Supplementary Material for *Parameter-free Robust Optimization for the Maximum-Sharpe Portfolio Problem*

Deepayan Chakrabarti
Information, Risk, and Operations Management,
McCombs School of Business, Austin, TX
`deepay@utexas.edu`

December 17, 2020

1 Description of Datasets and Competing Methods

The list of datasets and competing methods are presented in Tables 1 and 2 respectively.

Figure 1 shows simulation results on the value-weighted datasets. The results are qualitatively similar to those in Figure 3 in the manuscript.

Figure 2 shows how the average one-year Sharpe ratio of AlphaRob varies with the training sample size n (results are on real-world monthly returns and not simulated returns). For comparison, we also show the results for CS, Min Var (NLS), and Min Var (L2). AlphaRob generally has the best Sharpe ratio or is close to the best. The sole exception is again the `10FFVW` dataset. We note that for small n and large p, the robust optimization of CS becomes infeasible for the parameter values suggested by Ceria and Stubbs (2006).

2 Proofs of Theorems 1 to 6.

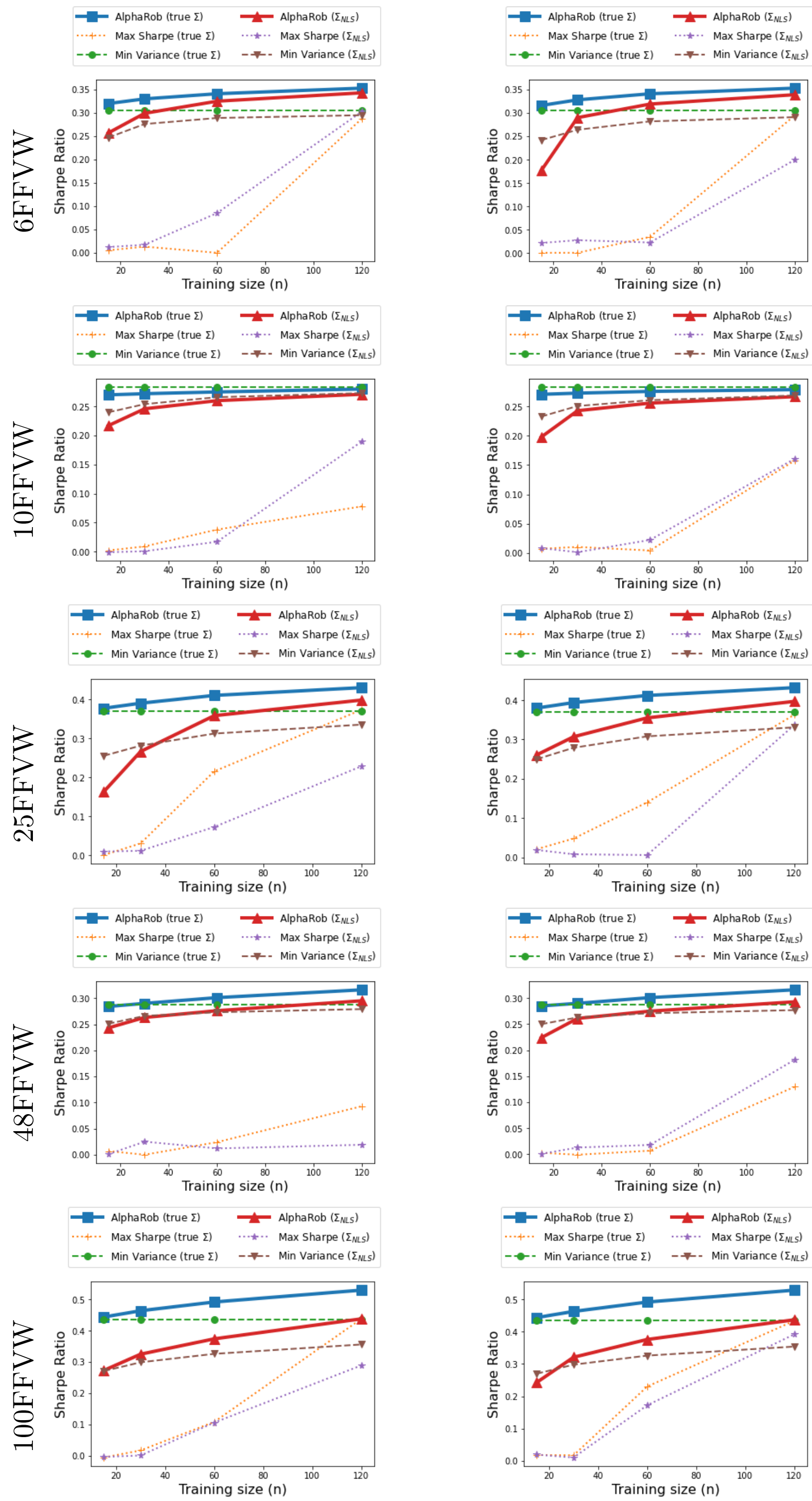
In the following, we will abbreviate $E_{\mathcal{D}_n}[.]$ to $E_n[.]$ and $Var_{\mathcal{D}_n}(.)$ to $Var_n(.)$.

Dataset	*Abbreviation*	p
Six portfolios of firms sorted by size and book-to-market	6FFEW, 6FFVW	6
Ten industry portfolios representing U.S. stock market	10FFEW, 10FFVW	10
Twenty-five portfolios of firms sorted by size and book-to-market	25FFEW, 25FFVW	25
Forty-eight industry portfolios representing U.S. stock market	48FFEW, 48FFVW	48
One hundred portfolios of firms sorted by size and book-to-market	100FFEW, 100FFVW	96
Top 200 market-value individual stocks with annual updates	200Stocks	200
Top 500 market-value individual stocks with annual updates	500Stocks	500

Table 1: *List of Datasets:* We use EW (equally-weighted) and VW (value-weighted) to indicate the corresponding weighting type in the abbreviation. For 100FFEW and 100FFVW, we ignore 4 assets with missing values in the data, so $p = 96$.

