Optimal Portfolio Choice with Fat Tails and Parameter Uncertainty

Raymond Kan, Nathan Lassance

Journal of Financial and Quantitative Analysis, 2025

Interpreter: Weihao Zhao

School of Economics and Management WHU

2025-09-15

Motivation

Parameter uncertainty/Estimation error is core challenge in portfolio optimization

- Most combining rules assume multivariate normality, while real returns are usually fat tailed;
- Assuming normal distribution may lead to bias in estimating risk and OOS predicting.

Few exact finite sample analyses on fat tails under elliptical distribution

- El Karoui (2010, 2013) provides high-dimensional asymptotic results;
- How to derive optimal combination coeffts of two/three funds under fat tails (including finite samples and high-dimensional asymptotic) has not been fully studied.

Theoretically superior combination rules are outperformed by simple strategies

- One of the important reasons may be the neglect of the fat tail effect.

Research quesion

1. How does parameter uncertainty affect out-of-sample optimal portfolio performance under fat tails/elliptical distribution?

2. How to derive and estimate the modified two/three-fund combination rules based on this distribution?

- If classical rules are ineffective, how to adjust them to obtain a optimal combination weight?

Research objectives

- Extended the existing two-/three-fund optimal rules under elliptical distributions;
- Compared theoretical properties and performance in finite sample and asymptotic scenarios.

Contributions

1. Literature on parameter uncertainty and portfolio optimization

Prior: Most optimal combination rules assume multivariate normality.(Kan et al., 2007)

- Lack of exact finite-sample expressions under ellipses.(El Karoui., 2010, 2013)

Extend: Explicitly introducing "fat tails" into the optimal solution of combination coefficients.

- Provide both exact expressions for **finite samples** and **asymptotic** analysis to fill the gap.
- Deriving optimal two-/three-fund rules under elliptical returns in both cases.

2. Literature on impact of fat tails on out-of-sample performance loss

Prior: Performance losses associated with ignoring fat tails are small.(Tu et al., 2004)

Extend: Neglecting fat tails causes excessive coeffs. (overexposure to estimation errors).

- Partially explained why certain theories are superior but empirical evidence is inferior.

Hypothesis

H1: Ignoring fat tails underestimates the estimated risk and lead to impaired out-of-sample performance.

- Theoretical deductions are consistent with empirical evidence, showing that the growth of $\tilde{\sigma}_p$ exceeds that of $\tilde{\mu}_p$, resulting in a decrease in $U(\hat{w})$.

H2: Including fat tails in calibration significantly improves out-of-sample utility.

- Elliptical calibration significantly outperforms normal calibration in terms of annualized utility;
- Weight shrinkage (c, c_1, c_2 decrease) corresponds to an increase in utility.

Main Model Framework

Optimization problem setting: Maximizing mean-variance utility:

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

If the parameters are known, the optimal solution is: $\mathbf{w}^* = \frac{1}{\gamma}\Sigma^{-1}\mu$

In practice, sample mean and sample covariance are used to replace true values.

Parameter uncertainty (Estimation risk)

Plug in weight: $\hat{\mathbf{w}} = \hat{\Sigma}^{-1}\hat{\mu}$. **Affected by estimation errors, causes poor OOS utility**.

Two-fund rule: Combine risk-free assets with $\hat{\mathbf{w}}$: $\mathbf{w_2 f}(c) = \frac{c}{\gamma} \cdot \hat{\mathbf{w}}, \quad c \in \mathbb{R}$

Three-fund rule: Combining $\hat{\mathbf{w}}$ with the GMV portfolio $\mathbf{w_3 f}(c_1, c_2) = \frac{c_1}{\gamma}\hat{\mathbf{w}} + \frac{c_2}{\gamma}\mathbf{w}_{\text{GMV}}$

Objective: Selecting optimal c or (c_1, c_2) to maximize expected out-of-sample utility.

Fat-tail distribution

Elliptical distribution and τ representation

$$r_t = \mu + \left(\tau_t^2\right)^{1/2} Y_t, \quad Y_t \sim \mathcal{N}(0, I), \quad \mathbb{E}[\tau_t] = 1$$

- Using $\kappa = \mathbb{E}[\tau_t^2] - 1$ to measure kurtosis.
- $\tau_t$is the scaling factor, controlling for fat tails:
- $\tau_t \equiv 1 \Rightarrow$ Normal Distribution; $\kappa = 0$
- $\kappa > 0$ indicates a fat tail, which magnifies $\hat{\mu}$ and $\hat{\Sigma}$.
- When τ_t fluctuates greatly $\Rightarrow$ fat tailed, with more frequent extreme values.

In finite sample or asymptotic analysis, κ enters the second-order moment formula of sample statistics, determining how much deviation the expected out of sample mean and variance are.

