Estimation of Out-of-Sample Sharpe Ratio for High Dimensional Portfolio Optimization

Xuran Meng

Joint work with Yuan Cao and Weichen Wang

University of Michigan

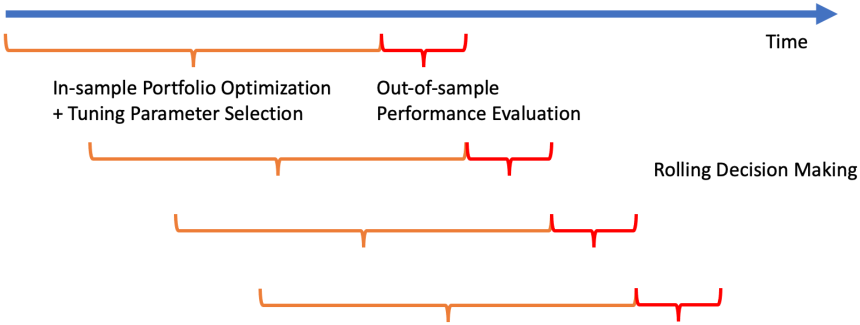
July 8, 2025

Outline

1. Introduction
2. Estimating OOS Sharpe
3. Extension 1: Estimating Efficient Frontier
4. Extension 2: Optimization Over Regularization
5. Extension 3: Estimating OOS Sharpe with Sample Mean
6. Real Data Experiments
7. Conclusion

Portfolio Management

Portfolio Management



Mean-Variance Portfolio

Definition (Mean-variance portfolio)

Given p risky assets with mean $\mathbf{r} \in \mathbb{R}^p$ and covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$, the mean-variance portfolio optimizes the allocation vector $\mathbf{w}$:

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^p} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \quad \text{s. t.} \quad \mathbf{w}^\top \boldsymbol{\mu} = \mu_0.$$

Here, we denote by $\boldsymbol{\mu} = \mathbf{r} - r_0 \mathbf{1}$ the excess return of the risky assets, and $\mu_0 > 0$ is the targeted excess return of the portfolio.

The solution is $\mathbf{w}^* \propto \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$.
We assume $\boldsymbol{\mu}$ is known and $\boldsymbol{\Sigma}$ needs to be estimated.

Ridge Regularized Mean-Variance Portoflio

ℓ_2-Regularized-MV: Consider the optimization with regularization $\mathbf{Q}$:

$$\mathbf{w}^* = \text{argmin}_{\mathbf{w}\in\mathbb{R}^p}\, \mathbf{w}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})\mathbf{w} \quad \text{s. t.} \quad \mathbf{w}^\top\boldsymbol{\mu} = \mu_0,$$

where $\mathbf{Q}$ is positive definite. The optimal $\mathbf{w}^*$ satisfies

$$\mathbf{w}^* \propto (\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\mu}.$$

Ridge Regularized Mean-Variance Portoflio

ℓ_2-Regularized-MV: Consider the optimization with regularization $\mathbf{Q}$:

$$\mathbf{w}^* = \mathrm{argmin}_{\mathbf{w}\in\mathbb{R}^p}\, \mathbf{w}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})\mathbf{w} \quad \mathrm{s.\,t.} \quad \mathbf{w}^\top\boldsymbol{\mu} = \mu_0,$$

where $\mathbf{Q}$ is positive definite. The optimal $\mathbf{w}^*$ satisfies

$$\mathbf{w}^* \propto (\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\mu}.$$

OOS Sharpe Ratio of $\mathbf{w}^*$:

$$SR(\mathbf{Q}) = \frac{\mathbb{E}_{\widetilde{\mathbf{R}}}[\mathbf{w}^{*\top}(\widetilde{\mathbf{R}} - r_0\mathbf{1})]}{\sqrt{\mathrm{Var}_{\widetilde{\mathbf{R}}}[\mathbf{w}^{*\top}(\widetilde{\mathbf{R}} - r_0\mathbf{1})]}} = \frac{\boldsymbol{\mu}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\mu}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\mu}}},$$

where $\widetilde{\mathbf{R}}$ is an out-of-sample point with mean $\mathbf{r} = \boldsymbol{\mu} + r_0\mathbf{1}$ and cov $\boldsymbol{\Sigma}$.

