposition 3. The estimates $\hat{\eta}$ and $\hat{\varphi}$ in (56)–(57) assume returns are multivariate t-distributed, and the number of degrees of freedom ν is estimated by maximum likelihood. The estimates $\tilde{\eta}$ and $\tilde{\varphi}$ in (58)–(59) use the sample distribution of τ_t by El Karoui (2010, 2013) in (60). The boxplots are obtained by simulating 10,000 return vectors of size T from a multivariate t-distribution with $(\mu, \Sigma) = (0_N, I_N)$, $N = 25$, $\nu = (4, 6, 8)$, and $T = (60, 120, 240)$. The dashed horizontal lines depict the true value of η and φ. The star symbol in each boxplot depicts the average estimated value.

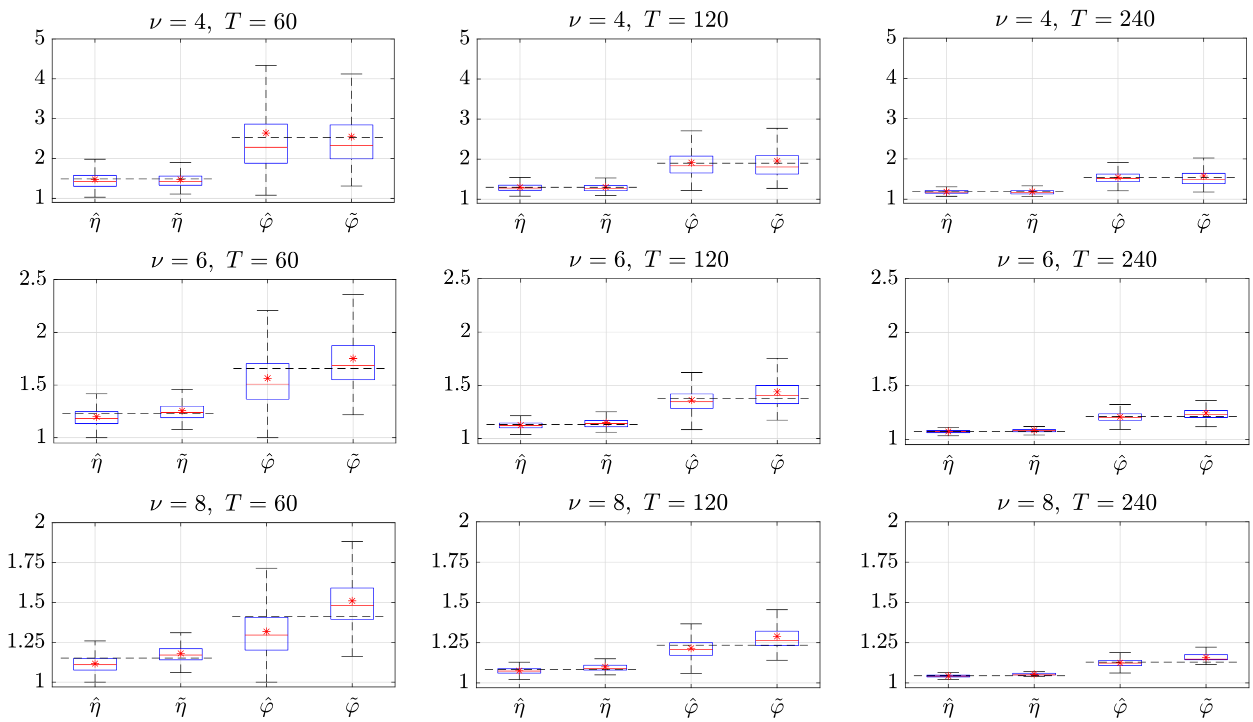


Figure 7: Parameters (η, φ) and (k_1, k_2, k_3) Calibrated to the Multivariate t and Elliptical Distributions.

This figure depicts boxplots of the estimates of the parameters (η, φ) and (k_1, k_2, k_3). These parameters are calibrated either to the multivariate t-distribution whose number of degrees of freedom ν is estimated by maximum likelihood, which corresponds to $(\hat{\eta}, \hat{\varphi})$ and $(\hat{k}_1, \hat{k}_2, \hat{k}_3)$, or to the multivariate elliptical distribution using the sample distribution of τ_t by El Karoui (2010, 2013), which corresponds to $(\tilde{\eta}, \tilde{\varphi})$ and $(\tilde{k}_1, \tilde{k}_2, \tilde{k}_3)$. The boxplots depict the variability of these parameters across all the estimation windows. We consider the dataset of 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to July 2023, and a sample size $T = 60$, 120, and 240 months.

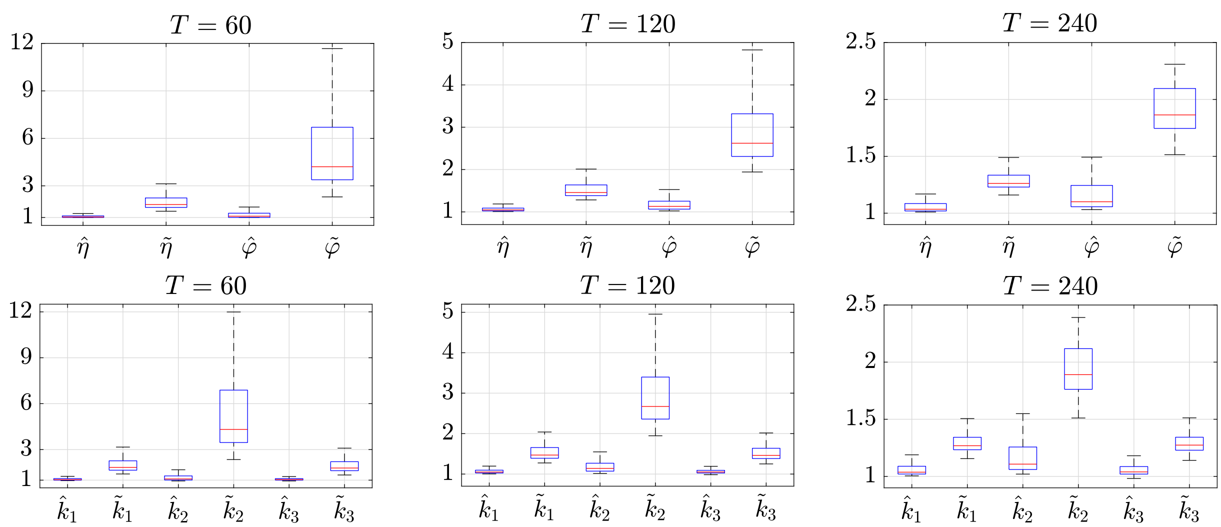


Figure 8: Net-of-Cost Annualized Out-of-Sample Utility of the Two-Fund and Three-Fund Rules.

This figure depicts, across the six datasets described in Section VI.B, the net-of-cost annualized out-of-sample utility delivered by the two-fund and three-fund rules calibrated to 1) the multivariate normal distribution, 2) the multivariate t-distribution whose number of degrees of freedom ν is estimated by maximum likelihood, 3) the multivariate elliptical distribution using the sample distribution of τ_t by El Karoui (2010, 2013). We compute the combination coefficients by using the exact finite-sample formula in Proposition 7, and we use sample sizes $T = 60$, 120, and 240 months. Formulas for the estimated two-fund and three-fund combination coefficients are available in Table 1. The utilities are net of proportional transaction costs of 10 basis points.

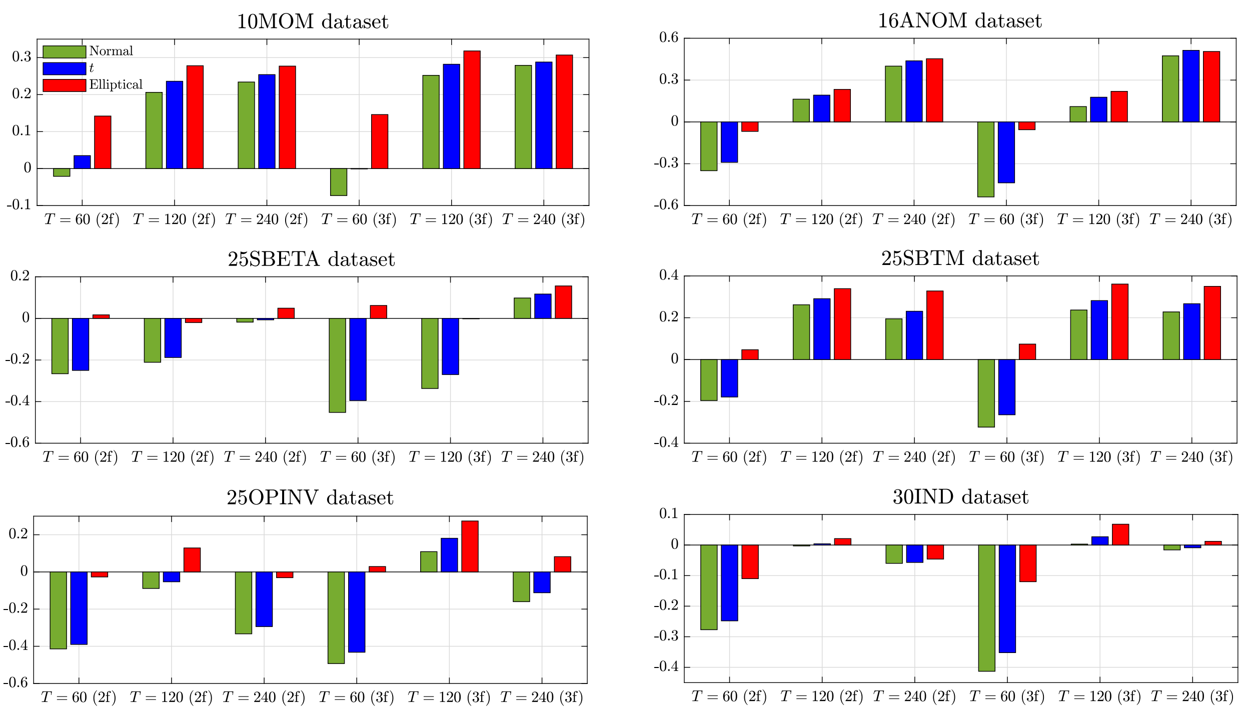


Table 1: Formulas for Estimated Two-Fund and Three-Fund Combination Coefficients

This table reports formulas for the estimators of the optimal combination coefficients in the two-fund rule ($\hat{c}$) and the three-fund rule $(\hat{c}_1, \hat{c}_2)$. The combination coefficients are calibrated either to the multivariate normal, t, or elliptical distribution using the asymptotic formula of the exact finite-sample formula. The different inputs entering the formulas for $\hat{c}$ and $(\hat{c}_1, \hat{c}_2)$ are defined in Section V.

Calibration method	Two-fund coefficient	Three-fund coefficients
Normal (exact)	$\hat{c} = \frac{(T-N-1)(T-N-4)}{T(T-2)} \left(\frac{\hat{\theta}_a^2}{\hat{\theta}_a^2 + \frac{N}{T}} \right)$	$(\hat{c}_1, \hat{c}_2) = \frac{(T-N-1)(T-N-4)}{T(T-2)} \left(\frac{\hat{\psi}_a^2}{\hat{\psi}_a^2 + \frac{N}{T}}, \frac{\hat{\mu}_g \frac{N}{T}}{\hat{\psi}_a^2 + \frac{N}{T}} \right)$
t (asymp)	$\hat{c} = \left(1 - \frac{N}{T}\right)^2 \frac{\hat{\theta}_a^2}{(\hat{\varphi}/\hat{\eta})\hat{\theta}_a^2 + \frac{N}{T}}$	$(\hat{c}_1, \hat{c}_2) = \left(1 - \frac{N}{T}\right)^2 \left(\frac{\hat{\psi}_a^2}{(\hat{\varphi}/\hat{\eta})\hat{\psi}_a^2 + \frac{N}{T}}, \frac{\hat{\mu}_g (\hat{\eta}/\hat{\varphi}) \frac{N}{T}}{(\hat{\varphi}/\hat{\eta})\hat{\psi}_a^2 + \frac{N}{T}} \right)$
t (exact)	$\hat{c} = \frac{(T-N-1)(T-N-4)}{T(T-2)} \left(\frac{\hat{k}_1 \hat{\theta}_a^2}{\hat{k}_2 \hat{\theta}_a^2 + \hat{k}_3 \frac{N}{T}} \right)$	$(\hat{c}_1, \hat{c}_2) = \frac{(T-N-1)(T-N-4)}{T(T-2)} \left(\frac{\hat{k}_1 \hat{\psi}_a^2}{\hat{k}_2 \hat{\psi}_a^2 + \hat{k}_3 \frac{N}{T}}, \frac{\hat{\mu}_g \hat{k}_1 (\hat{k}_3/\hat{k}_2) \frac{N}{T}}{\hat{k}_2 \hat{\psi}_a^2 + \hat{k}_3 \frac{N}{T}} \right)$
Elliptical (asymp)	$\hat{c} = \left(1 - \frac{N}{T}\right)^2 \frac{\hat{\theta}_a^2}{(\tilde{\varphi}/\tilde{\eta})\hat{\theta}_a^2 + \frac{N}{T}}$	$(\hat{c}_1, \hat{c}_2) = \left(1 - \frac{N}{T}\right)^2 \left(\frac{\hat{\psi}_a^2}{(\tilde{\varphi}/\tilde{\eta})\hat{\psi}_a^2 + \frac{N}{T}}, \frac{\hat{\mu}_g (\tilde{\eta}/\tilde{\varphi}) \frac{N}{T}}{(\tilde{\varphi}/\tilde{\eta})\hat{\psi}_a^2 + \frac{N}{T}} \right)$
Elliptical (exact)	$\hat{c} = \frac{(T-N-1)(T-N-4)}{T(T-2)} \left(\frac{\tilde{k}_1 \hat{\theta}_a^2}{\tilde{k}_2 \hat{\theta}_a^2 + \tilde{k}_3 \frac{N}{T}} \right)$	$(\hat{c}_1, \hat{c}_2) = \frac{(T-N-1)(T-N-4)}{T(T-2)} \left(\frac{\tilde{k}_1 \hat{\psi}_a^2}{\tilde{k}_2 \hat{\psi}_a^2 + \tilde{k}_3 \frac{N}{T}}, \frac{\hat{\mu}_g \tilde{k}_1 (\tilde{k}_3/\tilde{k}_2) \frac{N}{T}}{\tilde{k}_2 \hat{\psi}_a^2 + \tilde{k}_3 \frac{N}{T}} \right)$

Table 2: Out-of-Sample Performance with Simulated Multivariate t-Distributed Data.