Expected out-of-sample utility

$$\mathbb{E}[U(w(c))] = c \cdot (\text{expected return}) - \frac{\gamma}{2} c^2 \cdot (\text{expected variance})$$

Expected mean and **variance** is represented by κ, η, ϕ(**high-dim asymptotic**), k_1, k_2, k_3(**exact finite sample**).

Two-fund rule: Optimal coefficient $c^\star$ is similar to:

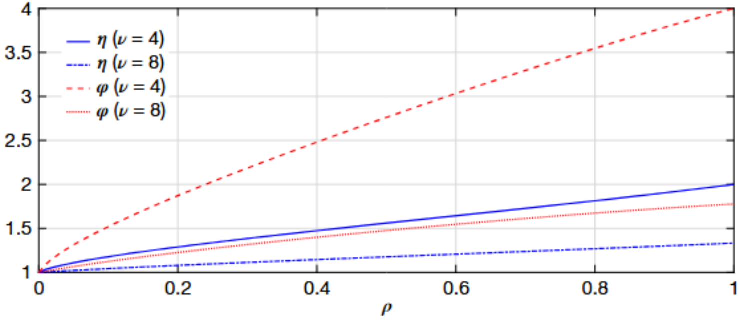
$$c^\star = \frac{\text{expected return}}{\gamma \cdot \text{expected variance}}$$

Three-fund rule: Similar logic, obtain $(c_1^\star, c_2^\star)$, The result also shows a shrinkage in the coefficient of plug-in combination and GMV.

Fat tail amplifies estimation risk $\Rightarrow$ Must be more conservative (shrink weights).
Optimal shrinkage $\Rightarrow$ Estimate statistical features of τ to obtain feasible calibration rules.

H1: Fat tail amplifies the estimation error of parameters

Either smaller ν leads to a fatter tail or larger value of ρ leads to a larger estimation error.

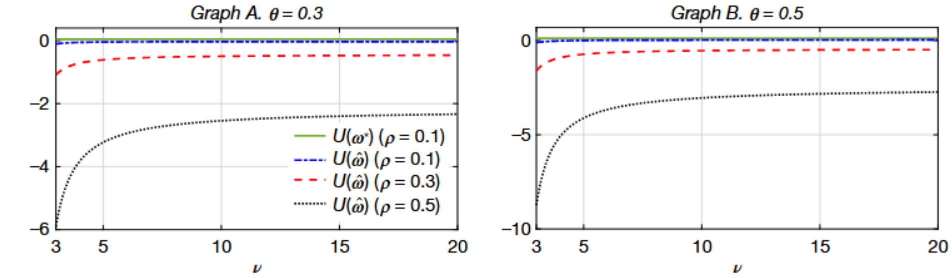


Fat tail has a greater impact on **out-of-sample variance** than on **out-of-sample mean**.

H1: Fat-tail has a significant negative impact on OOS utility $\hat{w}$.

$$\mathbb{E}[U(w(c))] = c \cdot (\text{expected return}) - \frac{\gamma}{2}c^2 \cdot (\text{expected variance})$$

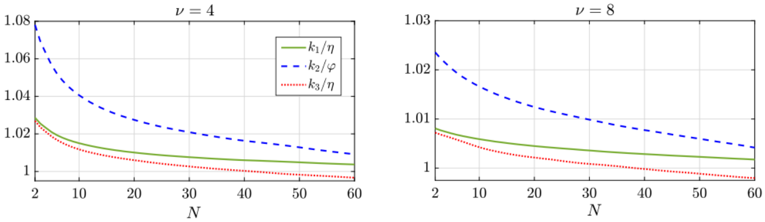
The growth of $\tilde{\sigma}_p$ exceeds that of $\tilde{\mu}_p$ resulting in a decrease in $U(\hat{w})$.



Exact Finite sample versus asymptotic

Asymptotic approximation underestimates exact value of the parameters (k_1, k_2, k_3).

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.

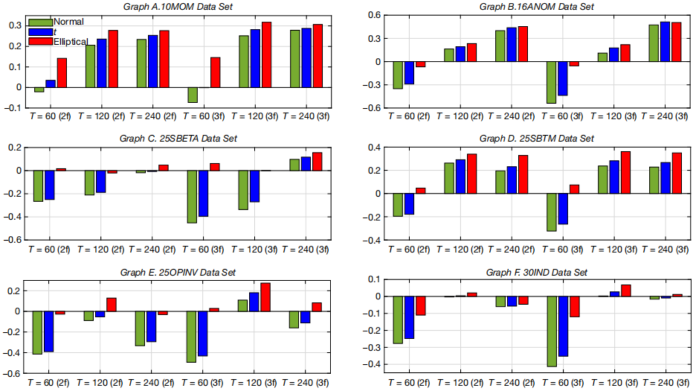


Asymptotic approx. underestimates impact of fat tails on out-of-sample performance.