Open Problems

★*Is it possible to estimate the out-of-sample Sharpe ratio with some regularization using in-sample data?*

★*Can we then optimize the estimator over the regularization parameter to enhance the out-of-sample Sharpe ratio?*

In-sample Optimism

How about we just use $\widehat{\mathbf{\Sigma}}$ to estimate $\mathbf{\Sigma}$ in $SR(\mathbf{Q})$?

- Assume n iid p-dim return vectors $\boldsymbol{R}_i \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}), i = 1, \ldots, n$, $\boldsymbol{\mu}$ is known and $p/n \to c < 1$.
- The optimized MV portfolio with $\mathbf{Q} = \mathbf{0}$ is based on the sample covariance $\widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{R}_i - \boldsymbol{\mu})(\boldsymbol{R}_i - \boldsymbol{\mu})^\top$, i.e. $\mathbf{w}^* \propto \widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}$.
- Its in-sample SR is $\sqrt{\boldsymbol{\mu}^\top\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}$ while its out-of-sample SR is $\boldsymbol{\mu}^\top\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}/\sqrt{\boldsymbol{\mu}^\top\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}$.
- Our theorem will show the latter is approximately $(1-c)\sqrt{\boldsymbol{\mu}^\top\widehat{\mathbf{\Sigma}}^{-1}\boldsymbol{\mu}}$, so the in-sample SR is $1/(1-c)$ times larger.
- When c is close to 1, the portfolio performance will be significantly exaggerated.

Estimating OOS Sharpe

Assumptions

1. Observed sample data $\mathbf{R} \in \mathbb{R}^{n\times p}$ satisfies
$$\mathbf{R} = \mathbf{1}_n \mathbf{r}^\top + \mathbf{X},$$
where $\mathbf{X} = \mathbf{Z}\mathbf{\Sigma}^{\frac{1}{2}} \in \mathbb{R}^{n\times p}$. The elements in $\mathbf{Z} \in \mathbb{R}^{n\times p}$ are i.i.d zero mean, variance 1 and finite $(8+\varepsilon)$-order moment for some $\varepsilon > 0$.
2. The portfolio dimension p and the sample size n both tend to infinity, with $p/n \to c > 0$.
3. There exist constants $c_{\mathbf{Q}}, C_{\mathbf{Q}} > 0$ such that $c_{\mathbf{Q}} \leq \lambda_{\min}(\mathbf{Q})$ and $\|\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Q}\mathbf{\Sigma}^{-\frac{1}{2}}\|_{\text{op}} \leq C_{\mathbf{Q}}$ for any sequences (n, p). In addition, we allow $\mathbf{Q} = \mathbf{0}$ when $c < 1$.

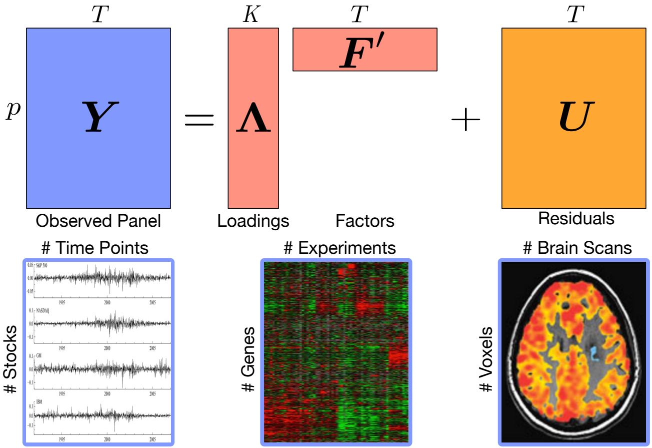
Assumptions

4. $\mathbf{\Sigma}$ is well scaled as $\|\mathbf{\Sigma}/p\|_{\mathrm{tr}} \le C$ for some constant $C > 0$. Denote by $\lambda_1 \ge \cdots \ge \lambda_p$ the eigenvalues of $\mathbf{\Sigma}$. $\lambda_p \ge c_1$ for some constant $c_1 > 0$. One of the following cases must hold:
   - (a) (Bounded spectrum) There exists $C_1 > 0$ such that $\lambda_1 \le C_1$.
   - (b) (Arbitrary number of diverging spikes when $c < 1$) $p/n \to c < 1$ and we allow arbitrary number of top eigenvalues to go to infinity.
   - (c) (Fixed number of diverging spikes when $c \ge 1$) $p/n \to c \ge 1$ and we let the number of diverging spikes be K, K is fixed and $\lambda_1 \le C\lambda_K^2$ for some constant $C > 0$.