This table reports the annualized expected out-of-sample utility (EU) and the mean value of the combination coefficients for the two-fund and three-fund rules. We simulate multivariate t-distributed returns with $\nu = 4$, 6, and 8 degrees of freedom and (μ_g, ψ, θ) calibrated to a dataset of 25 portfolios of firms sorted on size and book-to-market spanning July 1926 to July 2023. The risk-aversion coefficient is $\gamma = 1$ and the sample size is $T = 60$, 120, and 240 months. The combination coefficients are calibrated to 1) the multivariate normal distribution, 2) the multivariate t-distribution with ν estimated by maximum likelihood, and 3) the multivariate elliptical distribution using the sample distribution of τ_t by El Karoui (2010, 2013), and 4) a five-fold cross-validation method. For the second and third case, the combination coefficients are computed using the fixed N/T asymptotic formula and the exact finite-sample formula. Formulas for the estimated two-fund and three-fund combination coefficients are available in Table 1. We also report the EU for the case in which the optimal combination coefficients are known. The largest EU for the case with estimated combination coefficients is depicted in bold.

		Known combination coefficients			Estimated combination coefficients					
		Normal (exact)	t (asymp)	t (exact)	Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
A) $\nu = 4$										
Two-fund rule										
$T = 60$	EU	0.071	0.071	0.072	−0.155	−0.075	−0.029	0.004	**0.045**	−0.001
	Mean $\hat{c}$	0.054	0.054	0.049	0.092	0.084	0.075	0.071	0.056	0.035
$T = 120$	EU	0.145	0.148	0.149	0.043	0.092	0.100	0.125	**0.128**	0.081
	Mean $\hat{c}$	0.184	0.167	0.161	0.221	0.194	0.186	0.161	0.152	0.124
$T = 240$	EU	0.232	0.237	0.237	0.183	0.209	0.211	**0.219**	**0.219**	0.190
	Mean $\hat{c}$	0.369	0.329	0.322	0.385	0.339	0.333	0.292	0.285	0.262
Three-fund rule										
$T = 60$	EU	0.076	0.118	0.120	−0.284	−0.099	−0.038	0.013	**0.058**	−0.014
	Mean $\hat{c}_1$	0.039	0.041	0.036	0.082	0.076	0.068	0.065	0.051	0.028
	Mean $\hat{c}_2/\hat{\mu}_g$	0.264	0.160	0.139	0.221	0.134	0.116	0.083	0.053	0.055
$T = 120$	EU	0.170	0.198	0.198	0.011	0.109	0.120	0.152	**0.154**	0.089
	Mean $\hat{c}_1$	0.139	0.130	0.126	0.190	0.170	0.163	0.143	0.135	0.100
	Mean $\hat{c}_2/\hat{\mu}_g$	0.465	0.298	0.282	0.415	0.265	0.250	0.163	0.149	0.152
$T = 240$	EU	0.258	0.273	0.273	0.188	0.231	0.234	**0.243**	**0.243**	0.201
	Mean $\hat{c}_1$	0.296	0.270	0.265	0.318	0.286	0.280	0.251	0.246	0.220
	Mean $\hat{c}_2/\hat{\mu}_g$	0.495	0.348	0.337	0.472	0.334	0.324	0.236	0.228	0.227
B) $\nu = 6$										
Two-fund rule										
$T = 60$	EU	0.063	0.062	0.063	−0.087	−0.081	−0.039	0.016	**0.035**	0.000
	Mean $\hat{c}$	0.054	0.058	0.051	0.087	0.089	0.079	0.069	0.059	0.036
$T = 120$	EU	0.138	0.138	0.138	0.071	0.085	0.091	0.115	**0.117**	0.078
	Mean $\hat{c}$	0.184	0.179	0.172	0.207	0.199	0.192	0.164	0.157	0.127
$T = 240$	EU	0.229	0.230	0.230	0.194	0.202	0.203	**0.211**	**0.211**	0.187
	Mean $\hat{c}$	0.369	0.353	0.347	0.366	0.350	0.343	0.303	0.297	0.274
Three-fund rule										
$T = 60$	EU	0.108	0.117	0.119	−0.159	−0.119	−0.055	0.032	**0.054**	0.000
	Mean $\hat{c}_1$	0.039	0.042	0.038	0.079	0.081	0.072	0.063	0.054	0.028
	Mean $\hat{c}_2/\hat{\mu}_g$	0.264	0.211	0.185	0.224	0.184	0.162	0.086	0.068	0.061
$T = 120$	EU	0.191	0.198	0.198	0.073	0.105	0.116	0.148	**0.150**	0.091
	Mean $\hat{c}_1$	0.139	0.137	0.132	0.179	0.173	0.167	0.145	0.140	0.102
	Mean $\hat{c}_2/\hat{\mu}_g$	0.465	0.377	0.359	0.425	0.347	0.331	0.202	0.190	0.160
$T = 240$	EU	0.271	0.275	0.275	0.218	0.231	0.233	**0.242**	**0.242**	0.205
	Mean $\hat{c}_1$	0.296	0.287	0.282	0.303	0.292	0.287	0.257	0.253	0.223
	Mean $\hat{c}_2/\hat{\mu}_g$	0.495	0.424	0.413	0.488	0.421	0.411	0.293	0.285	0.278

Table 2: Out-of-Sample Performance with Simulated Multivariate t-Distributed Data (continued).

		Known combination coefficients			Estimated combination coefficients					
		Normal (exact)	t (asymp)	t (exact)	Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
C) $\nu = 8$										
Two-fund rule										
$T = 60$	EU	0.060	0.060	0.060	−0.069	−0.084	−0.042	0.017	**0.032**	−0.004
	Mean $\hat{c}$	0.054	0.059	0.053	0.086	0.092	0.082	0.069	0.061	0.037
$T = 120$	EU	0.135	0.135	0.135	0.071	0.085	0.092	0.116	**0.118**	0.076
	Mean $\hat{c}$	0.184	0.183	0.176	0.209	0.200	0.193	0.165	0.158	0.129
$T = 240$	EU	0.229	0.229	0.229	0.197	0.201	0.202	**0.210**	**0.210**	0.186
	Mean $\hat{c}$	0.369	0.361	0.355	0.362	0.354	0.348	0.308	0.302	0.279
Three-fund rule										
$T = 60$	EU	0.115	0.118	0.120	−0.124	−0.126	−0.062	0.035	**0.052**	−0.010
	Mean $\hat{c}_1$	0.039	0.043	0.038	0.077	0.083	0.074	0.063	0.056	0.029
	Mean $\hat{c}_2/\hat{\mu}_g$	0.264	0.234	0.208	0.226	0.208	0.184	0.090	0.076	0.060
$T = 120$	EU	0.196	0.199	0.199	0.092	0.107	0.117	0.147	**0.149**	0.089
	Mean $\hat{c}_1$	0.139	0.140	0.135	0.175	0.174	0.168	0.147	0.141	0.098
	Mean $\hat{c}_2/\hat{\mu}_g$	0.465	0.410	0.392	0.429	0.381	0.365	0.220	0.208	0.178
$T = 240$	EU	0.274	0.275	0.276	0.224	0.231	0.233	**0.242**	**0.242**	0.201
	Mean $\hat{c}_1$	0.296	0.292	0.287	0.300	0.295	0.290	0.262	0.257	0.227
	Mean $\hat{c}_2/\hat{\mu}_g$	0.495	0.451	0.442	0.490	0.449	0.440	0.315	0.306	0.286

Table 3: Out-of-Sample Performance of Two-Fund and Three-Fund Rules with Empirical Data.

This table reports the annualized out-of-sample utility (EU) and the mean value of the combination coefficients for the two-fund and three-fund rules, across the six datasets described in Section VI.B. The EU is either gross or net of proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$ and the sample size is $T = 60$, 120, and 240 months. The combination coefficients are calibrated to 1) the multivariate normal distribution, 2) the multivariate t-distribution with ν estimated by maximum likelihood, 3) the multivariate elliptical distribution using the sample distribution of τ_t by El Karoui (2010, 2013), and 4) a five-fold cross-validation method. For the second and third case, the combination coefficients are computed using the fixed N/T asymptotic formula and the exact finite-sample formula. Formulas for the estimated two-fund and three-fund combination coefficients are available in Table 1. The largest EU in each case is depicted in bold. For each dataset, we report the number of degrees of freedom estimated by maximum likelihood, $\hat{\nu}$.

		Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
A) 10MOM dataset ($\hat{\nu} = 4.15$)							
Two-fund rule							
$T = 60$	Gross EU	0.052	0.057	0.105	0.186	**0.202**	0.139
	Net EU	−0.021	−0.017	0.035	0.122	**0.142**	0.061
	Mean $\hat{c}$	0.284	0.288	0.267	0.233	0.215	0.155
$T = 120$	Gross EU	0.254	0.270	0.282	0.313	**0.317**	0.267
	Net EU	0.206	0.223	0.236	0.272	**0.278**	0.217
	Mean $\hat{c}$	0.445	0.435	0.420	0.375	0.360	0.337
$T = 240$	Gross EU	0.267	0.279	0.285	0.302	**0.306**	0.289
	Net EU	0.234	0.248	0.254	0.273	**0.277**	0.256
	Mean $\hat{c}$	0.626	0.606	0.594	0.548	0.536	0.489
Three-fund rule							
$T = 60$	Gross EU	0.003	0.008	0.070	0.185	**0.206**	0.137
	Net EU	−0.073	−0.066	−0.001	0.121	**0.146**	0.050
	Mean $\hat{c}_1$	0.242	0.246	0.229	0.202	0.187	0.124
	Mean $\hat{c}_2/\hat{\mu}_g$	0.406	0.363	0.331	0.217	0.193	0.143
$T = 120$	Gross EU	0.298	0.315	0.326	0.354	**0.356**	0.280
	Net EU	0.252	0.270	0.282	0.315	**0.318**	0.227
	Mean $\hat{c}_1$	0.361	0.357	0.345	0.311	0.299	0.281
	Mean $\hat{c}_2/\hat{\mu}_g$	0.455	0.398	0.379	0.288	0.271	0.181
$T = 240$	Gross EU	0.309	0.310	0.317	0.329	**0.333**	0.311
	Net EU	0.279	0.281	0.288	0.302	**0.307**	0.279
	Mean $\hat{c}_1$	0.498	0.489	0.480	0.447	0.438	0.402
	Mean $\hat{c}_2/\hat{\mu}_g$	0.408	0.355	0.344	0.292	0.282	0.211

Table 3: Out-of-Sample Performance of Two-Fund and Three-Fund Rules with Empirical Data (continued).

		Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
B) 16ANOM dataset ($\hat{\nu} = 5.46$)							
Two-fund rule							
$T = 60$	Gross EU	−0.198	−0.228	−0.145	0.012	0.043	**0.108**
	Net EU	−0.350	−0.380	−0.290	−0.108	−0.068	**−0.021**
	Mean $\hat{c}$	0.203	0.215	0.197	0.169	0.150	0.098
$T = 120$	Gross EU	0.242	0.258	0.267	0.291	**0.294**	0.213
	Net EU	0.163	0.181	0.192	0.227	**0.233**	0.150
	Mean $\hat{c}$	0.402	0.348	0.335	0.317	0.304	0.180
$T = 240$	Gross EU	0.458	0.488	0.493	**0.500**	**0.500**	0.406
	Net EU	0.400	0.433	0.438	0.451	**0.453**	0.352
	Mean $\hat{c}$	0.618	0.575	0.564	0.500	0.489	0.449
Three-fund rule							
$T = 60$	Gross EU	−0.385	−0.428	−0.292	0.010	0.057	**0.108**
	Net EU	−0.538	−0.577	−0.437	−0.111	−0.056	**−0.031**
	Mean $\hat{c}_1$	0.146	0.173	0.159	0.129	0.118	0.067
	Mean $\hat{c}_2/\hat{\mu}_g$	0.348	0.305	0.277	0.174	0.154	0.091
$T = 120$	Gross EU	0.189	0.235	0.252	0.272	**0.279**	0.155
	Net EU	0.110	0.158	0.177	0.209	**0.219**	0.087
	Mean $\hat{c}_1$	0.285	0.288	0.278	0.246	0.238	0.139
	Mean $\hat{c}_2/\hat{\mu}_g$	0.442	0.376	0.359	0.268	0.254	0.105
$T = 240$	Gross EU	0.528	0.560	**0.565**	0.549	0.550	0.426
	Net EU	0.474	0.507	**0.513**	0.503	0.505	0.366
	Mean $\hat{c}_1$	0.552	0.486	0.477	0.461	0.452	0.400
	Mean $\hat{c}_2/\hat{\mu}_g$	0.307	0.306	0.298	0.201	0.194	0.075
C) 25SBETA dataset ($\hat{\nu} = 7.92$)							
Two-fund rule							
$T = 60$	Gross EU	−0.073	−0.153	−0.060	0.125	**0.140**	0.039
	Net EU	−0.266	−0.358	−0.250	−0.011	**0.017**	−0.119
	Mean $\hat{c}$	0.107	0.117	0.104	0.074	0.066	0.043
$T = 120$	Gross EU	−0.105	−0.108	−0.085	0.052	**0.061**	0.021
	Net EU	−0.211	−0.213	−0.188	−0.032	**−0.020**	−0.055
	Mean $\hat{c}$	0.232	0.232	0.223	0.177	0.170	0.089
$T = 240$	Gross EU	0.045	0.050	0.054	0.098	0.100	**0.118**
	Net EU	−0.018	−0.012	−0.007	0.046	0.049	**0.069**
	Mean $\hat{c}$	0.355	0.350	0.344	0.291	0.286	0.180
Three-fund rule							
$T = 60$	Gross EU	−0.235	−0.367	−0.183	0.175	**0.191**	0.035
	Net EU	−0.452	−0.589	−0.395	0.032	**0.062**	−0.153
	Mean $\hat{c}_1$	0.082	0.090	0.081	0.061	0.055	0.028
	Mean $\hat{c}_2/\hat{\mu}_g$	0.221	0.225	0.199	0.079	0.068	0.059
$T = 120$	Gross EU	−0.234	−0.212	−0.171	0.066	**0.077**	−0.054
	Net EU	−0.337	−0.313	−0.270	−0.016	**−0.002**	−0.139
	Mean $\hat{c}_1$	0.175	0.176	0.170	0.140	0.135	0.064
	Mean $\hat{c}_2/\hat{\mu}_g$	0.429	0.390	0.373	0.200	0.189	0.094
$T = 240$	Gross EU	0.156	0.169	0.173	0.201	**0.202**	0.150
	Net EU	0.098	0.112	0.117	0.155	**0.156**	0.101
	Mean $\hat{c}_1$	0.259	0.256	0.252	0.218	0.214	0.143
	Mean $\hat{c}_2/\hat{\mu}_g$	0.532	0.491	0.481	0.325	0.316	0.165

Table 3: Out-of-Sample Performance of Two-Fund and Three-Fund Rules with Empirical Data (continued).

		Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
D) 25SBTM dataset ($\hat{v} = 4.03$)							
Two-fund rule							
$T = 60$	Gross EU	−0.044	−0.130	−0.030	0.143	**0.157**	0.097
	Net EU	−0.196	−0.283	−0.179	0.022	**0.047**	−0.046
	Mean $\hat{c}$	0.103	0.113	0.101	0.075	0.066	0.042
$T = 120$	Gross EU	0.366	0.378	0.392	**0.421**	0.418	0.379
	Net EU	0.262	0.274	0.291	0.338	**0.339**	0.280
	Mean $\hat{c}$	0.262	0.259	0.249	0.193	0.185	0.154
$T = 240$	Gross EU	0.259	0.285	0.294	0.381	0.384	**0.392**
	Net EU	0.195	0.222	0.231	0.323	**0.328**	0.327
	Mean $\hat{c}$	0.440	0.423	0.415	0.343	0.336	0.282
Three-fund rule							
$T = 60$	Gross EU	−0.149	−0.247	−0.097	0.176	**0.191**	0.100
	Net EU	−0.323	−0.415	−0.264	0.048	**0.074**	−0.075
	Mean $\hat{c}_1$	0.085	0.093	0.083	0.065	0.058	0.031
	Mean $\hat{c}_2/\hat{\mu}_g$	0.218	0.220	0.195	0.075	0.065	0.052
$T = 120$	Gross EU	0.342	0.366	0.384	**0.441**	0.439	0.379
	Net EU	0.237	0.262	0.282	**0.361**	**0.361**	0.269
	Mean $\hat{c}_1$	0.205	0.205	0.197	0.160	0.154	0.126
	Mean $\hat{c}_2/\hat{\mu}_g$	0.399	0.354	0.339	0.176	0.166	0.089
$T = 240$	Gross EU	0.290	0.316	0.328	0.400	**0.404**	0.394
	Net EU	0.228	0.255	0.267	0.345	**0.350**	0.325
	Mean $\hat{c}_1$	0.356	0.347	0.341	0.292	0.286	0.235
	Mean $\hat{c}_2/\hat{\mu}_g$	0.434	0.374	0.366	0.231	0.224	0.169
E) 25OPINV dataset ($\hat{v} = 7.35$)							
Two-fund rule							
$T = 60$	Gross EU	−0.248	−0.375	−0.227	0.070	**0.099**	**0.099**
	Net EU	−0.414	−0.539	−0.390	−0.068	**−0.027**	−0.058
	Mean $\hat{c}$	0.103	0.113	0.101	0.073	0.065	0.047
$T = 120$	Gross EU	0.030	0.037	0.063	0.211	**0.217**	0.196
	Net EU	−0.089	−0.081	−0.053	0.119	**0.129**	0.087
	Mean $\hat{c}$	0.261	0.260	0.250	0.186	0.179	0.143
$T = 240$	Gross EU	−0.271	−0.253	−0.233	0.016	0.028	**0.111**
	Net EU	−0.333	−0.314	−0.294	−0.044	−0.031	**0.052**
	Mean $\hat{c}$	0.435	0.426	0.418	0.342	0.336	0.233
Three-fund rule							
$T = 60$	Gross EU	−0.300	−0.445	−0.245	0.130	**0.161**	0.088
	Net EU	−0.493	−0.628	−0.432	−0.014	**0.029**	−0.096
	Mean $\hat{c}_1$	0.082	0.091	0.081	0.062	0.056	0.034
	Mean $\hat{c}_2/\hat{\mu}_g$	0.220	0.228	0.202	0.076	0.066	0.057
$T = 120$	Gross EU	0.231	0.274	0.300	**0.362**	0.361	0.303
	Net EU	0.109	0.153	0.181	0.272	**0.274**	0.183
	Mean $\hat{c}_1$	0.200	0.202	0.195	0.151	0.146	0.108
	Mean $\hat{c}_2/\hat{\mu}_g$	0.404	0.370	0.354	0.184	0.174	0.141
$T = 240$	Gross EU	−0.097	−0.071	−0.051	0.132	0.138	**0.218**
	Net EU	−0.160	−0.132	−0.112	0.076	0.082	**0.151**
	Mean $\hat{c}_1$	0.313	0.310	0.305	0.256	0.251	0.158
	Mean $\hat{c}_2/\hat{\mu}_g$	0.478	0.436	0.427	0.282	0.274	0.303

Table 3: Out-of-Sample Performance of Two-Fund and Three-Fund Rules with Empirical Data (continued).

		Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
F) 30IND dataset ($\hat{\nu} = 5.42$)							
Two-fund rule							
$T = 60$	Gross EU	−0.197	−0.253	−0.168	−0.085	−0.043	**−0.007**
	Net EU	−0.277	−0.342	−0.248	−0.159	−0.110	**−0.063**
	Mean $\hat{c}$	0.052	0.059	0.051	0.049	0.042	0.015
$T = 120$	Gross EU	0.035	0.033	0.040	0.050	**0.053**	0.048
	Net EU	−0.003	−0.005	0.004	0.016	**0.021**	0.017
	Mean $\hat{c}$	0.119	0.122	0.117	0.110	0.105	0.050
$T = 240$	Gross EU	−0.042	−0.041	−0.039	−0.032	−0.029	**−0.010**
	Net EU	−0.060	−0.060	−0.057	−0.049	−0.046	**−0.021**
	Mean $\hat{c}$	0.163	0.163	0.160	0.152	0.150	0.047
Three-fund rule							
$T = 60$	Gross EU	−0.307	−0.381	−0.248	−0.094	−0.042	**0.043**
	Net EU	−0.413	−0.496	−0.352	−0.181	−0.120	**−0.034**
	Mean $\hat{c}_1$	0.047	0.053	0.046	0.044	0.039	0.011
	Mean $\hat{c}_2/\hat{\mu}_g$	0.170	0.180	0.155	0.094	0.081	0.039
$T = 120$	Gross EU	0.047	0.056	0.068	0.097	**0.103**	0.067
	Net EU	0.003	0.013	0.027	0.060	**0.068**	0.027
	Mean $\hat{c}_1$	0.097	0.099	0.095	0.091	0.087	0.030
	Mean $\hat{c}_2/\hat{\mu}_g$	0.444	0.405	0.387	0.269	0.256	0.143
$T = 240$	Gross EU	0.004	0.007	0.011	0.026	**0.029**	0.019
	Net EU	−0.016	−0.013	−0.009	0.009	**0.012**	0.002
	Mean $\hat{c}_1$	0.104	0.105	0.103	0.100	0.099	0.012
	Mean $\hat{c}_2/\hat{\mu}_g$	0.649	0.593	0.582	0.460	0.450	0.362

Table 4: Out-of-Sample Performance of Additional Benchmark Portfolios with Empirical Data.

This table reports the annualized out-of-sample utility (EU) for the equally weighted portfolio (EW), the optimal combination of the equally weighted portfolio with the risk-free asset in (68) (EWRF), the sample global minimum-variance portfolio in (69) (SGMV), and the sample mean-variance portfolio in (15) (SMV), across the six datasets described in Section VI.B. The EU is either gross or net of proportional transaction costs of 10 basis points. The risk-aversion coefficient is $\gamma = 1$ and the sample size is $T = 60$, 120, and 240 months.

		$T = 60$		$T = 120$		$T = 240$	
		Gross EU	Net EU	Gross EU	Net EU	Gross EU	Net EU
10MOM	EW	0.074	0.073	0.065	0.065	0.065	0.065
	EWRF	−0.067	−0.071	0.067	0.065	0.034	0.032
	SGMV	0.089	0.082	0.080	0.077	0.086	0.084
	SMV	−2.770	−2.812	−0.389	−0.466	0.040	−0.003
16ANOM	EW	0.045	0.045	0.059	0.059	0.058	0.057
	EWRF	−0.214	−0.218	−0.027	−0.029	0.046	0.045
	SGMV	0.075	0.062	0.079	0.073	0.106	0.103
	SMV	−8.548	−8.021	−1.356	−1.481	0.010	−0.070
25SBETA	EW	0.068	0.067	0.083	0.083	0.080	0.079
	EWRF	−0.156	−0.160	0.025	0.023	0.106	0.104
	SGMV	0.068	0.049	0.059	0.051	0.075	0.071
	SMV	−26.56	−21.11	−4.537	−4.511	−0.995	−1.105
25SBTM	EW	0.086	0.086	0.085	0.084	0.078	0.077
	EWRF	−0.038	−0.043	0.087	0.085	0.064	0.062
	SGMV	0.087	0.064	0.090	0.080	0.097	0.092
	SMV	−25.64	−19.62	−3.099	−3.091	−0.771	−0.846
25OPINV	EW	0.064	0.063	0.079	0.078	0.075	0.074
	EWRF	−0.157	−0.161	0.007	0.005	0.088	0.087
	SGMV	0.080	0.057	0.099	0.088	0.105	0.100
	SMV	−33.03	−24.11	−4.702	−4.553	−2.368	−2.362
30IND	EW	0.076	0.075	0.076	0.076	0.071	0.071
	EWRF	−0.040	−0.045	0.101	0.099	0.049	0.047
	SGMV	0.048	0.033	0.053	0.047	0.057	0.054
	SMV	−52.81	−42.79	−4.700	−4.699	−1.469	−1.533

References

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

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

Beck, E.; D. Kozbur; and M. Wolf. "Hedging forecast combinations with an application to the random forest." *arXiv working paper*, (2023).

Bentler, P. M., and M. Berkane. "Greatest lower bound to the elliptical theory kurtosis parameter." *Biometrika*, 73 (1986), 240–241.

Bryzgalova, S.; M. Pelger; and J. Zhu. "Forest through the trees: Building cross-sections of stock returns." *The Journal of Finance*, (forthcoming) (2023).

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

Chopra, V. K., and W. T. Ziemba. "The effect of errors in means, variances, and covariances on optimal portfolio choice." *The Journal of Portfolio Management*, 19 (1993), 6–11.

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

DeMiguel, V.; A. Martín-Utrera; and F. Nogales. "Size matters: Optimal calibration of shrinkage estimators for portfolio selection." *Journal of Banking and Finance*, 37 (2013), 3018–3034.

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

Denk, S., and G. Löffler. "Predicting the equity premium with combination forecasts: A reappraisal." *The Review of Asset Pricing Studies*, 14 (2024), 545–577.

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

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

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

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

Jagannathan, R., and T. Ma. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *The Journal of Finance*, 58 (2003), 1651–1683.

Jobson, J., and B. Korkie. "Estimation for Markowitz efficient portfolios." *Journal of the American Statistical Association*, 75 (1980), 544–554.

Jorion, P. "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis*, 21 (1986), 279–292.

Kan, R., and D. Smith. "The distribution of the sample minimum-variance frontier." *Management Science*, 54 (2008), 1364–1380.

Kan, R., and X. Wang. "Optimal portfolio choice with unknown benchmark efficiency." *Management Science*, 70 (2023), 6117–6138.

Kan, R.; X. Wang; and X. Zheng. "In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models." *Journal of Financial Economics*, 155 (2024).

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

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

Lassance, N.; A. Martín-Utrera; and M. Simaan. “The risk of expected utility under parameter uncertainty.” *Management Science*, 70 (2024a), 7644–7663.

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

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

Martellini, L., and V. Ziemann. “Improved estimates of higher-order comoments and implications for portfolio selection.” *The Review of Financial Studies*, 23 (2010), 1467–1502.

Michaud, R. “The Markowitz optimization enigma: Is ‘optimized’ optimal?” *Financial Analysts Journal*, 45 (1989), 31–42.

Novy-Marx, R., and M. Velikov. “A taxonomy of anomalies and their trading costs.” *The Review of Financial Studies*, 29 (2016), 104–147.

Okhrin, Y., and W. Schmid. “Distributional properties of portfolio weights.” *Journal of Econometrics*, 134 (2006), 235–256.

Rapach, D.; J. Strauss; and G. Zhou. “Out-of-sample equity premium prediction: Combination forecasts and links to the real economy.” *The Review of Financial Studies*, 23 (2010), 821–862.

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

Tu, J., and G. Zhou. "Data-generating process uncertainty: What difference does it make in portfolio decisions?" *Journal of Financial Economics*, 72 (2004), 385–421.

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

Yuan, M., and G. Zhou. "Why naive 1/N diversification is not so naive, and how to beat it?" *Journal of Financial and Quantitative Analysis*, (forthcoming) (2023).