H2: Including fat tails in calibration significantly improves utility

Net-of-Cost Annualized Out-of-Sample Utility of the Two-Fund and Three-Fund Rules

Figure 8 depicts, across the six data sets 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 i) the multivariate normal distribution, ii) the multivariate t-distribution whose number of degrees of freedom ν is estimated by maximum likelihood, and iii) 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.



研究局限性与未来展望

局限性（及其对结果的影响）

1. **多元椭圆分布**（肥尾）为核心建模。椭圆保有线性变换封闭性、使得均值—方差仍为最优判则。

真实数据可能呈现明显非椭圆性（例如显著偏度、非对称重尾、或更复杂的尾部分布），则基于椭圆假设得到的 η、ϕ、k 参数及最优 c 值可能出现偏误，校准效果与样本外改进可能下降。

2. **椭圆分布本质为对称分布，不能捕捉收益分布的偏斜。**在存在系统性右/左偏的情形下，均值—方差框架本身的适用性与效用评估可能受到限制，且最优收缩方向/幅度可能需不同的修正（例如偏斜性可能使均值估计误差相对更关键）。

潜在拓展

扩展到非椭圆与有偏分布（引入偏度/skewness），研究在非对称分布（如 skewed-t、混合分布或其他非椭圆族）下估计风险与最优组合系数如何变化。

Appendix 两基金规则

两基金组合权重：$\hat{w}_{2f}(c) = \frac{c}{\gamma}\,\hat{\Sigma}^{-1}\hat{\mu}$

样本外期望收益：$\tilde{\mu}_p(c) = \hat{w}_{2f}(c)^\top \mu = \frac{c}{\gamma}\,\hat{\mu}^\top\hat{\Sigma}^{-1}\mu$

取期望得：$\mathbb{E}[\tilde{\mu}_p(c)] = \frac{c}{\gamma}\,\tilde{\mu}_1$

样本外期望方差：$\tilde{\sigma}_p^2(c) = \hat{w}_{2f}(c)^\top \Sigma \hat{w}_{2f}(c) = \frac{c^2}{\gamma^2}\,\hat{\mu}^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}$

取期望得：$\mathbb{E}[\tilde{\sigma}_p^2(c)] = \frac{c^2}{\gamma^2}\,\tilde{\sigma}_1^2$

因此期望样本外效用为：

$$\mathbb{E}[U(\hat{w}_{2f}(c))] = \frac{1}{\gamma}\left(c\,\tilde{\mu}_1 - \tfrac{1}{2}c^2\,\tilde{\sigma}_1^2\right)$$

Appendix 三基金规则

三基金组合权重：$\hat{w}_{3f}(c_1, c_2) = \frac{c_1}{\gamma}\,\hat{\Sigma}^{-1}\hat{\mu} + \frac{c_2}{\gamma}\,\hat{\Sigma}^{-1}1_N$

期望收益：$\tilde{\mu}_p(c_1, c_2) = \frac{1}{\gamma}(c_1\,\hat{\mu}^\top\hat{\Sigma}^{-1}\mu + c_2\,\mu^\top\hat{\Sigma}^{-1}1_N)$

取期望：$\mathbb{E}[\tilde{\mu}_p] = \frac{1}{\gamma}(c_1\,\tilde{\mu}_1 + c_2\,\tilde{\mu}_2)$

期望方差：$\tilde{\sigma}_p^2(c_1, c_2) = \frac{1}{\gamma^2}\left(c_1^2\,\hat{\mu}^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu} + c_2^2\,1_N^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N + 2c_1c_2\,\hat{\mu}^\top\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}1_N\right)$

取期望：$\mathbb{E}[\tilde{\sigma}_p^2] = \frac{1}{\gamma^2}(c_1^2\,\tilde{\sigma}_1^2 + c_2^2\,\tilde{\sigma}_2^2 + 2c_1c_2\,\tilde{\sigma}_{12})$

因此期望样本外效用为：

$$\mathbb{E}[U(\hat{w}_{3f})] = \frac{1}{\gamma}\left(c_1\tilde{\mu}_1 + c_2\tilde{\mu}_2 - \frac{1}{2}(c_1^2\tilde{\sigma}_1^2 + c_2^2\tilde{\sigma}_2^2 + 2c_1c_2\tilde{\sigma}_{12})\right)$$

两基金最优系数

两基金的闭式解：目标等价地最大化：

$$f(c) = c\,\tilde{\mu}_1 - \tfrac{1}{2}c^2\,\tilde{\sigma}_1^2$$

一阶条件（导数为零）：

$$f'(c) = \tilde{\mu}_1 - c\,\tilde{\sigma}_1^2 = 0 \quad \Longrightarrow \quad c^\star = \frac{\tilde{\mu}_1}{\tilde{\sigma}_1^2}$$