Factor models

$$y_{jt} = \boldsymbol{\lambda}_j' \, \mathbf{f}_t + u_{jt}, \quad j = 1, ..., p, \quad t = 1, ..., T \text{ or } n \,, \quad \mathbf{f}_t \in \mathbb{R}^K.$$

Matrix form: $\boldsymbol{Y} = \boldsymbol{\Lambda}\,\mathbf{F}' + \boldsymbol{U}$. $\qquad \mathbb{E}\,\mathbf{f}_t = \mathbf{0}, \;\; \mathbb{E}\, u_{jt} = 0.$



Factor models

$$y_{jt} = \boldsymbol{\lambda}_j^\top \mathbf{f}_t + u_{jt}, \;\; j = 1, ..., p, \;\; t = 1, ..., T \text{ or } n \,, \;\; \mathbf{f}_t \in \mathbb{R}^K.$$

Vector form: $\boldsymbol{y}_t = \boldsymbol{\Lambda}\, \mathbf{f}_t + \boldsymbol{u}_t$.
Covariance structure: $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{y}_t) = \boldsymbol{\Lambda}\, \mathrm{Cov}(\mathbf{f}_t)\boldsymbol{\Lambda}^\top + \mathrm{Cov}(\boldsymbol{u}_t)$.

- Typically, we assume $\mathrm{Cov}(\boldsymbol{u}_t)$ is a diagonal matrix (strict factor model) and $\mathrm{Cov}(\mathbf{f}_t), \mathrm{Cov}(\boldsymbol{u}_t)$ are well-conditioned.
- For $j \le K$, $\lambda_j(\boldsymbol{\Sigma}) \asymp \lambda_j(\boldsymbol{\Lambda}\, \mathrm{Cov}(\mathbf{f}_t)\boldsymbol{\Lambda}^\top) \asymp \lambda_j(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top) \asymp \lambda_j(\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda})$ $\asymp \lambda_j(\sum_{j=1}^p \boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top) \asymp p$ by law of large numbers.
- For $j > K$, $\lambda_j(\boldsymbol{\Sigma}) \asymp \lambda_j(\mathrm{Cov}(\boldsymbol{u}_t)) \asymp 1$.
- Therefore, we have fixed number K of diverging spikes with $\lambda_j \asymp p$ as $p \to \infty$.

Sharpe Estimation

Define the following quantities:

$$T_{n,1}(\mathbf{Q}) = \operatorname{tr}\left[\left(\frac{\mathbf{X}^\top \mathbf{X}}{n} + \mathbf{Q}\right)^{-1} \mathbf{A}\right],$$

$$T_{n,2}(\mathbf{Q}) = \operatorname{tr}\left[\left(\frac{\mathbf{X}^\top \mathbf{X}}{n} + \mathbf{Q}\right)^{-1} \boldsymbol{\Sigma} \left(\frac{\mathbf{X}^\top \mathbf{X}}{n} + \mathbf{Q}\right)^{-1} \mathbf{A}\right],$$

for some deterministic matrix $\mathbf{A} \in \mathbb{R}^{p \times p}$.

By setting $\mathbf{A} = \boldsymbol{\mu}\boldsymbol{\mu}^\top$, Sharpe Ratio (SR) could be expressed as

$$SR(\mathbf{Q}) = \frac{T_{n,1}(\mathbf{Q})}{\sqrt{|T_{n,2}(\mathbf{Q})|}}.$$

Sharpe Estimation

Theorem

Suppose Assumptions 1-4 hold. For any given $\mathbf{Q}$, *a good estimator* $\widehat{SR}(\mathbf{Q})$ *for* $SR(\mathbf{Q})$ *is as follows.*

$$\widehat{SR}(\mathbf{Q}) = \frac{T_{n,1}(\mathbf{Q})}{\sqrt{\left|\widehat{T}_{n,2}(\mathbf{Q})\right|}}, \quad \textit{where} \quad \widehat{T}_{n,2}(\mathbf{Q}) = \frac{\operatorname{tr}(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{A}}{\left(1 - \frac{c}{p}\operatorname{tr}\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\right)^2}.$$