Name	**Description**
Equal Weight	Each asset has the same weight
Portfolios using robust covariance estimators	
Min Var (NLS)	Non-linear shrinkage for covariance (Ledoit and Wolf, 2012, 2017)
Min Var (L2)	Linear shrinkage for covariance (Ledoit and Wolf, 2004; DeMiguel et al., 2009)
Est. Max Sharpe (NLS)	Based on the sample mean and non-linear shrinkage of covariance
Est. Max Sharpe (L2)	Based on the sample mean and linear shrinkage of covariance
Combination portfolios (each with risk aversion $\gamma = 1$ and $\gamma = 3$)	
AA	Combines minimum-variance and mean-variance portfolios using ambiguity aversion (Garlappi et al., 2007)
EQL MV-min	Combines minimum-variance and mean-variance using quadratic-loss calibration estimated using the bootstrap (DeMiguel et al., 2013)
TZ	Combines mean-variance and equal-weight portfolios (Tu and Zhou, 2011)
Portfolios needing out-of-sample parameter fitting	
CS*	Robust portfolio to maximize worst-case expected reward (Ceria and Stubbs, 2006)
PARR*	Portfolios built from partially completed conjugate descent (DeMiguel et al., 2009)

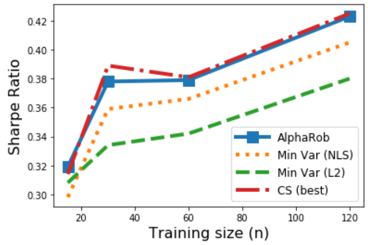
Table 2: List of competing methods: Methods whose parameters need to be set using out-of-sample data are starred.



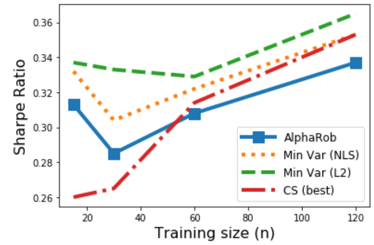
(a) Gaussian distribution

(b) t distribution

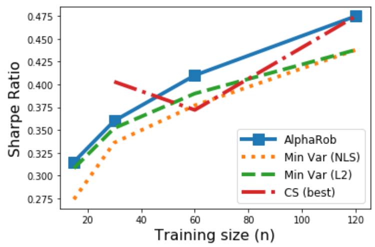
Figure 1: Sharpe ratios after simulating returns for different number of assets. The two columns show simulations using a Gaussian distribution and a heavy-tailed t-distribution respectively.



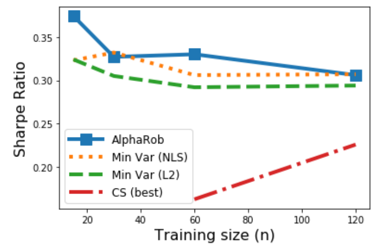
(a) 6FFVW



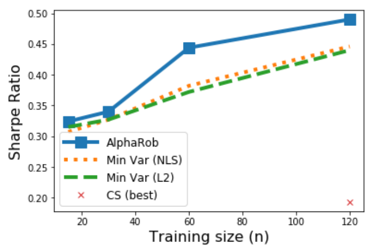
(b) 10FFVW



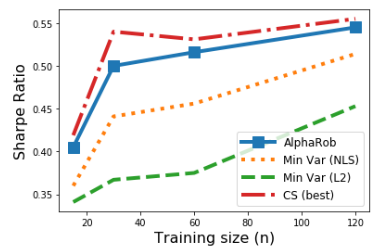
(c) 25FFVW



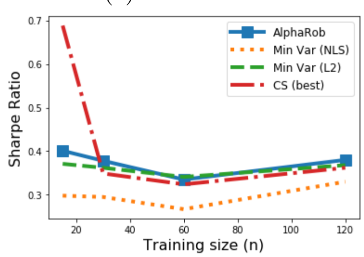
(d) 48FFVW



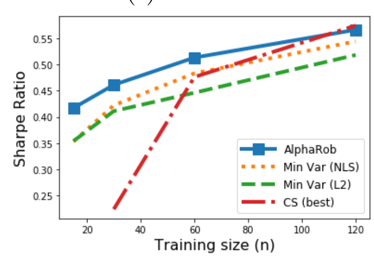
(e) 100FFVW



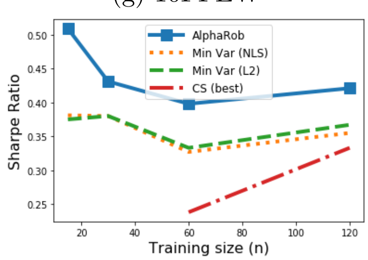
(f) 6FFEW



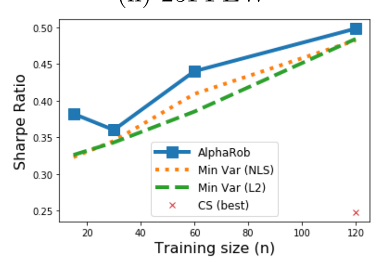
(g) 10FFEW



(h) 25FFEW



(i) 48FFEW



(j) 100FFEW

Figure 2: Average one-year Sharpe ratios as a function of the training size n. For some datasets and training sizes, the CS method gave no results and hence is not shown.

Proof of Theorem 1. From Eq. 1 and the definition of $\boldsymbol{x}$,

$$\mathcal{S}_x\left(\boldsymbol{x}\right) = \frac{E_{\mathcal{D}_n,\boldsymbol{r}}\left[\boldsymbol{w}^T\boldsymbol{r}\right]}{\sqrt{Var_{\mathcal{D}_n,\boldsymbol{r}}\left[\boldsymbol{w}^T\boldsymbol{r}\right]}} \qquad\qquad \text{where } \boldsymbol{w} = \frac{\Sigma^{-1/2}\boldsymbol{x}}{\mathbf{1}^T\Sigma^{-1/2}\boldsymbol{x}} = \frac{\Sigma^{-1/2}\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}.$$

Here $\mathcal{D}_n$ consists of n i.i.d. samples from a distribution $f(.)$, $\boldsymbol{r}$ is another independent sample drawn from $f(.)$, and $\boldsymbol{\mu}$ and Σ are the mean and covariance matrix under $f(.)$. Now,

$$E_{\mathcal{D}_n,\boldsymbol{r}}[\boldsymbol{w}^T\boldsymbol{r}] = E_nE_{\boldsymbol{r}}[\boldsymbol{w}^T\boldsymbol{r}] = E_n[\boldsymbol{w}^T\boldsymbol{\mu}] = E_n\left[\frac{\boldsymbol{x}^T\Sigma^{-1/2}\boldsymbol{\mu}}{\mathbf{1}^T\Sigma^{-1/2}\boldsymbol{x}}\right] = E_n\left[\frac{\boldsymbol{x}^T\boldsymbol{z}}{\boldsymbol{m}^T\boldsymbol{x}}\right].$$