Supplementary Material to

Optimal Portfolio Choice with Fat Tails and Parameter Uncertainty

Raymond Kan and Nathan Lassance

This supplementary material to the main paper contains four sections. Section I illustrates the estimation accuracy of the number of degrees of freedom ν and the distribution of τ_t. Section II studies the impact of fat tails on the accuracy of the adjusted estimators of θ^2 and ψ^2. Section III reports the tables containing the results for the additional empirical tests. Section IV contains the proofs of all theoretical results in the main body of the paper.

I. Accuracy of Estimators of ν and τ_t

In this section, we illustrate the estimation accuracy of the number of degrees of freedom ν and the distribution of τ_t, which underlie the two calibration methods proposed in Section V in the main body of the paper to estimate the optimal two-fund and three-fund combination coefficients.

Figure 1 studies the estimation accuracy for ν, which we estimate by maximum likelihood. We set $N = 25$, a population value of $\nu = (4, 6, 8)$, and we depict boxplots of $\hat{\nu}$ across 10,000 simulations of multivariate t-distributed returns for a sample size $T = (60, 120, 240)$. We set $(\mu, \Sigma) = (0_N, I_N)$ without loss of generality because ν does not depend on (μ, Σ). Figure 1 shows that as ν increases, and thus the returns are closer to multivariate normal, it becomes more difficult to estimate ν. Specifically, the boxplots get wider as ν increases. However, in comparison to the volatility of $\hat{\nu}$, the bias of $\hat{\nu}$ is more reasonable, and close to zero for $T = 120$ and 240.

Figure 2 illustrates how the sample distribution of τ_t by El Karoui (2010, 2013), $\hat{\tau}_t$ in (60), converges to the true distribution of τ_t as N increases. We assume returns are multivariate t-distributed, in which case we can show that the exact density function of τ_t in (2) is

$$f_{\tau_t}(x) = \frac{(\nu/2-1)^{\frac{\nu}{2}} e^{-\frac{\nu-2}{2x}}}{\Gamma(\nu/2) x^{\frac{\nu+2}{2}}}. \tag{A1}$$

Figure 1: Boxplots of Estimates of the Number of Degrees of Freedom ν.

This figure depicts boxplots of maximum-likelihood estimates of the number of degrees of freedom ν of the multivariate t-distribution. The boxplots are obtained by simulating 10,000 times T return vectors from a multivariate t-distribution with $(\mu, \Sigma) = (0_N, I_N)$, $N = 25$, $\nu = (4, 6, 8)$, and $T = (60, 120, 240)$. The dotted horizontal lines depict the true value of ν.

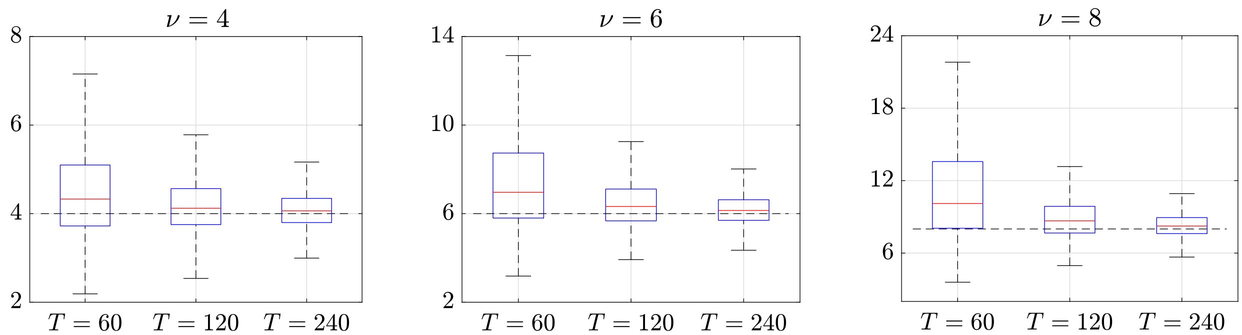
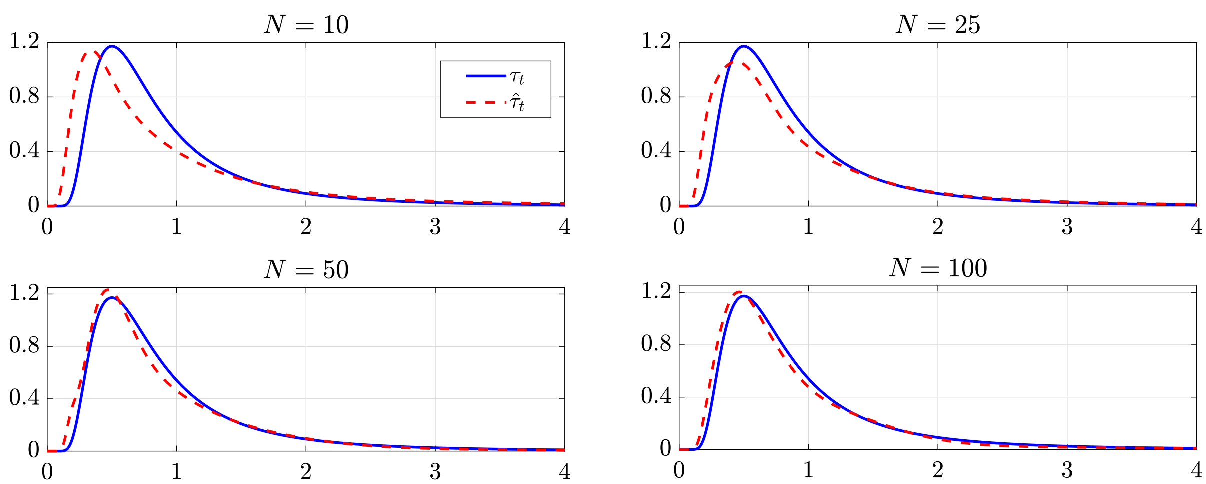


Figure 2: Comparison of the True Distribution of τ_t with the Sample Distribution.

This figure compares the sample distribution of τ_t by El Karoui (2010, 2013), $\hat{\tau}_t$ in (60), with the true distribution of τ_t. We assume returns are multivariate t-distributed, in which case the exact density function of τ_t is given by (A1). The density function of $\hat{\tau}_t$ is found using a kernel density estimator. We set $T = 120$, $\nu = 6$, and an increasing number of assets N that goes from 10 to 100.



Then, we set $T = 120$, $\nu = 6$, and we compare the true density function of τ_t in (A1) with that of $\hat{\tau}_t$ found using a kernel density estimator. We do the comparison for an increasing number of assets N that goes from 10 to 100. Figure 2 shows indeed that the sample distribution and the true distribution get closer to one another as N increases even for a finite T. Moreover, the sample distribution is reasonably accurate even for rather small values of N such as $N = 25$.

II. Impact of Fat Tails on Adjusted Estimators of θ^2 and ψ^2

In the main body of the paper, we estimate θ^2 and ψ^2 via their adjusted estimators θ_a^2 and ψ_a^2 in (54)–(55). These estimators are proposed by Kan and Zhou (2007) and are designed to have minimum root mean square error (RMSE). Unlike the unbiased estimators $\hat{\theta}_{unb}^2$ and $\hat{\psi}_{unb}^2$, given by the first term in (54)–(55), the adjusted estimators are non-negative. Moreover, the adjusted estimators deliver a lower RMSE than the trimmed estimators $\max(\hat{\theta}_{unb}^2, 0)$ and $\max(\hat{\psi}_{unb}^2, 0)$.

However, the adjusted estimators are derived under the multivariate normal distributional assumption, whereas we assume that returns are multivariate elliptical. Therefore, it is of interest to study how fat tails impact the RMSE of $\hat{\theta}_a^2$ and $\hat{\psi}_a^2$. For that purpose, we conduct the following simulation. We simulate $M = 1{,}000{,}000$ times T returns from a multivariate t-distribution with ν degrees of freedom and (μ, Σ) calibrated to a dataset of $N = 25$ portfolios of firms sorted on size and book-to-market spanning July 1926 to July 2023. This choice of (μ, Σ) yields $\theta = 0.302$ and $\psi = 0.250$ in the population. For each simulation $m = 1, \ldots, M$, we obtain adjusted estimates $\hat{\theta}_{a,m}^2$ and $\hat{\psi}_{a,m}^2$, and compute the RMSE as

$$\text{(A2)} \qquad \text{RMSE}(\hat{\theta}_a^2) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\theta}_{a,m}^2 - \theta^2)^2} \quad \text{and} \quad \text{RMSE}(\hat{\psi}_a^2) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\psi}_{a,m}^2 - \psi^2)^2}.$$

Figure 3: Root Mean Squared Error of θ_a^2 and ψ_a^2

This figure depicts the root mean squared error (RMSE) of the adjusted estimators $\hat{\theta}_a^2$ and $\hat{\psi}_a^2$ when the asset returns are multivariate t-distributed with ν degrees of freedom, where ν varies between 4 and 20. We consider a sample size $T = 60$, 120, and 240 months. We calibrate the population value of θ and ψ to a dataset of $N = 25$ portfolios of firms sorted on size and book-to-market spanning July 1926 to July 2023, which yields $\theta = 0.302$ and $\psi = 0.250$. The RMSE is obtained over one million simulations.

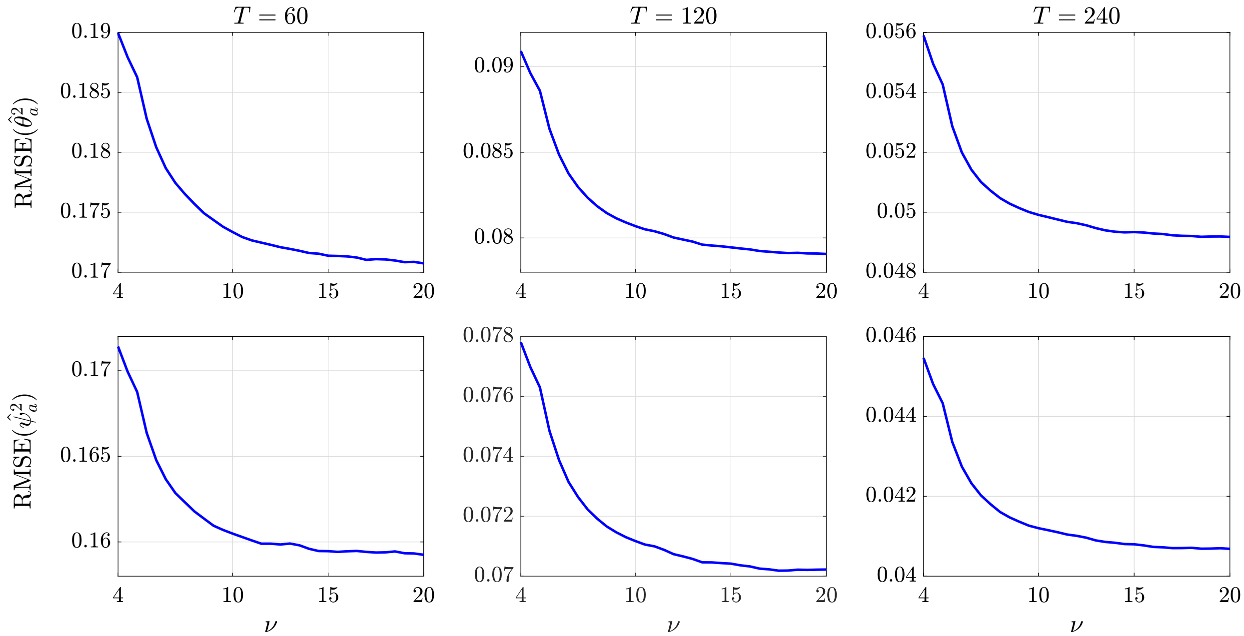


Figure 3 depicts RMSE($\hat{\theta}_a^2$) and RMSE($\hat{\psi}_a^2$) for ν varying between 4 and 20 and $T = 60$, 120, and 240 months. Figure 3 shows that the RMSE does not increase much as ν gets smaller and tails get fatter. For example, when going from $\nu = 20$ (close to normal) to $\nu = 6$ (excess kurtosis of three), RMSE($\hat{\theta}_a^2$) goes from 0.171 to 0.180 for $T = 60$, 0.078 to 0.084 for $T = 120$, and 0.049 to 0.052 for $T = 240$, respectively. The conclusion is similar for RMSE($\hat{\psi}_a^2$). Moreover, Figure 3 shows that the RMSE of $\hat{\theta}_a^2$ and $\hat{\psi}_a^2$ is particularly large when $T = 60$ months, which worsens the estimation accuracy for the two-fund and three-fund combination coefficients. This partly explains why, in Figure 3 of the main body of the manuscript, the empirical performance is generally much worse (and sometimes negative) when $T = 60$ months relative to $T = 120$ and 240 months.

III. Tables for the Additional Empirical Tests

In this section, we report tables containing additional empirical results. Specifically, Table 1 reports the skewness and excess kurtosis of the two-fund and three-fund rules, Table 2 reports the in-sample versus out-of-sample performance of the two-fund and three-fund rules discussed in Section VI.C.1 in the main body of the paper, Table 3 reports the results for the combination of the sample GMV portfolio with the risk-free asset discussed in Section VI.C.2, and Table 4 reports the results for daily data discussed in Section VI.C.3.

Table 1: Skewness and Excess Kurtosis of Two-Fund and Three-Fund Rules.

This table reports the monthly skewness and excess kurtosis of the net-of-cost out-of-sample returns of the two-fund and three-fund rules across the six datasets described in Section VI.B in the main body of the paper. The combination coefficients are calibrated either to the multivariate normal distribution or to the multivariate elliptical distribution using the exact finite-sample formula. Formulas for the estimated two-fund and three-fund combination coefficients are available in Table 1 in the main body of the paper. See the notes of Table 3 in the main body of the paper for details.