If $\mathbf{A}$ *is semi-positive definite, it holds that*

$$\widehat{SR}(\mathbf{Q})/SR(\mathbf{Q}) \overset{a.s}{\to} 1.$$

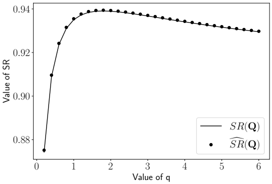
If additionally $\|\mathbf{A}\|_{\mathrm{tr}}$ *is bounded, then* $SR(\mathbf{Q})$ *is almost surely bounded and*

$$\widehat{SR}(\mathbf{Q}) - SR(\mathbf{Q}) \overset{a.s}{\to} 0.$$

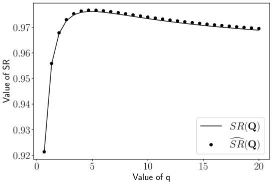
Simulations: Sharpe Estimation

1. Fix $n = 1500$, consider $p = 750$ (ratio $c = 1/2$) and $p = 2250$ (ratio $c = 3/2$).
2. $\mathbf{\Sigma} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) + 2 \cdot \mathbf{1}\mathbf{1}^\top$, where $\{\lambda_i\}_{i=1}^p$ are generated from a truncated $\Gamma^{-1}(1,1)$ distribution, truncated with the interval $[0.01, 9]$, and then ranked in decreasing order.
3. $r_0 = 0$, $\boldsymbol{\mu} = \sqrt{5/p} \cdot (\mathbf{1}(S_+) - \mathbf{1}(S_-)) \in \mathbb{R}^p$. S_+ and S_- are randomly selected subsets of $[p]$ with $|S_+| = |S_-| = p/10$ and $S_+ \cup S_- = \emptyset$.
4. $\mathbf{Q} = q \cdot \mathbf{Q}_0$ where $\mathbf{Q}_0 = \mathrm{diag}(3, ..., 3, 1, ..., 1)$, where the numbers of 3 and 1 entries are both $p/2$. We will vary q.
5. Repeat 1000 times.

Simulations: Sharpe Estimation



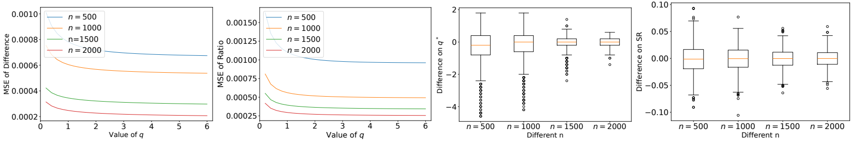
(a) $(n, p) = (1500, 750)$



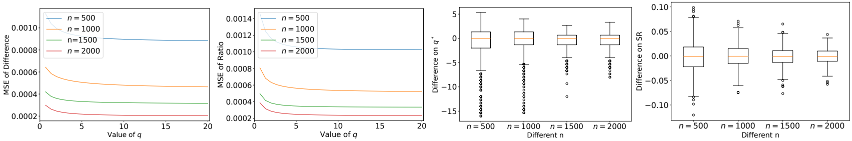
(b) $(n, p) = (1500, 2250)$

Figure 1: Simulation results in the basic settings. Figure 1a shows the case when $c = 1/2$ and Figure 1b depicts the case when $c = 3/2$. The x-axis corresponds to different q values, and the y-axis is the value of SR. The black solid line connects the values of $SR(q \cdot \mathbf{Q}_0)$, while the solid points represent the proposed statistics $\widehat{SR}(q \cdot \mathbf{Q}_0)$.