Also,

$$\begin{aligned} Var(\boldsymbol{w}^T\boldsymbol{r}) &= E_n\left[Var_{\boldsymbol{r}}(\boldsymbol{w}^T\boldsymbol{r})\right] + Var_n\left(E_{\boldsymbol{r}}[\boldsymbol{w}^T\boldsymbol{r}]\right) \\ &= E_n\left[\boldsymbol{w}^T\Sigma\boldsymbol{w}\right] + Var_n\left(\boldsymbol{w}^T\boldsymbol{\mu}\right) \\ &= E_n\left[\frac{\boldsymbol{x}^T\boldsymbol{x}}{(\boldsymbol{m}^T\boldsymbol{x})^2}\right] + Var_n\left(\frac{\boldsymbol{x}^T\boldsymbol{z}}{\boldsymbol{m}^T\boldsymbol{x}}\right). \end{aligned}$$

The theorem statement follows. $\square$

Proof of Theorem 2. The estimated mean asset return $\hat{\boldsymbol{\mu}}$ is the average of n independent and identically distributed returns from a distribution with mean $\boldsymbol{\mu}$ and covariance Σ. So, $E\left[\hat{\boldsymbol{z}}\right] = \Sigma^{-1/2}\boldsymbol{\mu} = \boldsymbol{z}$ and $Var\left(\hat{\boldsymbol{z}}\right) = (1/n)\cdot I_{p\times p}$, where $I_{p\times p}$ is the identity matrix of that size. Since $\boldsymbol{x} = \boldsymbol{m} + \alpha\cdot\hat{\boldsymbol{z}}$, we have $E_n[\boldsymbol{x}] = \boldsymbol{m} + \alpha\cdot\boldsymbol{z} = \boldsymbol{\chi}$ and $Var_n(\boldsymbol{x}) = \frac{\alpha^2}{n}\cdot I_{p\times p}$. Now we do a Taylor expansion of the function $f(\boldsymbol{x}) = (\boldsymbol{z}^T\boldsymbol{x})/(\boldsymbol{m}^T\boldsymbol{x})$ around $\boldsymbol{x} = \boldsymbol{\chi}$ and then take expectations. After some algebraic manipulations, we get

$$E_n\left[\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right] = \frac{\boldsymbol{z}^T\boldsymbol{\chi}}{\boldsymbol{m}^T\boldsymbol{\chi}} + \frac{\alpha^3\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)}{n\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^3} + o\left(\frac{1}{n}\right).$$

Here, we use the fact that higher-order moments of $\hat{\boldsymbol{z}}$ are $o(1/n)$. By another Taylor expansion of the function $f(\boldsymbol{x}) = \left((\boldsymbol{z}^T\boldsymbol{x})/(\boldsymbol{m}^T\boldsymbol{x})\right)^2$ around $\boldsymbol{x} = \boldsymbol{\chi}$ and taking expectations, we get

$$
\begin{aligned}
E_n\left[\left(\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right)^2\right] &= \left(\frac{\boldsymbol{z}^T\boldsymbol{\chi}}{\boldsymbol{z}^T\boldsymbol{\chi}}\right)^2\left[1+\frac{\alpha^2}{n}\left\{\frac{\|\boldsymbol{z}\|^2}{(\boldsymbol{m}^T\boldsymbol{\chi})^2} - \frac{4s}{(\boldsymbol{z}^T\boldsymbol{\chi})(\boldsymbol{m}^T\boldsymbol{\chi})} + \frac{3\|\boldsymbol{m}\|^2}{(\boldsymbol{m}^T\boldsymbol{\chi})^2}\right\}\right] + o\left(\frac{1}{n}\right) \\
\Rightarrow Var_n\left(\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right) &= E_n\left[\left(\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right)^2\right] - \left(E_n\left[\frac{\boldsymbol{z}^T\boldsymbol{x}}{\boldsymbol{m}^T\boldsymbol{x}}\right]\right)^2 \\
&= \frac{\alpha^2}{n}\cdot\frac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)\|\boldsymbol{\chi}\|^2}{(\boldsymbol{m}^T\boldsymbol{\chi})^4} + o\left(\frac{1}{n}\right).
\end{aligned}
$$

A Taylor expansion of the function $f(\boldsymbol{x}) = (\boldsymbol{x}^T\boldsymbol{x})/(\boldsymbol{m}^T\boldsymbol{x})^2$ around $\boldsymbol{x} = \boldsymbol{\chi}$ yields

$$
\begin{aligned}
E_n\left[\frac{\boldsymbol{x}^T\boldsymbol{x}}{(\boldsymbol{m}^T\boldsymbol{x})^2}\right] &= \frac{\|\boldsymbol{\chi}\|^2}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2} + \frac{\alpha^2}{n\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}\left((p-4) + \frac{3\|\boldsymbol{m}\|^2\|\boldsymbol{\chi}\|^2}{(\boldsymbol{m}^T\boldsymbol{\chi})^2}\right) + o\left(\frac{1}{n}\right) \\
&= \frac{\|\boldsymbol{\chi}\|^2}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2} + \frac{\alpha^2}{n\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}\left((p-1) + \frac{3\alpha^2\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}\right) + o\left(\frac{1}{n}\right).
\end{aligned}
$$

Combining these and applying Theorem 1, we get

$$
\begin{aligned}
\mathcal{S}_x(\boldsymbol{x}) &= \frac{E_n\left[\dfrac{\boldsymbol{z}^T x}{\boldsymbol{m}^T x}\right]}{\sqrt{E_n\left[\dfrac{x'x}{(\boldsymbol{m}^T x)^2}\right] + Var_n\left(\dfrac{\boldsymbol{z}^T x}{\boldsymbol{m}^T x}\right)}} \\
&= \frac{(\boldsymbol{z}^T\boldsymbol{\chi}) + \dfrac{\alpha^3}{n}\cdot\dfrac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}}{\sqrt{\|\boldsymbol{\chi}\|^2 + \dfrac{\alpha^2}{n}\left[(p-1) + \dfrac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)\cdot\left(3\alpha^2 + \|\boldsymbol{\chi}\|^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2}\right]}} + o\left(\frac{1}{n}\right).
\end{aligned}
$$