		10MOM		**16ANOM**		**25SBETA**	
		Normal	Elliptical	Normal	Elliptical	Normal	Elliptical
Two-fund rule							
$T = 60$	Skewness	1.675	0.722	0.263	0.281	−0.912	−0.832
	Exc. kurtosis	18.60	9.977	4.266	4.774	16.98	12.40
$T = 120$	Skewness	0.727	0.662	0.381	0.460	−0.285	−0.509
	Exc. kurtosis	5.334	4.554	3.224	3.575	12.62	8.634
$T = 240$	Skewness	−0.153	−0.171	0.829	0.794	−0.914	−0.975
	Exc. kurtosis	4.179	3.952	3.310	3.088	4.483	4.742
Three-fund rule							
$T = 60$	Skewness	1.255	0.635	0.115	0.181	−0.126	−0.692
	Exc. kurtosis	12.74	7.936	4.508	5.508	13.28	10.289
$T = 120$	Skewness	0.472	0.480	−0.259	−0.172	−0.448	−0.716
	Exc. kurtosis	3.558	3.465	3.860	3.904	10.26	7.234
$T = 240$	Skewness	−0.237	−0.240	0.307	0.384	−1.046	−1.102
	Exc. kurtosis	3.298	3.391	2.089	2.134	3.502	4.122

		25SBTM		**25OPINV**		**30IND**	
		Normal	Elliptical	Normal	Elliptical	Normal	Elliptical
Two-fund rule							
$T = 60$	Skewness	0.786	1.101	−0.990	−0.666	−1.547	0.185
	Exc. kurtosis	16.90	15.61	17.08	15.06	49.79	25.36
$T = 120$	Skewness	0.833	0.691	−0.560	−0.298	0.756	0.730
	Exc. kurtosis	5.131	4.794	7.867	7.259	7.602	7.398
$T = 240$	Skewness	0.433	0.393	−0.868	−1.001	−0.716	−0.704
	Exc. kurtosis	7.349	6.358	10.62	10.12	7.515	7.364
Three-fund rule							
$T = 60$	Skewness	0.157	0.945	−0.811	−0.639	−0.966	0.252
	Exc. kurtosis	15.54	15.32	12.12	14.20	22.06	16.17
$T = 120$	Skewness	0.461	0.386	0.079	0.015	0.733	0.926
	Exc. kurtosis	3.342	3.261	3.703	3.416	8.738	10.25
$T = 240$	Skewness	0.368	0.345	−0.408	−0.601	−0.151	−0.226
	Exc. kurtosis	4.445	4.656	6.241	6.555	4.297	5.007

Table 2: In-Sample versus Out-of-Sample Performance of Two-Fund and Three-Fund Rules.

This table reports, for the two-fund and three-fund rules and across the six datasets described in Section VI.B in the main body of the paper, the difference between 1) the average in-sample annualized mean return, volatility, and utility over all estimation windows of size $T = 60$, 120, and 240 months versus 2) the corresponding out-of-sample realized statistic. The difference for the volatility is always negative and we report it positively. Thus, the lower the differences, the better. The combination coefficients are calibrated either to the multivariate normal, t, or elliptical distribution using the exact finite-sample formula. Formulas for the estimated two-fund and three-fund combination coefficients are available in Table 1 in the main body of the paper. See the notes of Table 3 in the main body of the paper for details.

		10MOM			**16ANOM**			**25SBETA**		
		Normal	t	Elliptical	Normal	t	Elliptical	Normal	t	Elliptical
Two-fund rule										
$T = 60$	Mean	0.757	0.704	0.577	1.483	1.391	0.961	1.152	1.124	0.672
	Vol	0.567	0.523	0.352	0.609	0.581	0.388	0.716	0.704	0.391
	Utility	1.209	1.092	0.776	1.976	1.839	1.180	1.676	1.629	0.839
$T = 120$	Mean	0.567	0.529	0.449	0.867	0.808	0.636	0.887	0.848	0.639
	Vol	0.326	0.308	0.244	0.317	0.281	0.241	0.489	0.479	0.299
	Utility	0.792	0.728	0.583	1.081	0.981	0.768	1.245	1.188	0.793
$T = 240$	Mean	0.425	0.401	0.364	0.638	0.593	0.504	0.650	0.629	0.508
	Vol	0.238	0.230	0.206	0.196	0.168	0.157	0.219	0.212	0.165
	Utility	0.615	0.573	0.504	0.848	0.758	0.642	0.773	0.744	0.586
Three-fund rule										
$T = 60$	Mean	0.797	0.740	0.592	1.452	1.362	0.900	1.076	1.050	0.613
	Vol	0.554	0.515	0.347	0.778	0.735	0.474	0.952	0.926	0.425
	Utility	1.325	1.186	0.809	2.299	2.100	1.218	2.033	1.942	0.817
$T = 120$	Mean	0.531	0.493	0.422	0.859	0.797	0.634	0.863	0.813	0.610
	Vol	0.276	0.267	0.217	0.343	0.291	0.249	0.600	0.575	0.320
	Utility	0.759	0.693	0.554	1.142	1.010	0.787	1.445	1.334	0.797
$T = 240$	Mean	0.395	0.381	0.347	0.562	0.520	0.456	0.487	0.472	0.393
	Vol	0.183	0.191	0.177	0.169	0.146	0.139	0.224	0.213	0.162
	Utility	0.566	0.541	0.479	0.751	0.670	0.581	0.676	0.639	0.491

		25SBTM			**25OPINV**			**30IND**		
		Normal	t	Elliptical	Normal	t	Elliptical	Normal	t	Elliptical
Two-fund rule										
$T = 60$	Mean	1.093	1.063	0.672	1.196	1.165	0.706	0.719	0.704	0.553
	Vol	0.771	0.767	0.426	0.906	0.897	0.510	0.637	0.610	0.459
	Utility	1.662	1.620	0.860	1.928	1.880	0.952	1.044	1.006	0.731
$T = 120$	Mean	0.840	0.801	0.582	1.099	1.054	0.717	0.442	0.432	0.388
	Vol	0.542	0.506	0.338	0.567	0.545	0.344	0.332	0.324	0.285
	Utility	1.269	1.176	0.769	1.562	1.477	0.908	0.566	0.550	0.481
$T = 240$	Mean	0.679	0.645	0.523	0.809	0.780	0.615	0.321	0.316	0.293
	Vol	0.484	0.451	0.336	0.630	0.618	0.427	0.203	0.199	0.183
	Utility	1.125	1.036	0.755	1.474	1.409	0.956	0.374	0.367	0.337
Three-fund rule										
$T = 60$	Mean	1.075	1.021	0.635	1.108	1.074	0.644	0.781	0.750	0.563
	Vol	0.898	0.881	0.450	1.036	1.012	0.531	0.769	0.726	0.513
	Utility	1.923	1.820	0.851	2.165	2.068	0.921	1.304	1.219	0.797
$T = 120$	Mean	0.757	0.719	0.538	0.851	0.809	0.593	0.349	0.338	0.309
	Vol	0.581	0.541	0.344	0.565	0.524	0.305	0.457	0.437	0.358
	Utility	1.298	1.181	0.739	1.397	1.270	0.765	0.629	0.588	0.470
$T = 240$	Mean	0.610	0.586	0.486	0.631	0.606	0.503	0.258	0.253	0.235
	Vol	0.461	0.434	0.330	0.549	0.537	0.385	0.226	0.223	0.199
	Utility	1.065	0.983	0.719	1.272	1.196	0.823	0.383	0.366	0.319

Table 3: Out-of-Sample Performance of the Combination of the Sample GMV Portfolio with the Risk-Free Asset.

This table reports the annualized gross and net-of-cost out-of-sample utility (EU), and the mean value of the combination coefficient, for the scaled GMV portfolio in Section VI.C.2 across the six datasets described in Section VI.B in the main body of the paper. The largest EU in each case is depicted in bold. See the notes of Table 3 in the main body of the paper for details.

		Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
A) 10MOM dataset							
$T=60$	Gross EU	0.041	0.043	0.068	0.108	**0.115**	0.090
	Net EU	0.000	0.003	0.031	0.079	**0.088**	0.054
	Mean $\hat{c}$	0.648	0.609	0.560	0.419	0.380	0.236
$T=120$	Gross EU	0.124	0.131	0.136	0.145	**0.147**	0.121
	Net EU	0.105	0.113	0.119	0.130	**0.133**	0.102
	Mean $\hat{c}$	0.816	0.755	0.724	0.599	0.570	0.374
$T=240$	Gross EU	0.078	0.069	0.072	0.077	**0.079**	0.055
	Net EU	0.068	0.059	0.062	0.069	**0.071**	0.045
	Mean $\hat{c}$	0.906	0.844	0.825	0.739	0.720	0.524
B) 16ANOM dataset							
$T=60$	Gross EU	−0.270	−0.303	−0.209	0.028	**0.061**	0.029
	Net EU	−0.350	−0.377	−0.282	−0.030	**0.007**	−0.033
	Mean $\hat{c}$	0.494	0.480	0.438	0.288	0.260	0.090
$T=120$	Gross EU	0.070	0.110	0.120	0.123	**0.129**	0.090
	Net EU	0.037	0.079	0.090	0.100	**0.106**	0.066
	Mean $\hat{c}$	0.727	0.677	0.650	0.491	0.468	0.200
$T=240$	Gross EU	0.262	0.281	**0.283**	0.268	0.267	0.207
	Net EU	0.242	0.263	**0.265**	0.253	0.252	0.190
	Mean $\hat{c}$	0.859	0.811	0.795	0.660	0.644	0.470
C) 25SBETA dataset							
$T=60$	Gross EU	−0.148	−0.242	−0.113	0.154	**0.155**	0.053
	Net EU	−0.293	−0.387	−0.251	0.075	**0.084**	−0.045
	Mean $\hat{c}$	0.303	0.315	0.280	0.140	0.123	0.077
$T=120$	Gross EU	−0.178	−0.156	−0.124	0.054	**0.059**	−0.070
	Net EU	−0.231	−0.206	−0.174	0.018	**0.023**	−0.104
	Mean $\hat{c}$	0.604	0.566	0.543	0.340	0.324	0.128
$T=240$	Gross EU	0.141	0.151	0.153	**0.169**	0.168	0.055
	Net EU	0.111	0.122	0.125	**0.148**	0.147	0.031
	Mean $\hat{c}$	0.791	0.747	0.733	0.543	0.530	0.275

Table 3: Out-of-Sample Performance of the Combination of the Sample GMV Portfolio with the Risk-Free Asset (continued).

		Normal (exact)	t (asymp)	t (exact)	Elliptical (asymp)	Elliptical (exact)	Cross validation
D) 25SBTM dataset							
$T = 60$	Gross EU	−0.021	−0.072	0.020	**0.181**	0.180	0.133
	Net EU	−0.150	−0.200	−0.103	0.107	**0.114**	0.036
	Mean $\hat{c}$	0.303	0.312	0.277	0.140	0.123	0.077
$T = 120$	Gross EU	0.221	0.240	0.251	**0.268**	0.266	0.215
	Net EU	0.157	0.179	0.192	**0.227**	0.226	0.166
	Mean $\hat{c}$	0.604	0.559	0.536	0.335	0.319	0.191
$T = 240$	Gross EU	0.204	0.215	0.219	**0.237**	**0.237**	0.227
	Net EU	0.168	0.182	0.186	0.211	**0.212**	0.198
	Mean $\hat{c}$	0.791	0.721	0.707	0.522	0.510	0.382
E) 25OPINV dataset							
$T = 60$	Gross EU	−0.090	−0.144	−0.042	0.147	**0.151**	0.045
	Net EU	−0.242	−0.297	−0.187	0.063	**0.075**	−0.063
	Mean $\hat{c}$	0.303	0.319	0.284	0.138	0.122	0.082
$T = 120$	Gross EU	0.160	0.203	0.217	**0.246**	0.242	0.161
	Net EU	0.088	0.133	0.149	**0.201**	0.199	0.099
	Mean $\hat{c}$	0.604	0.572	0.549	0.336	0.320	0.223
$T = 240$	Gross EU	0.245	0.266	**0.269**	**0.269**	**0.269**	0.243
	Net EU	0.203	0.225	0.229	**0.239**	**0.239**	0.206
	Mean $\hat{c}$	0.791	0.746	0.732	0.537	0.525	0.449
F) 30IND dataset							
$T = 60$	Gross EU	−0.119	−0.140	−0.080	0.009	0.022	**0.034**
	Net EU	−0.194	−0.221	−0.152	−0.043	−0.023	**−0.013**
	Mean $\hat{c}$	0.217	0.233	0.201	0.138	0.119	0.047
$T = 120$	Gross EU	0.058	0.069	0.079	0.104	**0.107**	0.079
	Net EU	0.029	0.041	0.052	0.083	**0.087**	0.056
	Mean $\hat{c}$	0.541	0.504	0.482	0.360	0.344	0.170
$T = 240$	Gross EU	0.041	0.044	0.047	0.061	**0.063**	0.025
	Net EU	0.025	0.030	0.033	0.049	**0.051**	0.011
	Mean $\hat{c}$	0.754	0.698	0.685	0.561	0.549	0.374

Table 4: Out-of-Sample Performance with Daily Data.

This table reports the gross and net-of-cost annualized out-of-sample utility (EU), and the mean value of the combination coefficients, for the scaled GMV portfolio, the two-fund rule, and the three-fund rules across the six datasets described in Section VI.B in the main body of the paper. We use sample sizes of $T = 5$, 10, and 20 years of daily returns. The largest EU in each case is depicted in bold. See the notes of Table 3 in the main body of the paper for details.