Simulations: Sharpe Estimation



(a) MSE of diff (b) MSE of ratio (c) q^* diff (d) SR* diff



(e) MSE of diff (f) MSE of ratio (g) q^* diff (h) SR* diff

Figure 2: Simulation results with increasing n. Figures 2a-2d and Figures 2e-2h correspond to $c = 1/2$ and $c = 3/2$. Figures 2a, 2e show $\sum_{b=1}^{1000}(SR_b(q\cdot\mathbf{Q}_0) - \widehat{SR}_b(q\cdot\mathbf{Q}_0))^2/1000$ for different q's. Figures 2b, 2f show $\sum_{b=1}^{1000}(SR_b(q\cdot\mathbf{Q}_0)/\widehat{SR}_b(q\cdot\mathbf{Q}_0) - 1)^2/1000$ for different q's. Figures 2c, 2g give boxplot of $\operatorname{argmax}_q SR_b(q\cdot\mathbf{Q}_0) - \operatorname{argmax}_q \widehat{SR}_b(q\cdot\mathbf{Q}_0)$ for different n's. Figures 2d, 2h give boxplot of $\max_q SR_b(q\cdot\mathbf{Q}_0) - \max_q \widehat{SR}_b(q\cdot\mathbf{Q}_0)$ for different n's.

Estimating Efficient Frontier

Efficient Frontier

When No Risk-free Asset: Given target return $\mu_0 > 0$, the regularized portfolio optimization is given by

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbf{w}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})\mathbf{w}, \quad \text{s.t. } \mathbf{w}^\top\mathbf{r} = \mu_0 \text{ and } \mathbf{w}^\top\mathbf{1} = 1.$$

The optimal $\mathbf{w}^*$ is given by $\mathbf{w}^* = \mathbf{g} + \mu_0 \cdot \mathbf{h}$, where

$$\begin{aligned}
&\mathbf{g} = D^{-1}\Big[B(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1} - A(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{r}\Big], \\
&\mathbf{h} = D^{-1}\Big[C(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{r} - A(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1}\Big], \\
&A = \mathbf{r}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1},\ B = \mathbf{r}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{r}, \\
&C = \mathbf{1}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1},\ D = BC - A^2.
\end{aligned}$$

Efficient Frontier

When No Risk-free Asset: Given target return $\mu_0 > 0$, the regularized portfolio optimization is given by

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathbf{w}^\top(\widehat{\mathbf{\Sigma}} + \mathbf{Q})\mathbf{w}, \quad \text{s.t. } \mathbf{w}^\top\mathbf{r} = \mu_0 \text{ and } \mathbf{w}^\top\mathbf{1} = 1.$$

The optimal $\mathbf{w}^*$ is given by $\mathbf{w}^* = \mathbf{g} + \mu_0 \cdot \mathbf{h}$, where

$$\begin{aligned}
&\mathbf{g} = D^{-1}\Big[B(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1} - A(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{r}\Big],\\
&\mathbf{h} = D^{-1}\Big[C(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{r} - A(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1}\Big],\\
&A = \mathbf{r}^\top(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1},\ B = \mathbf{r}^\top(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{r},\\
&C = \mathbf{1}^\top(\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1}\mathbf{1},\ D = BC - A^2.
\end{aligned}$$

Efficient Frontier: The curve (σ_0, μ_0) as we change target return μ_0, where

$$\sigma_0^2 = \mathbf{w}^{*\top}\mathbf{\Sigma}\mathbf{w}^* = (\mathbf{g} + \mu_0 \cdot \mathbf{h})^\top\mathbf{\Sigma}(\mathbf{g} + \mu_0 \cdot \mathbf{h}),$$

is the variance. Our objective is to estimate σ_0 for any given $\mathbf{Q}$ and μ_0.

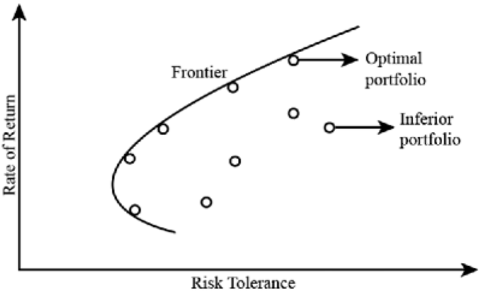
Efficient Frontier

Efficient Frontier with known $\boldsymbol{\Sigma}$:

$$\sigma_0^2 = (\mathbf{g} + \mu_0 \cdot \mathbf{h})^\top \boldsymbol{\Sigma} (\mathbf{g} + \mu_0 \cdot \mathbf{h}) = \mathbf{g}^\top \boldsymbol{\Sigma} \mathbf{g} + 2\mu_0 \mathbf{g}^\top \boldsymbol{\Sigma} \mathbf{h} + \mu_0^2 \mathbf{h}^\top \boldsymbol{\Sigma} \mathbf{h}.$$