□

Proof of Theorem 3. We will first show that

$$
\begin{aligned}
0 \leq &\quad \frac{\alpha^3}{n}\cdot\frac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2(\boldsymbol{z}^T\boldsymbol{\chi})} \quad \leq \frac{1/\theta^2 - 1}{n\|\boldsymbol{z}\|^2} \\
(1-\theta^2)\|\boldsymbol{z}\|^2 \leq &\quad \frac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)\cdot\left(3\alpha^2 + \|\boldsymbol{\chi}\|^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2} \quad \leq (1-\theta^2)\cdot\frac{\left(\|\boldsymbol{z}\|^2 + 3\right)}{\theta^2}.
\end{aligned}
$$

Using $\boldsymbol{\chi} = \boldsymbol{m} + \alpha \cdot \boldsymbol{z}$, we have

$$\|\boldsymbol{\chi}\|^2 = \|\boldsymbol{m}\|^2 + 2\alpha\theta\|\boldsymbol{m}\|\|\boldsymbol{z}\| + \alpha^2\|\boldsymbol{z}\|^2 \qquad \boldsymbol{m}^T\boldsymbol{\chi} = \|\boldsymbol{m}\|^2 + \alpha s \qquad \boldsymbol{z}^T\boldsymbol{\chi} = s + \alpha\|\boldsymbol{z}\|^2.$$

Using $\boldsymbol{m}^T\boldsymbol{z} = s = \|\boldsymbol{m}\|\|\boldsymbol{z}\|\theta$, we have

$$\frac{\alpha^3}{n} \cdot \frac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2\left(\boldsymbol{z}^T\boldsymbol{\chi}\right)} = \frac{\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2\left(1-\theta^2\right)}{n\left(s + \|\boldsymbol{m}\|^2/\alpha\right)^2\left(\|\boldsymbol{z}\|^2 + s/\alpha\right)} \in \left[0, \frac{\left(1/\theta^2 - 1\right)}{n\|\boldsymbol{z}\|^2}\right],$$

which proves the first bound. Also,

$$\begin{aligned}
\frac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right)\cdot\left(3\alpha^2 + \|\boldsymbol{\chi}\|^2\right)}{\left(\boldsymbol{m}^T\boldsymbol{\chi}\right)^2} &= \frac{\|\boldsymbol{z}\|^2(1-\theta^2)(\|\boldsymbol{m}\|^2 + 2\alpha\|\boldsymbol{m}\|\|\boldsymbol{z}\|\theta + \alpha^2(3 + \|\boldsymbol{z}\|^2)}{\left(\|\boldsymbol{m}\| + \alpha\theta\|\boldsymbol{z}\|\right)^2} \\
&= \|\boldsymbol{z}\|^2(1-\theta^2)\left(1 + \frac{\left(3 + \|\boldsymbol{z}\|^2(1-\theta^2)\right)}{\left(\|\boldsymbol{m}\|/\alpha + \theta\|\boldsymbol{z}\|\right)^2}\right) \\
&\in \left[(1-\theta^2)\|\boldsymbol{z}\|^2, (1-\theta^2)\cdot\frac{(3 + \|\boldsymbol{z}\|^2)}{\theta^2}\right],
\end{aligned}$$

proving the second bound. Using these two bounds in Theorem 2 followed by algebraic manipulations yields the desired result. □

Proof of Theorem 4. The numerator of Eq. 9 does not depend on α. We will show that the denominator has a single minimum at $\alpha = \alpha^\star$. The denominator equals $1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)$ (Eq. 10):

$$\begin{aligned}
q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right) &= \frac{\left(\|\boldsymbol{m}\|^2\|\boldsymbol{z}\|^2 - s^2\right) + \alpha^2(p/n)\|\boldsymbol{z}\|^2}{\left(s + \alpha\|\boldsymbol{z}\|^2\right)^2} \\
&= \frac{\left(1-\theta^2\right) + \alpha^2 A}{\left(\theta + \alpha B\right)^2} \qquad \text{where } A = \frac{p}{n\|\boldsymbol{m}\|^2},\ B = \frac{\|\boldsymbol{z}\|}{\|\boldsymbol{m}\|}.
\end{aligned} \tag{1}$$

Taking the derivative with respect to α,

$$\frac{d}{d\alpha} q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right) = 2 \cdot \frac{\alpha A\theta - \left(1-\theta^2\right)B}{\left(\theta + \alpha B\right)^3}. \tag{2}$$

Since $\theta > 0$, $\alpha \geq 0$, and $B \geq 0$, this has a single inflection point at $\alpha = \alpha^\star$ as defined in the theorem statement. Also, the derivative is positive for $\alpha > \alpha^\star$ and negative for $\alpha < \alpha^\star$, so this is a minimum. The theorem follows. $\square$

Proof of Theorem 5. We have $E[\hat{\boldsymbol{z}}] = \boldsymbol{z}$ and $Var(\hat{\boldsymbol{z}}) = (1/n) \cdot I_{p\times p}$. By a Taylor expansion of the function $f(\boldsymbol{x}) = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$ around $\boldsymbol{x} = \boldsymbol{z}$ and taking expectations, we get

$$\begin{aligned}
E\|\hat{\boldsymbol{z}}\| &= \|\boldsymbol{z}\| + \frac{1}{2} \cdot E\left[(\hat{\boldsymbol{z}} - \boldsymbol{z})^T \frac{1}{\|\boldsymbol{z}\|}\left(I_{p\times p} - \frac{1}{\|\boldsymbol{z}\|^2} \cdot \boldsymbol{z}\boldsymbol{z}^T\right)(\hat{\boldsymbol{z}} - \boldsymbol{z})\right] + o\left(\frac{1}{n}\right) \\
&= \|\boldsymbol{z}\| + \frac{1}{2\|\boldsymbol{z}\|} \cdot Trace\left[\left(I_{p\times p} - \frac{1}{\|\boldsymbol{z}\|^2} \cdot \boldsymbol{z}\boldsymbol{z}^T\right) \cdot Var\left(\hat{\boldsymbol{z}}\right)\right] + o\left(\frac{1}{n}\right) \\
&= \|\boldsymbol{z}\| + \frac{p-1}{2n\|\boldsymbol{z}\|} + o\left(\frac{1}{n}\right).
\end{aligned}$$