		Normal (exact)	t (asymp)	Elliptical (asymp)	Normal (exact)	t (asymp)	Elliptical (asymp)
		A) 10MOM dataset ($\hat{\nu} = 3.42$)			**B) 25SBTM dataset** ($\hat{\nu} = 2.99$)		
	Scaled GMV portfolio						
$T = 5$ years	Gross EU	2.348	2.400	**2.582**	3.085	3.220	**3.894**
	Net EU	2.232	2.285	**2.472**	2.835	2.973	**3.657**
	Mean $\hat{c}$	0.982	0.953	0.896	0.958	0.936	0.810
$T = 10$ years	Gross EU	2.267	2.312	**2.381**	3.665	3.701	**4.154**
	Net EU	2.190	2.236	**2.306**	3.489	3.526	**3.983**
	Mean $\hat{c}$	0.991	0.966	0.932	0.979	0.958	0.871
$T = 20$ years	Gross EU	2.328	2.367	**2.446**	3.045	3.077	**3.441**
	Net EU	2.258	2.297	**2.377**	2.849	2.882	**3.254**
	Mean $\hat{c}$	0.995	0.971	0.959	0.990	0.963	0.918
	Two-fund rule						
$T = 5$ years	Gross EU	2.159	2.200	**2.401**	2.843	2.905	**3.687**
	Net EU	1.968	2.011	**2.220**	2.500	2.563	**3.352**
	Mean $\hat{c}$	0.541	0.536	0.508	0.459	0.456	0.407
$T = 10$ years	Gross EU	2.312	2.350	**2.438**	3.424	3.454	**3.970**
	Net EU	2.192	2.231	**2.321**	3.157	3.188	**3.710**
	Mean $\hat{c}$	0.586	0.579	0.559	0.518	0.515	0.473
$T = 20$ years	Gross EU	2.464	2.508	**2.606**	2.568	2.600	**3.024**
	Net EU	2.359	2.403	**2.503**	2.299	2.332	**2.765**
	Mean $\hat{c}$	0.681	0.675	0.660	0.639	0.637	0.603
	Three-fund rule						
$T = 5$ years	Gross EU	2.209	2.267	**2.479**	2.814	2.956	**3.769**
	Net EU	2.039	2.098	**2.317**	2.490	2.635	**3.453**
	Mean $\hat{c}_1$	0.342	0.340	0.326	0.281	0.280	0.259
	Mean $\hat{c}_2/\hat{\mu}_g$	0.640	0.613	0.570	0.678	0.657	0.551
$T = 10$ years	Gross EU	2.350	2.400	**2.477**	3.684	3.725	**4.236**
	Net EU	2.241	2.292	**2.371**	3.435	3.478	**3.993**
	Mean $\hat{c}_1$	0.409	0.407	0.395	0.358	0.356	0.335
	Mean $\hat{c}_2/\hat{\mu}_g$	0.581	0.558	0.537	0.621	0.602	0.536
$T = 20$ years	Gross EU	2.514	2.560	**2.651**	2.875	2.911	**3.326**
	Net EU	2.417	2.464	**2.557**	2.615	2.653	**3.077**
	Mean $\hat{c}_1$	0.527	0.524	0.514	0.488	0.486	0.466
	Mean $\hat{c}_2/\hat{\mu}_g$	0.469	0.446	0.445	0.502	0.477	0.451

Table 4: Out-of-Sample Performance with Daily Data (continued).

		Normal (exact)	t (asymp)	Elliptical (asymp)	Normal (exact)	t (asymp)	Elliptical (asymp)
		C) 25OPINV dataset ($\hat{\nu} = 6.27$)			**D) 30IND dataset** ($\hat{\nu} = 3.86$)		
	Scaled GMV portfolio						
$T = 5$ years	Gross EU	5.951	6.064	**7.248**	2.499	2.685	**3.453**
	Net EU	5.511	5.627	**6.829**	2.291	2.479	**3.256**
	Mean $\hat{c}$	0.958	0.944	0.810	0.951	0.923	0.822
$T = 10$ years	Gross EU	4.928	4.994	**5.871**	2.771	2.906	**3.243**
	Net EU	4.593	4.660	**5.550**	2.643	2.778	**3.119**
	Mean $\hat{c}$	0.979	0.968	0.866	0.975	0.952	0.883
$T = 20$ years	Gross EU	−0.539	−0.470	**0.430**	3.057	3.167	**3.461**
	Net EU	−0.855	−0.785	**0.132**	2.943	3.054	**3.350**
	Mean $\hat{c}$	0.990	0.982	0.912	0.988	0.967	0.927
	Two-fund rule						
$T = 5$ years	Gross EU	5.758	5.835	**7.012**	2.100	2.246	**3.007**
	Net EU	5.157	5.236	**6.437**	1.790	1.936	**2.708**
	Mean $\hat{c}$	0.621	0.617	0.540	0.477	0.472	0.435
$T = 10$ years	Gross EU	4.725	4.766	**5.671**	2.650	2.739	**3.059**
	Net EU	4.300	4.343	**5.260**	2.452	2.542	**2.867**
	Mean $\hat{c}$	0.681	0.677	0.612	0.530	0.525	0.496
$T = 20$ years	Gross EU	−0.692	−0.633	**0.300**	2.699	2.789	**3.084**
	Net EU	-1.068	-1.008	**−0.059**	2.548	2.638	**2.936**
	Mean $\hat{c}$	0.750	0.747	0.695	0.600	0.595	0.574
	Three-fund rule						
$T = 5$ years	Gross EU	5.866	5.982	**7.296**	2.217	2.412	**3.247**
	Net EU	5.339	5.458	**6.786**	1.949	2.145	**2.989**
	Mean $\hat{c}_1$	0.313	0.312	0.283	0.228	0.228	0.218
	Mean $\hat{c}_2/\hat{\mu}_g$	0.645	0.632	0.527	0.722	0.695	0.603
$T = 10$ years	Gross EU	4.718	4.789	**5.752**	2.754	2.888	**3.226**
	Net EU	4.331	4.403	**5.377**	2.594	2.729	**3.070**
	Mean $\hat{c}_1$	0.361	0.360	0.329	0.200	0.200	0.195
	Mean $\hat{c}_2/\hat{\mu}_g$	0.618	0.608	0.537	0.775	0.752	0.689
$T = 20$ years	Gross EU	−0.954	−0.882	**0.084**	2.938	3.049	**3.346**
	Net EU	-1.308	-1.235	**−0.253**	2.812	2.923	**3.223**
	Mean $\hat{c}_1$	0.477	0.476	0.452	0.159	0.158	0.157
	Mean $\hat{c}_2/\hat{\mu}_g$	0.513	0.507	0.461	0.829	0.809	0.770

IV. Proofs of Theoretical Results

Proof of Proposition 1

The expected out-of-sample utility of the two-fund rule $\hat{w}_{2f}(c)$ is

$$\mathbb{E}[U(\hat{w}_{2f}(c))] = \frac{c}{\gamma}\tilde{\mu}_1 - \frac{c^2}{2\gamma}\tilde{\sigma}_1^2, \tag{A3}$$

which yields the optimal $c^\star$ in (21). The expected out-of-sample utility of the three-fund rule $\hat{w}_{3f}(c_1, c_2)$ is

$$\mathbb{E}[U(\hat{w}_{3f}(c_1, c_2))] = \frac{c_1}{\gamma}\tilde{\mu}_1 + \frac{c_2}{\gamma}\tilde{\mu}_2 - \frac{c_1^2}{2\gamma}\tilde{\sigma}_1^2 - \frac{c_2^2}{2\gamma}\tilde{\sigma}_2^2 - \frac{c_1 c_2}{\gamma}\tilde{\sigma}_{12}, \tag{A4}$$

which yields the optimal $(c_1^\star, c_2^\star)$ in (22) and completes the proof.

Proof of Proposition 2

Part 1. Because $\hat{\mu}$ and $\hat{\Sigma}$ are asymptotically unbiased for a fixed N, the sample mean-variance portfolio $\hat{w}$ is asymptotically unbiased too. To find its asymptotic covariance matrix, we write $\hat{w}$ as a function of $\hat{\mu}$ and $\text{vec}(\hat{\Sigma}^{-1})$ as

$$\hat{w} = \frac{1}{\gamma}(\hat{\mu}^\top \otimes I_N)\text{vec}(\hat{\Sigma}^{-1}). \tag{A5}$$

Moreover, the derivative of $\text{vec}(\hat{\Sigma}^{-1})$ with respect to $\text{vec}(\hat{\Sigma})$ is

$$\frac{\partial \text{vec}(\hat{\Sigma}^{-1})}{\partial \text{vec}(\hat{\Sigma})^\top} = -\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1}. \tag{A6}$$

Therefore, using the delta method, we can find the asymptotic covariance matrix of $\hat{w}$ from the asymptotic covariance matrix of $(\hat{\mu}, \text{vec}(\hat{\Sigma}))$, which from Muirhead (1982, p.82, 89) is

$$\text{Avar}\begin{bmatrix} \hat{\mu} \\ \text{vec}(\hat{\Sigma}) \end{bmatrix} = \begin{bmatrix} \Sigma & 0_{N\times N^2} \\ 0_{N^2\times N} & (1+\kappa)(I_{N^2}+K_N)(\Sigma\otimes\Sigma)+\kappa\,\text{vec}(\Sigma)\text{vec}(\Sigma)^\top \end{bmatrix}, \tag{A7}$$

where K_N is an $N^2 \times N^2$ commutation matrix such that $K_N \text{vec}(A) = \text{vec}(A^\top)$ for an $N \times N$ matrix A. Specifically, given (A5)–(A6), we have

$$\frac{\partial \hat{w}}{\partial \hat{\mu}^\top} = \frac{1}{\gamma}\hat{\Sigma}^{-1}, \tag{A8}$$

$$\frac{\partial \hat{w}}{\partial \text{vec}(\hat{\Sigma})^\top} = -\frac{1}{\gamma}(\hat{\Sigma}^{-1}\hat{\mu}) \otimes \hat{\Sigma}^{-1}, \tag{A9}$$

and therefore the asymptotic covariance matrix of $\hat{w}$ is

$$\text{Avar}[\hat{w}] = \frac{1}{\gamma^2}\left[\Sigma^{-1}, -(\Sigma^{-1}\mu)\otimes\Sigma^{-1}\right]\text{Avar}\begin{bmatrix} \hat{\mu} \\ \text{vec}(\hat{\Sigma}) \end{bmatrix}\left[\Sigma^{-1}, -(\Sigma^{-1}\mu)\otimes\Sigma^{-1}\right]^\top, \tag{A10}$$

which after simplification corresponds to the desired result in (23).

Part 2. Given that $\hat{w}$ is asymptotically unbiased as shown in part 1, $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ are asymptotically unbiased too. To find the asymptotic covariance matrix of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$, we use the delta method. Let $h(w) = [w^\top \mu, w^\top \Sigma w]^\top$. Then, the Jacobian of h evaluated at $w^\star$ is

$\nabla h(w^\star) = [\mu^\top, 2w^{\star\top}\Sigma]^\top = [\mu^\top, 2\mu^\top/\gamma]^\top$. Therefore, the asymptotic covariance matrix of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ is $\text{Avar}[\tilde{\mu}_p, \tilde{\sigma}_p^2] = \nabla h(w^\star)^\top \text{Avar}[\hat{w}] \nabla h(w^\star)$, where $\text{Avar}[\hat{w}]$ is given by (23), which corresponds to the desired result in (24).

Part 3. Given $\hat{w} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}$ and $U(\hat{w}) = \hat{w}^\top \mu - \hat{w}^\top \Sigma \hat{w}$, we have

$$\text{(A11)} \qquad D = \left.\frac{\partial U(\hat{w})}{\partial \hat{w}}\right|_{\hat{w}=w^*} = \mu - \gamma\Sigma w^\star = 0_N,$$

$$\text{(A12)} \qquad H = \left.\frac{\partial^2 U(\hat{w})}{\partial \hat{w}\partial \hat{w}^\top}\right|_{\hat{w}=w^*} = -\gamma\Sigma.$$

Because $D = 0_N$, it holds that

$$\text{(A13)} \qquad T[U(\hat{w}) - U(w^\star)] \xrightarrow{d} \sum_{i=1}^{N} \lambda_i X_i,$$

where the X_i's are independent χ_1^2 random variables and the λ_i's are the eigenvalues of $HS/2$, where S is the asymptotic covariance matrix of $\hat{w}$ in (23):

$$\text{(A14)} \qquad S = \frac{1}{\gamma^2}\Big((1 + (1+\kappa)\theta^2)\Sigma^{-1} + (1+2\kappa)\Sigma^{-1}\mu\mu^\top\Sigma^{-1}\Big).$$

The matrix $HS/2$ is

$$\text{(A15)} \qquad \frac{1}{2}HS = -\frac{1+(1+\kappa)\theta^2}{2\gamma}\left(I_N + \frac{1+2\kappa}{1+(1+\kappa)\theta^2}\mu\mu^\top\Sigma^{-1}\right).$$

All the eigenvalues of I_N are one, and $\mu\mu^\top\Sigma^{-1}$ has $N-1$ zero eigenvalues and one eigenvalue equal to the trace of $\mu\mu^\top\Sigma^{-1}$, i.e., θ^2. Therefore, we have $\lambda_1 = \cdots = \lambda_{N-1} = -\frac{1+(1+\kappa)\theta^2}{2\gamma}$ and

$\lambda_N = -\frac{1+(1+\kappa)\theta^2}{2\gamma}\left(1+\frac{(1+2\kappa)\theta^2}{1+(1+\kappa)\theta^2}\right) = -\frac{1+(2+3\kappa)\theta^2}{2\gamma}$, which yields the desired result in (25).

Part 4. Let the risk-aversion coefficient $\gamma = 1$ for notational simplicity, which is without loss of generality because it is clear that $\tilde{\mu}_p$, $\tilde{\sigma}_p^2$, and $U(\hat{w})$ are proportional to $1/\gamma$, $1/\gamma^2$, and $1/(2\gamma)$, respectively. Let $\phi = [\mu^\top, w^{\star\top}]^\top$ and $\hat{\phi} = [\hat{\mu}^\top, \hat{w}^\top]^\top$. Note that $\hat{\phi}$ can be written as the generalized-method-of-moments estimator of ϕ based on the following moment conditions:

$$\mathbb{E}[g_t(\phi)] = \mathbb{E}\begin{bmatrix} r_t - \mu \\ C_t w^\star - \mu \end{bmatrix} = 0_{2N}, \tag{A16}$$

where $C_t = (r_t - \mu)(r_t - \mu)^\top$.