When we know true $\boldsymbol{\Sigma}$, we use $\boldsymbol{\Sigma}$ instead of $\widehat{\boldsymbol{\Sigma}} + \mathbf{Q}$ in $\mathbf{g}, \mathbf{h}, A, B, C, D$. Then the above is equivalent to

$$C\sigma_0^2 - C^2/D \cdot (\mu_0 - A/C)^2 = 1. \qquad \text{(Hyperbola)}$$



Assumptions

5. Let $s_0 > 0$ to be the unique solution of the equation.

$$s_0 = \frac{c}{p}\text{tr}\ \boldsymbol{\Sigma}\left(\frac{\boldsymbol{\Sigma}}{1+s_0} + \mathbf{Q}\right)^{-1}.$$

Define

$$\mathcal{A}_{rr} = \mathbf{r}^\top\left(\frac{\boldsymbol{\Sigma}}{1+s_0} + \mathbf{Q}\right)^{-1}\mathbf{r},$$
$$\mathcal{A}_{r1} = \mathbf{r}^\top\left(\frac{\boldsymbol{\Sigma}}{1+s_0} + \mathbf{Q}\right)^{-1}\mathbf{1},$$
$$\mathcal{A}_{11} = \mathbf{1}^\top\left(\frac{\boldsymbol{\Sigma}}{1+s_0} + \mathbf{Q}\right)^{-1}\mathbf{1}.$$

There exists a constant $\rho < 1$ such that $\mathcal{A}_{r1}^2/(\mathcal{A}_{11}\mathcal{A}_{rr}) \leq \rho < 1$.

Efficient Frontier Estimation

Theorem

Suppose that Assumptions 1-5 hold. Define

$$\widehat{\sigma}^2 = \frac{(\mathbf{g} + \mu_0 \mathbf{h})^\top \widehat{\mathbf{\Sigma}} (\mathbf{g} + \mu_0 \mathbf{h})}{(1 - c/p \cdot \mathrm{tr} \widehat{\mathbf{\Sigma}} (\widehat{\mathbf{\Sigma}} + \mathbf{Q})^{-1})^2},$$

where $\mathbf{g}$ *and* $\mathbf{h}$ *are defined as before, it holds that*

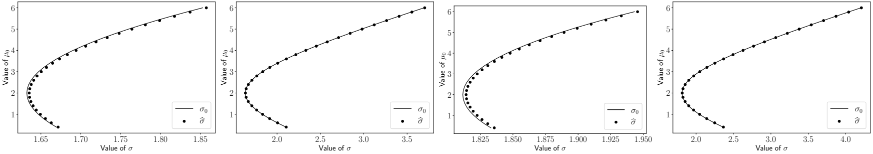
$$\widehat{\sigma}^2 / \sigma_0^2 \stackrel{a.s}{\to} 1.$$

Moreover, the following properties hold:

1. *If* $\mathcal{A}_{rr}$ *is bounded, for any* $r_0 = O(\mu_0)$ *it holds that* $\frac{\mu_0 - r_0}{\sigma_0} - \frac{\mu_0 - r_0}{\widehat{\sigma}} \stackrel{a.s}{\to} 0$.
2. *If* $\mu_0 \leq C\sqrt{\mathcal{A}_{rr}}$ *for some* $C > 0$, *then* $\widehat{\sigma}^2 - \sigma_0^2 \stackrel{a.s}{\to} 0$.