Similarly, taking a Taylor expansion of the function $f(\boldsymbol{x}) = (\boldsymbol{m}^T\boldsymbol{x})/(\|\boldsymbol{m}\|\|\boldsymbol{x}\|)$ around $\boldsymbol{x} = \boldsymbol{z}$ and taking expectations, we get

$$E\hat{\theta} = \theta - \frac{(p-1) \cdot \boldsymbol{m}^T\boldsymbol{z}}{2n\|\boldsymbol{m}\|\|\boldsymbol{z}\|^3} + o\left(\frac{1}{n}\right) = \theta\left(1 - \frac{p-1}{2n\|\boldsymbol{z}\|^2}\right) + o\left(\frac{1}{n}\right).$$

This proves the theorem. $\square$

Proof of Theorem 6. By Eq. 11, any element in $\{\|\boldsymbol{z}\| \mid (\|\boldsymbol{z}\|, \theta_\gamma) \in \Gamma_\gamma\}$ must satisfy $\|\boldsymbol{z}\| = \gamma p \cdot (n\|\boldsymbol{m}\| \cdot (1/\theta_\gamma - \theta_\gamma))^{-1}$, which proves the first statement.

Define $f_\gamma(\alpha, \theta) = h\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)$ as a function of α and θ, with $\|\boldsymbol{z}\|$ fixed to the unique value that gives $\alpha^\star(\|\boldsymbol{z}\|, \theta) = \gamma$. Also, let $g_\gamma(\alpha, \theta) = \frac{d}{d\theta}\log\left(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)\right)$. The second statement of the theorem will be proved if we show that $\frac{d}{d\theta}f_\gamma(\alpha, \theta) \geq 0$ for any $\alpha \geq 0$. Now,

$$\frac{d}{d\theta}f_\gamma(\alpha, \theta) \geq 0 \Leftrightarrow \frac{d}{d\theta}\log\left(1 + q\left(\gamma \mid \|\boldsymbol{z}\|, \theta\right)\right) - \frac{d}{d\theta}\log\left(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)\right) \geq 0 \Leftrightarrow g_\gamma(\gamma, \theta) \geq g_\gamma(\alpha, \theta). \tag{3}$$

To show Eq. 3, we will show that for any fixed θ, $\alpha = \gamma$ is a maximum of $g_\gamma(\alpha, \theta)$.

First, observe that

$$\begin{aligned}
\frac{d}{d\alpha} g_\gamma(\alpha,\theta) &= \frac{d^2}{d\alpha d\theta} \log\left(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)\right) \\
&= \frac{d}{d\theta} \frac{2(\alpha-\gamma)A\theta}{\left(\theta+\alpha B\right)^3 \left(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)\right)} \\
&= 2(\alpha-\gamma)A \cdot \frac{d}{d\theta}\left(\frac{\theta}{\left(\theta+\alpha B\right)^3 \left(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)\right)}\right) \qquad (4)
\end{aligned}$$

where the second equality is obtained from Eqs. 2 and 11. Note that B is a function of $\|\boldsymbol{z}\|$, which itself is a function of γ and θ. Thus, $\left[\frac{d}{d\alpha} g_\gamma(\alpha,\theta)\right]_{\alpha=\gamma} = 0$, so $g_\gamma(\alpha,\theta)$ has a mode at $\alpha=\gamma$ for any θ. To show that this mode is a maximum, we will show that $\frac{d}{d\alpha} g_\gamma(\alpha,\theta)$ is positive for $0<\alpha<\gamma$ and negative for $\alpha>\gamma$.

Let us fix $\alpha>0$, and use $q(\theta) = q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)$ and $B(\theta)$ to denote the fact that $q\left(. \mid \|\boldsymbol{z}\|, \theta\right)$ and $B(.)$ are functions of θ. Let $C(\theta) = \left(\theta+\alpha B(\theta)\right)^3 \left(1+q(\theta)\right)$. Continuing from Eq. 4,

$$\begin{aligned}
\frac{d}{d\theta}\left(\frac{\theta}{\left(\theta+\alpha B\right)^3\left(1+q(\theta)\right)}\right) &= \left(\frac{\theta+\alpha B(\theta)}{C(\theta)}\right)^2 D(\theta) \\
\text{where } D(\theta) = (\theta+\alpha B(\theta))(1+q(\theta)) &- 3\theta(1+q(\theta)) \cdot \frac{d}{d\theta}(\theta+\alpha B(\theta)) - \theta(\theta+\alpha B(\theta)) \cdot \frac{d}{d\theta} q(\theta)
\end{aligned}$$

Now, by Eqs. 11 and 1, we have

$$\begin{aligned}
B(\theta) &= \gamma A\left(\frac{\theta}{1-\theta^2}\right) \\
\Rightarrow \frac{d}{d\theta}\left(\theta+\alpha\cdot B(\theta)\right) &= 1+\gamma\alpha A \cdot \frac{1+\theta^2}{\left(1-\theta^2\right)^2} = 1+\frac{\alpha B(\theta)}{\theta}\left(1+\frac{2\theta^2}{1-\theta^2}\right) \qquad (5) \\
\frac{d}{d\theta} q(\theta) &= \frac{d}{d\theta}\left(\frac{1-\theta^2+\alpha^2 A}{\left(\theta+\alpha B(\theta)\right)^2}\right) = -\frac{2\theta}{\left(\theta+\alpha B(\theta)\right)^2} - 2\left(\frac{q(\theta)}{\theta+\alpha B(\theta)}\right)\cdot\frac{d}{d\theta}\left(\theta+\alpha B(\theta)\right). \qquad (6)
\end{aligned}$$