We derive the first-order bias of $\hat{\phi}$ by using a stochastic expansion of $\hat{\phi}$ based on the results of Bao and Ullah (2007, 2009), which suggest

$$\hat{\phi} = \phi + a_{-1/2} + a_{-1} + a_{-3/2} + O_p(T^{-2}), \tag{A17}$$

where

$$a_{-1/2} = -\mathbb{E}[H_1]^{-1}\bar{g}, \tag{A18}$$

$$a_{-1} = -\mathbb{E}[H_1]^{-1}V_1 a_{-1/2} - \frac{1}{2}\mathbb{E}[H_1]^{-1}\mathbb{E}[H_2](a_{-1/2}\otimes a_{-1/2}), \tag{A19}$$

$$\begin{aligned} a_{-3/2} = {} & -\mathbb{E}[H_1]^{-1}V_1 a_{-1} - \frac{1}{2}\mathbb{E}[H_1]^{-1}V_2(a_{-1/2}\otimes a_{-1/2}) \\ & -\frac{1}{2}\mathbb{E}[H_1]^{-1}\mathbb{E}[H_2](a_{-1/2}\otimes a_{-1} + a_{-1}\otimes a_{-1/2}) \\ & -\frac{1}{6}\mathbb{E}[H_1]^{-1}\mathbb{E}[H_3](a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}), \end{aligned} \tag{A20}$$

with $\bar{g} = \frac{1}{T}\sum_{t=1}^{T} g_t(\phi)$, $H_i = \nabla^i \bar{g}$, and $V_i = H_i - \mathbb{E}[H_i]$.[1]

We now provide explicit expressions of $a_{-1/2}$ and a_{-1}. For $a_{-3/2}$, we can show that its expectation is $O(T^{-2})$.[2] For $a_{-1/2}$, we have

$$H_1 = \begin{bmatrix} -I_N & 0_{N\times N} \\ -z_t I_N - (r_t-\mu)w^{\star\top} - I_N & C_t \end{bmatrix}, \tag{A21}$$

where $z_t = (r_t-\mu)^\top w^\star$, and thus,

$$\mathbb{E}[H_1] = \begin{bmatrix} -I_N & 0_{N\times N} \\ -I_N & \Sigma \end{bmatrix}, \tag{A22}$$

$$\mathbb{E}[H_1]^{-1} = \begin{bmatrix} -I_N & 0_{N\times N} \\ -\Sigma^{-1} & \Sigma^{-1} \end{bmatrix}. \tag{A23}$$

It follows that $a_{-1/2}$ is equal to

$$a_{-1/2} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} r_t-\mu \\ a_t \end{bmatrix}, \tag{A24}$$

where

$$a_t = \Sigma^{-1}(r_t-\mu)(1-z_t) + w^*, \tag{A25}$$

[1] $\nabla^i \bar{g}$ is the matrix of i-th order partial derivative of $\bar{g}(\phi)$ and is obtained recursively. Specifically, If $\bar{g}(\phi)$ is a k-vector function of ϕ, the j-th element of the l-th row of $A_i \equiv \nabla^i \bar{g}$ (a $k\times k^i$ matrix) is the $1\times k$ vector $a^i_{lj} = \partial a^{i-1}_{lj}/\partial\phi^\top$.

[2] An explicit expression of $\mathbb{E}[a_{-3/2}]$ is available upon request.

and it is obvious that $\mathbb{E}[a_{-1/2}] = 0_{2N}$.

For a_{-1}, we have from (A21) and (A22) that

$$V_1 = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} 0_{N\times N} & 0_{N\times N} \\ -z_t I_N - (r_t-\mu)w^{\star\top} & C_t - \Sigma \end{bmatrix}. \tag{A26}$$

Moreover,

$$H_2 = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} 0_{N\times 2N^2} & 0_{N\times 2N^2} \\ A_t & B_t \end{bmatrix}, \tag{A27}$$

where

$$A_t = I_N \otimes [w^{\star\top},\ -(r_t-\mu)^\top] + w^{\star\top}\otimes D_1^\top - (r_t-\mu)\mathrm{vec}(D_2)^\top, \tag{A28}$$

$$B_t = -(r_t-\mu)^\top \otimes D_1^\top - (r_t-\mu)\mathrm{vec}(D_1)^\top, \tag{A29}$$

with

$$D_1 = \begin{bmatrix} I_N \\ 0_{N\times N} \end{bmatrix}, \tag{A30}$$

$$D_2 = \begin{bmatrix} 0_{N\times N} \\ I_N \end{bmatrix}. \tag{A31}$$

Therefore,

(A32) $$\mathbb{E}[H_2] = \begin{bmatrix} 0_{N\times 2N^2} & 0_{N\times 2N^2} \\ I_N \otimes [w^{\star\top},\, 0_N^\top] + w^{\star\top} \otimes D_1^\top & 0_{N\times 2N^2} \end{bmatrix}.$$

It follows that a_{-1} is equal to

(A33) $$a_{-1} = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\begin{bmatrix} 0_N \\ z_s\Sigma^{-1}(r_t-\mu) - \Sigma^{-1}(C_s-\Sigma)a_t \end{bmatrix}.$$

Using the fact that r_s is independent of r_t if $s \neq t$, $\mathbb{E}[z_t] = 0$, $\mathbb{E}[a_t] = 0_N$, and $\mathbb{E}[C_t - \Sigma] = 0_{N\times N}$, the expectation of a_{-1} is

(A34) $$\mathbb{E}[a_{-1}] = \frac{1}{T}\begin{bmatrix} 0_N \\ \mathbb{E}[z_t\Sigma^{-1}(r_t-\mu)] - \mathbb{E}[\Sigma^{-1}(C_t-\Sigma)a_t] \end{bmatrix}.$$

Let $y_t = \Sigma^{-1/2}(r_t - \mu)$. Then,

(A35) $$\mathbb{E}[z_t\Sigma^{-1}(r_t-\mu)] = \mathbb{E}[y_t y_t^\top]w^\star = w^\star,$$

(A36) $$\mathbb{E}[\Sigma^{-1}(C_t-\Sigma)a_t] = \mathbb{E}[y_t y_t^\top]w^\star - \mathbb{E}[(y_t^\top y_t)y_t y_t^\top]w^\star = [1-(N+2)(1+\kappa)]w^\star,$$

where we use the fact that $\mathbb{E}[(y_t^\top y_t)y_t y_t^\top] = (N+2)(1+\kappa)I_N$ under the multivariate elliptical distribution assumption. Therefore,

(A37) $$\mathbb{E}[a_{-1}] = \begin{bmatrix} 0_N \\ \frac{1}{T}(N+2)(1+\kappa)w^\star \end{bmatrix}.$$

Using $\mathbb{E}[a_{-1/2}] = 0_{2N}$ and $\mathbb{E}[a_{-1}]$ in (A37), the first-order bias of $\hat{w}$ is

$$\mathbb{E}[\hat{w}] - w^{\star} = \frac{(N+2)(1+\kappa)w^{\star}}{T} + O(T^{-2}). \tag{A38}$$

It follows that

$$\mathbb{E}[\tilde{\mu}_p] - \mu_p = \frac{(N+2)(1+\kappa)\mu_p}{T} + O(T^{-2}), \tag{A39}$$

which corresponds to the desired result in (26) after adding back $1/\gamma$.

Turning to the first-order bias of $\tilde{\sigma}_p^2$, we use (A17) to obtain

$$\mathbb{E}[\tilde{\sigma}_p^2] - \sigma_p^2 = 2w^{\star\top}\Sigma\mathbb{E}[\tilde{a}_{-1}] + \mathbb{E}[\tilde{a}_{-1/2}^{\top}\Sigma\tilde{a}_{-1/2}] + O(T^{-2}), \tag{A40}$$

where $\tilde{a}_{-1/2}$ and $\tilde{a}_{-1}$ are the last N elements of $a_{-1/2}$ and a_{-1}, respectively. Given $\mathbb{E}[\tilde{a}_{-1}]$ in (A37), we have

$$2w^{\star\top}\Sigma\mathbb{E}[\tilde{a}_{-1}] = \frac{2(N+2)(1+\kappa)\theta^2}{T}. \tag{A41}$$

Moreover, given $\tilde{a}_{-1/2}$ in (A24), we have

$$\tilde{a}_{-1/2}^{\top}\Sigma\tilde{a}_{-1/2} = \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} a_s^{\top}\Sigma a_t. \tag{A42}$$

Therefore,

$$\mathbb{E}[\tilde{a}_{-1/2}^{\top}\Sigma\tilde{a}_{-1/2}] = \frac{1}{T}\mathbb{E}[a_t^{\top}\Sigma a_t]$$

$$= \frac{1}{T}\Big(\theta^2 + 2\mathbb{E}[(1-z_t)(r_t-\mu)^{\top}\Sigma^{-1}\mu] + \mathbb{E}[(1-z_t)^2(r_t-\mu)^{\top}\Sigma^{-1}(r_t-\mu)]\Big). \tag{A43}$$

It holds that

$$\mathbb{E}[(1-z_t)(r_t-\mu)^{\top}\Sigma^{-1}\mu] = -\theta^2, \tag{A44}$$

$$\mathbb{E}[(1-z_t)^2(r_t-\mu)^{\top}\Sigma^{-1}(r_t-\mu)] = \mathbb{E}[y_t^{\top}y_t] + \mathbb{E}[\mu^{\top}\Sigma^{-1/2}(y_t^{\top}y_t)y_t y_t^{\top}\Sigma^{-1/2}\mu]$$

$$= N + (N+2)(1+\kappa)\theta^2, \tag{A45}$$

and thus,

$$\mathbb{E}[\tilde{a}_{-1/2}^{\top}\Sigma\tilde{a}_{-1/2}] = \frac{N + [(N+2)(1+\kappa)-1]\theta^2}{T}. \tag{A46}$$

It follows that

$$\mathbb{E}[\tilde{\sigma}_p^2] - \sigma_p^2 = \frac{N + [3(N+2)(1+\kappa)-1]\theta^2}{T} + O(T^{-2}), \tag{A47}$$

which corresponds to the desired result in (27) after adding back $1/\gamma^2$. Finally, the first-order bias of $U(\hat{w})$ in (28) is directly obtained from (26)–(27). This completes the proof.

Proof of Proposition 3

Part 1. This result is a direct consequence of El Karoui (2010, equation (9)) after defining the parameter η as $(1-\rho)\mathfrak{s}$, where $\mathfrak{s}$ is defined in El Karoui (2010, equation (4)). El Karoui (2010) shows in p. 3506 that $\mathfrak{s} \geq 1/(1-\rho)$, and thus $\eta \geq 1$. Finally, we show in part 4 of this proposition that $\lim_{\rho\to 1} \eta = \mathbb{E}[1/\tau_t]$.

Part 2. From El Karoui (2013, equation (3.4)), we have

$$\tilde{\sigma}_p^2 = \frac{1}{\gamma^2}\hat{\mu}^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu} \xrightarrow{p} \frac{1}{\gamma^2}\mu^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mu + \frac{\eta\rho}{\gamma^2(1-\rho)^3}. \tag{A48}$$

Moreover, using the last equation in p. 748 of El Karoui (2013), we obtain

$$\mu^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mu \xrightarrow{p} \frac{\varphi\theta^2}{(1-\rho)^3}, \tag{A49}$$

where φ is defined as $(1-\rho)^3\xi$ with ξ defined in El Karoui (2013, equation (3.2)). El Karoui (2013, fact 3.1) shows that $\xi \geq \mathfrak{s}^2/(1-\rho)$, which in our notation is equivalent to $\varphi \geq \eta^2$. Finally, we show in part 4 of this proposition that $\lim_{\rho\to 1} \varphi = (\mathbb{E}[1/\tau_t])^2$.

Part 3. Equation (33) is a direct consequence of (29)–(31). As shown in part 4 of this proposition, the case of multivariate normally distributed returns corresponds to $\eta = \varphi = 1$. Therefore, the limit of $U(\hat{w})$ in (33) is smaller than that under normality if

$$\frac{2\eta}{1-\rho} - \frac{\varphi}{(1-\rho)^3} \leq \frac{2}{1-\rho} - \frac{1}{(1-\rho)^3}, \tag{A50}$$

which is equivalent to

$$\rho \geq 1 - \sqrt{\frac{\varphi - 1}{2(\eta - 1)}}. \tag{A51}$$

Condition (A51) always holds because the right-hand side of (A51) is negative:

$$1 - \sqrt{\frac{\varphi - 1}{2(\eta - 1)}} \leq 1 - \sqrt{\frac{\eta^2 - 1}{2(\eta - 1)}} = 1 - \sqrt{\frac{\eta + 1}{2}} \leq 0, \tag{A52}$$

where the first and last inequalities hold because $\varphi \geq \eta^2$ and $\eta \geq 1$, respectively.

Part 4. El Karoui (2010, p. 3506) and El Karoui (2013, Fact 3.1) show that $\eta = \varphi = 1$ when returns are multivariate normally distributed. Moreover, when $\rho \to 0$ we recover the fixed N asymptotic regime in Proposition 2, in which case $\eta = \varphi = 1$ too because $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ are asymptotically unbiased. Finally, we show that $\lim_{\rho \to 1} \eta = \mathbb{E}[1/\tau_t]$ and $\lim_{\rho \to 1} \varphi = (\mathbb{E}[1/\tau_t])^2$. For η, it is direct from its definition in (30). For φ, we apply L'Hopital's rule to obtain

$$\lim_{\rho \to 1} \varphi = \left(\lim_{\rho \to 1} \frac{2}{\eta^3} \frac{\partial \eta}{\partial \rho} + \mathbb{E}\left[\frac{\tau^2}{(1 - \rho + \rho \eta \tau_t)^2} \right] + 2\rho \mathbb{E}\left[\frac{\tau_t^2 (1 - \tau_t(\eta + \rho \frac{\partial \eta}{\partial \rho}))}{(1 - \rho + \rho \eta \tau_t)^3} \right] \right)^{-1}. \tag{A53}$$

We apply implicit differentiation on (30) to obtain

$$\frac{\partial \eta}{\partial \rho} = \frac{\mathbb{E}\left[\frac{1 - \eta \tau_t}{(1 - \rho + \rho \eta \tau_t)^2} \right]}{\mathbb{E}\left[\frac{\rho \tau_t}{(1 - \rho + \rho \eta \tau_t)^2} \right]}, \tag{A54}$$

whose limit as $\rho \to 1$ is

$$\lim_{\rho \to 1} \frac{\partial \eta}{\partial \rho} = \frac{\mathbb{E}[1/\tau_t^2]}{\mathbb{E}[1/\tau_t]} - \mathbb{E}[1/\tau_t]. \tag{A55}$$

Finally, using (A55) and $\lim_{\rho \to 1} \eta = \mathbb{E}[1/\tau_t]$, the limit in (A53) simplifies to $(\mathbb{E}[1/\tau_t])^2$. This completes the proof.