二阶导数：

$$f''(c) = -\tilde{\sigma}_{12} < 0$$

是方差型的期望量，通常为正)，因此所得为唯一的最大值。

γ 在此过程中已全部消去（因分子与分母均含与 γ 相同的次序)，故 $c^\star$ 与 γ 无关。

三基金最优系数

目标函数：$F(c_1, c_2) = c_1\tilde{\mu}_1 + c_2\tilde{\mu}_2 - \frac{1}{2}(c_1^2\tilde{\sigma}_1^2 + c_2^2\tilde{\sigma}_2^2 + 2c_1c_2\tilde{\sigma}_{12})$

对 c_1, c_2 求偏导并令零，得到线性方程组：

$$\begin{cases} \tilde{\mu}_1 - c_1\tilde{\sigma}_1^2 - c_2\tilde{\sigma}_{12} = 0 \\ \tilde{\mu}_2 - c_1\tilde{\sigma}_{12} - c_2\tilde{\sigma}_2^2 = 0 \end{cases}$$

以矩阵形式写作：

$$\begin{bmatrix} c_1^\star \\ c_2^\star \end{bmatrix} = \frac{1}{D}\begin{bmatrix} \tilde{\sigma}_2^2 & -\tilde{\sigma}_{12} \\ -\tilde{\sigma}_{12} & \tilde{\sigma}_1^2 \end{bmatrix}\begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix}, \quad D = \tilde{\sigma}_1^2\tilde{\sigma}_2^2 - \tilde{\sigma}_{12}^2$$

$$c_1^\star = \frac{\tilde{\mu}_1\tilde{\sigma}_2^2 - \tilde{\mu}_2\tilde{\sigma}_{12}}{\tilde{\sigma}_1^2\tilde{\sigma}_2^2 - \tilde{\sigma}_{12}^2}, \quad c_2^\star = \frac{\tilde{\mu}_2\tilde{\sigma}_1^2 - \tilde{\mu}_1\tilde{\sigma}_{12}}{\tilde{\sigma}_1^2\tilde{\sigma}_2^2 - \tilde{\sigma}_{12}^2}$$

二阶偏导矩阵（Hessian）为负定（等于上述 2×2 矩阵），因此这是唯一的全局最大值。

高维渐近下的 η, ϕ

当 N, T 同阶增长（$N/T \to \rho \in (0,1)$）时，样本协方差的谱分布收敛到确定极限。

此时，$\hat{\Sigma}^{-1}$ 与 $\hat{\mu}$ 的行为不能用经典 CLT 近似，需要随机矩阵理论。

- η 与 ϕ 是从随机矩阵极限导出的两个特征量，分别反映：
- η：样本均值与真实均值的偏差在高维下的放大因子；
- ϕ：样本协方差逆矩阵在高维下的偏差因子。

在高维渐近下，所有期望量（例如 $\tilde{\mu}_1, \tilde{\sigma}_1^2$）都能写成 μ, Σ 的函数乘以 η, ϕ。

因此，η, ϕ **浓缩了高维随机矩阵偏差的全部信息**，用它们就能表达期望均值/协方差的极限形式。

有限样本精确下的 k_1, k_2, k_3

当 N 固定、$T \to \infty$ 或者 T 也不大时，可以直接从样本均值、样本协方差的分布出发（在椭圆族下可表示为缩放的 Wishart 结构）。

k_1, k_2, k_3 是 **有限样本下对二阶矩的期望精确表达**（它们是 N, T, κ 的函数）。

k_1 控制 $\mathbb{E}[\hat{\mu}^\top \hat{\Sigma}^{-1} \mu]$ 的偏差；

k_2 控制 $\mathbb{E}[\hat{\mu}^\top \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{\mu}]$ 的偏差；

k_3 控制交叉项。

这些 k_i 相当于“修正系数”，把有限样本下的抽样噪声与厚尾特征捕捉进去，使期望量能被显式写出。

Appendix

为什么能用这些参数?

因为在椭圆分布下，样本均值与协方差的二阶矩可以精确或渐近地表示为真实参数与厚尾强度的函数。于是，用 κ, η, ϕ, k_1, k_2, k_3 就能浓缩表达这些期望量。

为什么要两种设定?

高维渐近给出一般规律和直观理解，有限样本公式保证在小样本或低维情况下的精确度。二者互为补充，使得理论既简洁又可用于实证。

实证数据

论文使用 6 个实际股票收益数据集（包括月度与日度数据），具体数据集在论文表/附录中列明（例如常见为 Fama-French 风格组合、行业组合或公司组合时间序列）。

窗口长度实验：T=60、120、240（月）进行滚动窗口估计并计算样本外表现。

日度数据同时使用以测试更精细的协方差估计与更强的厚尾性（日度数据往往更厚尾）。