Using these in the formula for $D(\theta)$, we have

$$
\begin{aligned}
D(\theta) &= (\theta + \alpha B(\theta))(1 + q(\theta)) + \frac{2\theta^2}{\theta + \alpha B(\theta)} - \theta\,(3 + q(\theta)) \cdot \frac{d}{d\theta}\,(\theta + \alpha B(\theta)) \\
&= (1 + q(\theta)) \left(\theta + \alpha B(\theta) - \theta \cdot \frac{d}{d\theta}(\theta + \alpha B(\theta)) \right) + \frac{2\theta^2}{\theta + \alpha B(\theta)} - 2\theta \cdot \frac{d}{d\theta}(\theta + \alpha B(\theta)) \\
&= (1 + q(\theta)) \left(-\frac{2\alpha\theta^2 B(\theta)}{1 - \theta^2} \right) + \frac{2\theta^2}{\theta + \alpha B(\theta)} - 2\theta - 2\alpha B(\theta) \left(1 + \frac{2\theta^2}{1 - \theta^2} \right) \\
&= -3 \left(\frac{\alpha\theta^2 B(\theta)}{1 - \theta^2} \right) (3 + q(\theta)) - 2 \left(\frac{\alpha B(\theta)(2\theta + \alpha B(\theta))}{\theta + \alpha B(\theta)} \right) \\
&< 0,
\end{aligned}
$$

where the first equality follows from Eq. 6, the third equality from Eq. 5, and the inequality from $q(\theta) > 0$, since $\theta \leq 1$ and $\alpha > 0$. Using this in Eq. 4 shows that $\frac{d}{d\alpha} g_\gamma(\alpha, \theta)$ is positive for $0 < \alpha < \gamma$ and negative for $\alpha > \gamma$. This completes the proof of the second statement. $\square$

3 Proof of Theorem 7.

We first prove several helpful lemmas. Define $u^{(t)}$ and $v^{(t)}$ to be the minimum and maximum of $h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)$ and $h\left(\alpha^{(t)} \mid \gamma_{hi}^{(t+1)}\right)$ (see Figure 2). Note that $\kappa^{(t)}$, $\kappa^{(t+1)}$, $u^{(t)}$, and $v^{(t)}$ are all less than or equal to zero.

Lemma 1. *If $\theta_- > 0$, $\|\boldsymbol{z}\|_+ < \infty$, and $0 \leq \alpha \leq \max(\mathcal{I}_\gamma)$, there exist constants $C_1 > 0$ and $C_2 > 0$ that depend on $\mathcal{I}_{\|\boldsymbol{z}\|}$ and $\mathcal{I}_\theta$ such that $C_1 \cdot |\alpha - \gamma| \leq \left| \frac{d}{d\alpha} h\left(\alpha \mid \gamma\right) \right| \leq C_2$.*

Proof.

$$
\frac{d}{d\alpha} h\left(\alpha \mid \gamma\right) = \frac{d}{d\alpha} \log \left(\frac{1 + q\left(\gamma \mid \|\boldsymbol{z}\|, \theta\right)}{1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)} \right) = -\frac{2(\alpha - \gamma)A\theta}{\left(1 + q\left(\alpha \mid \|\boldsymbol{z}\|, \theta\right)\right)\left(\theta + \alpha B\right)^3}.
$$

The desired bounds follows from $\theta \geq \theta_- > 0$, $\|\boldsymbol{z}\| \leq \|\boldsymbol{z}\|_+ < \infty$, and $0 \leq \alpha \leq \max(\mathcal{I}_\gamma) = \alpha^\star(\|\boldsymbol{z}\|_+, \theta_-) < \infty$. $\square$

Lemma 2. *We have:*

$$\forall t',\ \kappa^{(t')} \geq \kappa_{rob}, \qquad u^{(t)} \leq \kappa_{rob}, \qquad \kappa^{(t+1)} \leq v^{(t)} \leq \kappa^{(t)}.$$

Proof. First we will show that $\kappa^{(t')} \geq \kappa_{rob}$. Suppose this is not the case. Then, $\alpha^{(t')} \neq \alpha_{rob}$ and $h\left(\alpha^{(t')} \mid \gamma_{lo}^{(t')}\right) = h\left(\alpha^{(t')} \mid \gamma_{hi}^{(t')}\right) = \kappa^{(t')} < \kappa_{rob}$. If $\alpha^{(t')} < \alpha_{rob}$, then $h\left(\alpha_{rob} \mid \gamma_{lo}^{(t')}\right) < h\left(\alpha^{(t')} \mid \gamma_{lo}^{(t')}\right) < \kappa_{rob}$, where the first inequality follows from unimodality of $h\left(\alpha \mid \gamma_{lo}^{(t')}\right)$ (Theorem 4) and the fact that it achieves its maximum at $\gamma_{lo}^{(t')} < \alpha^{(t')} < \alpha_{rob}$. But by definition, $\kappa_{rob} = \min_\gamma h\left(\alpha_{rob} \mid \gamma\right)$. This leads to a contradiction. A similar argument holds when $\alpha^{(t')} > \alpha_{rob}$. Hence, $\kappa^{(t')} \geq \kappa_{rob}$.

If $u^{(t)} > \kappa_{rob}$, then $v^{(t)} > u^{(t)} > \kappa_{rob}$, by the definition of $u^{(t)}$ and $v^{(t)}$. By the definitions of $\gamma_{lo}^{(t+1)}$ and $\gamma_{hi}^{(t+1)}$ in Algorithm 1, the above statement implies that $\min_\gamma h\left(\alpha^{(t)} \mid \gamma\right) > \kappa_{rob}$ (recall that $h\left(\alpha \mid \gamma\right) \leq 0$ always). This contradicts the definition of α_{rob} and κ_{rob}. Hence, $u^{(t)} \leq \kappa_{rob}$.