Proof of Proposition 4

Using the results in the proof of Proposition 3, the quantities needed in Proposition 1 to identify the optimal combination coefficients are

$$\tilde{\mu}_1 = \frac{\eta\theta^2}{1-\rho}, \tag{A56}$$

$$\tilde{\mu}_2 = \frac{\eta}{1-\rho}\frac{\theta_g^2}{\mu_g}, \tag{A57}$$

$$\tilde{\sigma}_1^2 = \frac{\varphi\theta^2 + \eta\rho}{(1-\rho)^3}, \tag{A58}$$

$$\tilde{\sigma}_2^2 = \frac{\varphi}{(1-\rho)^3}\frac{\theta_g^2}{\mu_g^2}, \tag{A59}$$

$$\tilde{\sigma}_{12} = \frac{\varphi}{(1-\rho)^3}\frac{\theta_g^2}{\mu_g}. \tag{A60}$$

Plugging (A56)–(A60) into (21)–(22) delivers the optimal two-fund and three-fund combination coefficients in (34)–(36). Finally, these combination coefficients are smaller than those under normally distributed returns because $\varphi \geq \eta^2 \geq \eta$. This completes the proof.

Proof of Proposition 5

Part 1. Given $\tilde{\mu}_1$ and $\tilde{\sigma}_1^2$ in (A56) and (A58), we have from (A3) that the out-of-sample utility of the two-fund rule $\hat{w}_{2f}(c)$ converges to

$$U(\hat{w}_{2f}(c)) \xrightarrow{p} \frac{c}{\gamma}\frac{\eta\theta^2}{1-\rho} - \frac{c^2}{2\gamma}\frac{\varphi\theta^2 + \eta\rho}{(1-\rho)^3}. \tag{A61}$$

Plugging the limit of $c^\star$ in (34) into (A61) yields the desired result in (37). This utility is larger than that under the multivariate normal distribution if and only if

$$\text{(A62)} \qquad \frac{\eta\theta^2}{\frac{\varphi}{\eta}\theta^2+\rho} > \frac{\theta^2}{\theta^2+\rho},$$

which is equivalent to the condition $\theta^2 < \rho\eta(\eta-1)/(\varphi-\eta^2)$.

Part 2. Given $\tilde{\mu}_1$, $\tilde{\mu}_2$, $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$, and $\tilde{\sigma}_{12}$ in (A56)–(A60), we have from (A4) that the out-of-sample utility of the three-fund rule $\hat{w}_{3f}(c)$ converges to

(A63)

$$U(\hat{w}_{3f}(c_1,c_2)) \xrightarrow{p} \frac{c_1}{\gamma}\frac{\eta\theta^2}{1-\rho} + \frac{c_2}{\gamma}\frac{\eta}{1-\rho}\frac{\theta_g^2}{\mu_g} - \frac{c_1^2}{2\gamma}\frac{\varphi\theta^2+\eta\rho}{(1-\rho)^3} - \frac{c_2^2}{2\gamma}\frac{\varphi}{(1-\rho)^3}\frac{\theta_g^2}{\mu_g^2} - \frac{c_1c_2}{\gamma}\frac{\varphi}{(1-\rho)^3}\frac{\theta_g^2}{\mu_g}.$$

Plugging the limit of $(c_1^\star, c_2^\star)$ in (35)–(36) into (A63) yields the desired result in (38). This utility is larger than that under the multivariate normal distribution if and only if

$$\text{(A64)} \qquad \frac{\eta\psi^2}{\frac{\varphi}{\eta}\psi^2+\rho}\left(1+\frac{\eta\rho\theta_g^2}{\varphi\theta^2\psi^2}\right) > \frac{\psi^2}{\psi^2+\rho}\left(1+\frac{\rho\theta_g^2}{\theta^2\psi^2}\right),$$

which is equivalent to the condition $\eta(\psi^2+\rho)(\theta^2\psi^2+\frac{\eta}{\varphi}\theta_g^2\rho) > (\frac{\varphi}{\eta}\psi^2+\rho)(\theta^2\psi^2+\theta_g^2\rho)$. This completes the proof.

Proof of Proposition 6

Part 1. Given the distribution of τ_t in (2), the expectation in (30) evaluates to

$$\text{(A65)} \qquad \mathbb{E}\big[(1-\rho+\rho\eta\tau_t)^{-1}\big] = \frac{\nu}{2(1-\rho)} e^y E_{\frac{\nu}{2}+1}(y),$$

where $y = (\nu-2)\rho\eta/[2(1-\rho)]$. Using the recursive relation on the exponential integral,

$$\text{(A66)} \qquad E_{\frac{\nu}{2}+1}(y) = \frac{e^{-y} - yE_{\frac{\nu}{2}}(y)}{\nu/2}.$$

Therefore, (A65) simplifies to

$$\text{(A67)} \qquad \mathbb{E}\big[(1-\rho+\rho\eta\tau_t)^{-1}\big] = \frac{1-ye^yE_{\frac{\nu}{2}}(y)}{1-\rho},$$

and thus condition (30) means that η is the solution to (39). Finally, $\mathbb{E}[1/\tau_t] = \nu/(\nu-2)$, and thus $1 \le \eta \le \nu/(\nu-2)$ from Proposition 3.

Part 2. Given the distribution of τ_t in (2), the expectation in (32) evaluates to

$$\text{(A68)} \qquad \mathbb{E}\left[\frac{\tau_t^2}{(1-\rho+\rho\eta\tau_t)^2}\right] = \frac{1}{(1-\rho)^2}\left[-\left(\frac{\nu-2}{2}\right) + \left(\frac{\nu-2}{2}\right)^2\left(1+\frac{\rho\eta}{1-\rho}\right)e^yE_{\frac{\nu}{2}-1}(y)\right],$$

where $y = (\nu-2)\rho\eta/[2(1-\rho)]$. Using the recursive relation on the exponential integral,

$$\text{(A69)} \qquad E_{\frac{\nu}{2}-1}(y) = \frac{e^{-y} - (\frac{\nu}{2}-1)E_{\frac{\nu}{2}}(y)}{y}.$$

Moreover, it holds that $e^y E_{\frac{\nu}{2}}(y) = \rho / y$ from (39). Therefore,

$$e^y E_{\frac{\nu}{2}-1}(y) = \frac{1}{y} - \frac{\rho(\frac{\nu}{2}-1)}{y^2}. \tag{A70}$$

Plugging (A70) and $y = \frac{(\nu-2)\rho\eta}{2(1-\rho)}$ into (A68) yields

$$\mathbb{E}\left[\frac{\tau_t^2}{(1-\rho+\rho\eta\tau_t)^2}\right] = \frac{(\nu-2)(\eta-1)}{2\rho\eta^2}. \tag{A71}$$

Plugging (A71) into (32) delivers the desired formula for φ in (40). Finally, $(\mathbb{E}[1/\tau_t])^2 = \nu^2/(\nu-2)^2$, and thus $\eta^2 \leq \varphi \leq \nu^2/(\nu-2)^2$ from Proposition 3. This completes the proof.

Proof of Proposition 7

Given the definition of Y, Λ, M, and the stochastic representation for the multivariate elliptical distribution in (1), we can write the sample mean and covariance matrix as

$$\hat{\mu} = \mu + \frac{1}{T}\Sigma^{1/2}Y^\top \Lambda 1_T, \tag{A72}$$

$$\hat{\Sigma} = \frac{1}{T}\Sigma^{1/2}Y^\top \Lambda M \Lambda Y \Sigma^{1/2}. \tag{A73}$$

Therefore, the out-of-sample mean and variance of the sample mean-variance portfolio $\hat{w}$ are

$$\tilde{\mu}_p = \frac{1}{\gamma}\Big[T\mu^\top\Sigma^{-1/2}(Y^\top\Lambda M\Lambda Y)^{-1}\Sigma^{-1/2}\mu + 1_T^\top\Lambda Y(Y^\top\Lambda M\Lambda Y)^{-1}\Sigma^{-1/2}\mu\Big], \tag{A74}$$

$$\tilde{\sigma}_p^2 = \frac{1}{\gamma^2}\Big[T^2\mu^\top\Sigma^{-1/2}(Y^\top\Lambda M\Lambda Y)^{-2}\Sigma^{-1/2}\mu + 2T1_T^\top\Lambda Y(Y^\top\Lambda M\Lambda Y)^{-2}\Sigma^{-1/2}\mu$$

$$\text{(A75)} \qquad + 1_T^\top \Lambda Y (Y^\top \Lambda M \Lambda Y)^{-2} Y^\top \Lambda 1_T \Big].$$

By symmetry, the expectation of the second terms in (A74)–(A75) is zero because Y has zero mean. Therefore, the expected out-of-sample mean and variance are

$$\text{(A76)} \quad \mathbb{E}[\tilde{\mu}_p] = \frac{1}{\gamma} T \mu^\top \Sigma^{-1/2} \mathbb{E}\Big[(Y^\top \Lambda M \Lambda Y)^{-1}\Big] \Sigma^{-1/2} \mu,$$

$$\text{(A77)} \quad \mathbb{E}[\tilde{\sigma}_p^2] = \frac{1}{\gamma^2} T^2 \mu^\top \Sigma^{-1/2} \mathbb{E}\Big[(Y^\top \Lambda M \Lambda Y)^{-2}\Big] \Sigma^{-1/2} \mu + \frac{1}{\gamma^2} \mathbb{E}\Big[1_T^\top \Lambda Y (Y^\top \Lambda M \Lambda Y)^{-2} Y^\top \Lambda 1_T\Big].$$

By definition of k_3, the second term in (A77) is equal to

$$\text{(A78)} \qquad \frac{1}{\gamma^2} \mathbb{E}\Big[1_T^\top \Lambda Y (Y^\top \Lambda M \Lambda Y)^{-2} Y^\top \Lambda 1_T\Big] = \frac{NT(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \frac{k_3}{\gamma^2}.$$

Moreover, $Y^\top \Lambda M \Lambda Y = T \Sigma^{-1/2} \hat{\Sigma} \Sigma^{-1/2}$, and thus by symmetry $\mathbb{E}[(Y^\top \Lambda M \Lambda Y)^{-1}]$ and $\mathbb{E}[(Y^\top \Lambda M \Lambda Y)^{-2}]$ are both proportional to the identity matrix I_N. If we denote the proportionality constants by a_1 and a_2, then

$$\text{(A79)} \qquad \frac{1}{\gamma} T \mu^\top \Sigma^{-1/2} \mathbb{E}\Big[(Y^\top \Lambda M \Lambda Y)^{-1}\Big] \Sigma^{-1/2} \mu = T a_1 \mu_p,$$

$$\text{(A80)} \qquad \frac{1}{\gamma^2} T^2 \mu^\top \Sigma^{-1/2} \mathbb{E}\Big[(Y^\top \Lambda M \Lambda Y)^{-2}\Big] \Sigma^{-1/2} \mu = T^2 a_2 \sigma_p^2,$$

which proves (44)–(45) because $k_1 = (T-N-2)a_1$ and $k_2 = \frac{(T-N-1)(T-N-2)(T-N-4)}{T-2} a_2$ by definition. Finally, it is known from Kan and Zhou (2007) that when asset returns are multivariate normally distributed, the expectations of $\tilde{\mu}_p$ and $\tilde{\sigma}_p^2$ evaluate to (44) and (45), respectively, with $k_1 = k_2 = k_3 = 1$. This completes the proof.

Proof of Proposition 8

Using the results in the proof of Proposition 7, the quantities needed in Proposition 1 to identify the optimal combination coefficients are

$$\tilde{\mu}_1 = \frac{T}{T-N-2} k_1 \theta^2, \tag{A81}$$

$$\tilde{\mu}_2 = \frac{T}{T-N-2} k_1 \frac{\theta_g^2}{\mu_g}, \tag{A82}$$

$$\tilde{\sigma}_1^2 = \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} \left(k_2 \theta^2 + k_3 \frac{N}{T} \right), \tag{A83}$$

$$\tilde{\sigma}_2^2 = \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} k_2 \frac{\theta_g^2}{\mu_g^2}, \tag{A84}$$

$$\tilde{\sigma}_{12} = \frac{T^2(T-2)}{(T-N-1)(T-N-2)(T-N-4)} k_2 \frac{\theta_g^2}{\mu_g}. \tag{A85}$$

Plugging (A81)–(A85) into (21)–(22) delivers the optimal two-fund and three-fund combination coefficients in (50)–(52). This completes the proof.

Proof of Proposition 9

The proof of this proposition is similar to that for the optimal three-fund combination coefficients in Propositions 4 and 8 when we constrain the combination coefficient on the sample mean-variance portfolio to be $c_1 = 0$.

References

Bao, Y., and A. Ullah. "The second-order bias and mean squared error of estimators in time-series models." *Journal of Econometrics*, 140 (2007), 650–669.

Bao, Y., and A. Ullah. "Higher-order bias and MSE of nonlinear estimators." *Pakistan Journal of Statistics*, 25 (2009), 287–294.

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

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

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

Muirhead, R. *Aspects of Multivariate Statistical Theory*, New Jersey: John Wiley & Sons (1982).