Now suppose, without loss of generality, that the maximum of $h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)$ and $h\left(\alpha^{(t)} \mid \gamma_{hi}^{(t+1)}\right)$ was achieved by the former. Then, from the definition of $\gamma_{lo}^{(t+1)}$, we have $v^{(t)} = \min_{\gamma \leq \alpha^{(t)}} h\left(\alpha^{(t)} \mid \gamma\right) \leq h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t)}\right) = \kappa^{(t)}$. Also, $\kappa^{(t+1)}$ must be between $u^{(t)}$ and $v^{(t)}$, since $\kappa^{(t+1)}$ is achieved at the intersection of the curves indexed by $\gamma_{lo}^{(t+1)}$ and $\gamma_{hi}^{(t+1)}$. Hence, we must have $v^{(t)} \geq \kappa^{(t+1)}$. This proves the lemma. □

Lemma 3. *Under the conditions of Lemma 1, for any feasible α, α', and γ, we have*

$$|\alpha - \alpha'| \geq \frac{|h\left(\alpha \mid \gamma\right) - h\left(\alpha' \mid \gamma\right)|}{C_2}.$$

In particular,

$$\min\left(|\alpha^{(t)} - \gamma_{lo}^{(t+1)}|, |\alpha^{(t)} - \gamma_{hi}^{(t+1)}|\right) \geq \frac{|v^{(t)}|}{C_2}, \qquad |\alpha^{(t+1)} - \alpha^{(t)}| \geq \frac{v^{(t)} - u^{(t)}}{2 \cdot C_2}.$$

Proof. The first statement follows from $|h\left(\alpha \mid \gamma\right) - h\left(\alpha' \mid \gamma\right)| = \left|\int_\alpha^{\alpha'} \frac{d}{dx} h\left(x \mid \gamma\right) dx\right| \leq C_2 \cdot |\alpha - \alpha'|$, where we used Lemma 1 in the inequality. Applying this twice, we find

$$
\begin{aligned}
|\alpha^{(t)} - \gamma_{lo}^{(t+1)}| &\geq \frac{|h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right) - h\left(\gamma_{lo}^{(t+1)} \mid \gamma_{lo}^{(t+1)}\right)|}{C_2} = \frac{|h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)|}{C_2}, \\
|\alpha^{(t)} - \gamma_{hi}^{(t+1)}| &\geq \frac{|h\left(\alpha^{(t)} \mid \gamma_{hi}^{(t+1)}\right)|}{C_2}, \\
\Rightarrow \min\left(|\alpha^{(t)} - \gamma_{lo}^{(t+1)}|, |\alpha^{(t)} - \gamma_{hi}^{(t+1)}|\right) &\geq \frac{\min\left(|h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)|, |h\left(\alpha^{(t)} \mid \gamma_{hi}^{(t+1)}\right)|\right)}{C_2} = \frac{|v^{(t)}|}{C_2},
\end{aligned}
$$

proving the second statement. Repeating this argument, we see

$$
\begin{aligned}
|\alpha^{(t+1)} - \alpha^{(t)}| &\geq \frac{|h\left(\alpha^{(t+1)} \mid \gamma_{lo}^{(t+1)}\right) - h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)|}{C_2} = \frac{v^{(t)} - \kappa^{(t+1)}}{C_2}, \\
\text{and } |\alpha^{(t+1)} - \alpha^{(t)}| &\geq \frac{|h\left(\alpha^{(t+1)} \mid \gamma_{hi}^{(t+1)}\right) - h\left(\alpha^{(t)} \mid \gamma_{hi}^{(t+1)}\right)|}{C_2} = \frac{\kappa^{(t+1)} - u^{(t)}}{C_2},
\end{aligned}
$$

where we assumed without loss of generality that $v^{(t)} = h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)$, and we used the fact that $u^{(t)} \leq \kappa_{rob} \leq \kappa^{(t+1)} \leq v^{(t)}$ (Lemma 2). Summing these two statements, $|\alpha^{(t+1)} - \alpha^{(t)}| \geq \frac{v^{(t)} - u^{(t)}}{2 \cdot C_2}$. $\square$

Proof of Theorem 7. The fact that $\kappa_{rob} \leq \kappa^{(t')} \leq \kappa^{(1)} \leq 0$ follows from Lemma 2. Since the curves are unimodal (Theorem 4), $\kappa^{(1)} = 0$ iff $\gamma_{lo}^{(1)} = \gamma_{hi}^{(1)}$. By construction, this happens only if the interval $\mathcal{I}_\gamma$ is degenerate. Now assume, without loss of generality, that $v^{(t)} = h\left(\alpha^{(t)} \mid \gamma_{lo}^{(t+1)}\right)$. This implies that $\alpha^{(t+1)} \geq \alpha^{(t)}$, since $\alpha^{(t+1)}$ is the intersection of the curves indexed by $\gamma_{lo}^{(t+1)}$ and

$\gamma_{hi}^{(t+1)}$ (see Figure 2). Then,

$$\begin{aligned}
\kappa^{(t+1)} &= \kappa^{(t)} + \int_{\alpha^{(t)}}^{\alpha^{(t+1)}} \frac{d}{d\alpha} h\left(\alpha \mid \gamma_{lo}^{(t+1)}\right) \\
&\leq \kappa^{(t)} - |\alpha^{(t)} - \alpha^{(t+1)}| \cdot C_1 \cdot |\alpha^{(t)} - \gamma_{lo}^{(t+1)}| && \text{(Lemma 1)} \\
&\leq \kappa^{(t)} - \left(\frac{v^{(t)} - u^{(t)}}{2 \cdot C_2}\right) \cdot C_1 \cdot \frac{|v^{(t)}|}{C_2} && \text{(Lemma 3)} \\
&\leq \kappa^{(t)} \left(1 + \frac{C_1 \cdot \left(v^{(t)} - u^{(t)}\right)}{2 \cdot C_2^2}\right) && (|v^{(t)}| = -v^{(t)} \geq -\kappa^{(t)} \text{ by Lemma 2)} \\
\Rightarrow |\kappa^{(t+1)}| &\geq |\kappa^{(t)}| \left(1 + \frac{C_1 \cdot \left(v^{(t)} - u^{(t)}\right)}{2 \cdot C_2^2}\right) \\
&\geq |\kappa^{(t)}| \left(1 + \frac{C_1 \cdot \left(|\kappa_{rob}| - |v^{(t)}|\right)}{2 \cdot C_2^2}\right) && (u^{(t)} \leq \kappa_{rob} \text{ by Lemma 2)}
\end{aligned}$$

This shows that $|\kappa^{(t+1)}| \geq |\kappa^{(t)}|$, since $|\kappa_{rob}| \geq |\kappa^{(t+1)}| \geq |v^{(t)}|$ by Lemma 2. But, by Lemma 2, we also have $\kappa^{(t+1)} \leq v^{(t)}$. So,

$$\begin{aligned}
|\kappa^{(t+1)}| &\geq \max\left(|v^{(t)}|, |\kappa^{(t)}| \left(1 + C_3 \cdot \left(|\kappa_{rob}| - |v^{(t)}|\right)\right)\right) && \left(C_3 = \frac{C_1}{2 \cdot C_2^2}\right) \\
&\geq |\kappa^{(t)}| \left(\frac{1 + C_3 \cdot |\kappa_{rob}|}{1 + C_3 \cdot |\kappa^{(t)}|}\right),
\end{aligned}$$

where the second inequality is achieved by setting $v^{(t)}$ to equalize both terms in the maximum. Hence,

$$\begin{aligned}
|\kappa_{rob}| - |\kappa^{(t+1)}| &\leq \frac{|\kappa_{rob}| - |\kappa^{(t)}|}{1 + C_3 \cdot |\kappa^{(t)}|} \\
&\leq \frac{|\kappa_{rob}| - |\kappa^{(t)}|}{1 + C_3 \cdot |\kappa^{(1)}|} && \text{since } |\kappa^{(t+1)}| \geq |\kappa^{(t)}|.
\end{aligned}$$

The formula in the theorem statement follows easily. □

4 Combination Portfolios under the Mean-Variance Utility.

This common utility function linearly combines the mean portfolio return and its variance:

$$\mathbb{U}_w(\boldsymbol{w}) = E[\boldsymbol{w}^T \boldsymbol{r}] - \gamma \cdot Var(\boldsymbol{w}^T \boldsymbol{r}), \tag{7}$$

where the parameter $\gamma > 0$ represents the risk-reward tradeoff.

Let $\boldsymbol{x}$ be defined as in Section 3, and let $\mathbb{U}_x(\boldsymbol{x})$ refer to the expected utility of the portfolio $\boldsymbol{w}$ constructed from $\boldsymbol{x}$.

Theorem 1.

$$\begin{aligned}\mathbb{U}_x(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}) = &\left(\frac{\boldsymbol{z}^T \boldsymbol{\chi}}{\boldsymbol{m}^T \boldsymbol{\chi}}\right) \left[1 + \frac{\alpha^3}{n} \cdot \frac{\left(\|\boldsymbol{m}\|^2 \|\boldsymbol{z}\|^2 - s^2\right)}{\left(\boldsymbol{m}^T \boldsymbol{\chi}\right)^2 \left(\boldsymbol{z}^T \boldsymbol{\chi}\right)}\right] \\ &- \frac{\gamma}{(\boldsymbol{m}^T \boldsymbol{\chi})^2} \cdot \left[\|\boldsymbol{\chi}\|^2 + \frac{\alpha^2}{n}\left[(p-1) + \frac{\left(\|\boldsymbol{m}\|^2 \|\boldsymbol{z}\|^2 - s^2\right) \cdot \left(3\alpha^2 + \|\boldsymbol{\chi}\|^2\right)}{\left(\boldsymbol{m}^T \boldsymbol{\chi}\right)^2}\right]\right] + o\left(\frac{1}{n}\right),\end{aligned}$$

where $\boldsymbol{\chi} = \boldsymbol{m} + \alpha \cdot \boldsymbol{z}$.

Under the conditions of Theorem 3, this simplifies to

$$\mathbb{U}_x(\boldsymbol{m} + \alpha \cdot \hat{\boldsymbol{z}}) \approx \frac{\boldsymbol{z}^T \boldsymbol{\chi}}{\boldsymbol{m}^T \boldsymbol{\chi}} - \gamma \cdot \frac{\|\boldsymbol{\chi}\|^2 + \alpha^2 \cdot \frac{p}{n}}{(\boldsymbol{m}^T \boldsymbol{\chi})^2}.$$

The optimum is attained at

$$\alpha^\star_{MV} = \begin{cases} \frac{\|\boldsymbol{m}\|^2}{2\gamma\left(1 + \frac{p}{n} \cdot \left(\frac{\|\boldsymbol{m}\|^2}{\|\boldsymbol{m}\|^2 \|\boldsymbol{z}\|^2 - s^2}\right)\right) - s}, & \text{if } \gamma \left(1 + \frac{p}{n} \cdot \left(\frac{\|\boldsymbol{m}\|^2}{\|\boldsymbol{m}\|^2 \|\boldsymbol{z}\|^2 - s^2}\right)\right) \geq \frac{s}{2} \\ \infty, & \text{otherwise} \end{cases}$$

This behaves quite differently from the $\alpha^\star$ for the maximum-Sharpe portfolios (Eq. 11). The maximum-Sharpe $\alpha^\star$ grows linearly with n. However, for the expected utility in Eq. 7, there is a threshold behavior. When $\gamma \geq s/2$, $\alpha^\star_{MV}$ converges to $\|\boldsymbol{m}\|^2/(2\gamma - s)$ from below as n increases. When $\gamma < s/2$, $\alpha^\star_{MV}$ increases as n increases towards a threshold value, and jumps to ∞ above this threshold. This threshold is given by

$$n_{thresh} = \frac{2\gamma p}{(s-2\gamma)\|\boldsymbol{z}\|^2(1-\theta^2)},$$

where $\theta = s/(\|\boldsymbol{m}\|\|\boldsymbol{z}\|)$.

This can be intuitively explained by viewing Eq. 7 as the Lagrangian of the “mean-variance” optimization problem:

$$\text{maximize } E[\boldsymbol{w}^T\boldsymbol{r}] \text{ subject to } Var(\boldsymbol{w}^T\boldsymbol{r}) \leq \rho, \boldsymbol{w}^T\mathbf{1} = 1.$$

Larger values of γ correspond to tighter constraints on the variance (i.e., smaller ρ), and reduces the achievable mean return. Thus, for a large enough γ, the optimal α_{opt} converges to a constant with growing n. However, if γ is too small, the variance constraint is not active and the mean portfolio continuously grows with $\alpha^\star_{MV}$ as long as the expected return of each asset can be estimated relatively accurately (i.e., large enough n). For small n, uncertainty about the expected asset returns limits the achievable return of the overall portfolio.

References

S. Ceria and R. A. Stubbs. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management; London*, 7(2):109–127, July 2006.

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

V. DeMiguel, A. Martin-Utrera, and F. J. Nogales. Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking & Finance*, 37(8):3018–3034, 2013.

L. Garlappi, R. Uppal, and T. Wang. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81, 2007.

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

O. Ledoit and M. Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 04 2012.

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

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