Simulations: Efficient Frontier Estimation

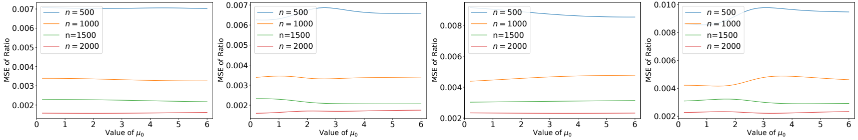
1. Fix $n = 1500$, consider $p = 750$ (ratio $c = 1/2$) and $p = 2250$ (ratio $c = 3/2$).
2. Generate $\boldsymbol{\xi} \in \mathbb{R}^p$ i.i.d. $\Gamma(1,1)$. $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) + 2 \cdot \mathbf{1}\mathbf{1}^\top + \boldsymbol{\xi}\boldsymbol{\xi}^\top$. $\boldsymbol{\Sigma}$ respresents a covariance matrix with two factors.
3. $r_0 = 0$, the mean vector $\boldsymbol{\mu}$ here has two choices: $\boldsymbol{\mu} = \boldsymbol{\mu}_1 = p^{\frac{1}{4}} \boldsymbol{\mu}_0 + 2 \cdot \mathbf{1}$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_0 + 2 \cdot \mathbf{1} + \boldsymbol{\xi}$. $\mathcal{A}_{rr}$ becomes unbounded when $\boldsymbol{\mu} = \boldsymbol{\mu}_1$, while it remains bounded when $\boldsymbol{\mu} = \boldsymbol{\mu}_2$.
4. $\mathbf{Q} = 0.2 \cdot \mathbf{Q}_0$.
5. μ_0 ranges from 0.2 to 6 with the increment of 0.2.
6. Repeat 1000 times.



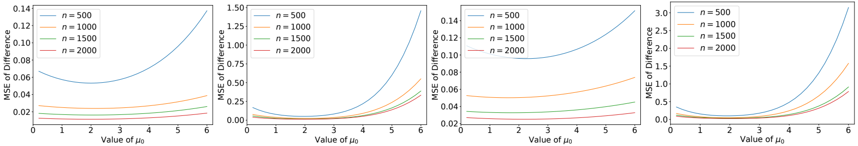
(a) $\boldsymbol{\mu} = \boldsymbol{\mu}_1, c = 1/2$ (b) $\boldsymbol{\mu} = \boldsymbol{\mu}_2, c = 1/2$ (c) $\boldsymbol{\mu} = \boldsymbol{\mu}_1, c = 3/2$ (d) $\boldsymbol{\mu} = \boldsymbol{\mu}_2, c = 3/2$

Figure 3: Efficient frontiers of $\mathbf{w}^*$. The x-axis is the volatility level, and the y-axis is the target return μ_0. The solid line characterizes the curve of (μ_0, σ_0), while the solid points represent the points in the curve of $(\mu_0, \widehat{\sigma})$.

Simulations: Efficient Frontier Estimation



(a) $\boldsymbol{\mu} = \boldsymbol{\mu}_1, c = 1/2$ (b) $\boldsymbol{\mu} = \boldsymbol{\mu}_2, c = 1/2$ (c) $\boldsymbol{\mu} = \boldsymbol{\mu}_1, c = 3/2$ (d) $\boldsymbol{\mu} = \boldsymbol{\mu}_2, c = 3/2$



(e) $\boldsymbol{\mu} = \boldsymbol{\mu}_1, c = 1/2$ (f) $\boldsymbol{\mu} = \boldsymbol{\mu}_2, c = 1/2$ (g) $\boldsymbol{\mu} = \boldsymbol{\mu}_1, c = 3/2$ (h) $\boldsymbol{\mu} = \boldsymbol{\mu}_2, c = 3/2$

Figure 4: Simulation results with increasing n. x-axis in all figures shows different values of μ_0. Figures 4a-4d show $\sum_{b=1}^{1000}(\widehat{\sigma}_b^2/\sigma_{0,b}^2 - 1)^2/1000$. Figures 4e-4h show $\sum_{b=1}^{1000}(\widehat{\sigma}_b^2 - \sigma_{0,b}^2)^2/1000$.

Optimization Over Q

Interesting Questions

★***Q1:*** *Given the maximum OOS Sharpe* $SR_{\max} = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}$*, can* $SR(\mathbf{Q})$ *approach* $SR_{\max}$*? How to predetermine the structure of* $\mathbf{Q}$*?*

★***Q2:*** *Can we optimize* $\mathbf{Q}$ *from* $\widehat{SR}(\mathbf{Q})$*? Define* $\widehat{\mathbf{Q}} = \operatorname{argmax}_{\mathbf{Q}} \widehat{SR}(\mathbf{Q})$*, will the performance of* $SR(\widehat{\mathbf{Q}})$ *be good?*

Answers of Q1

Theorem

Suppose that Assumptions 1-4 hold. Then for any given $\varepsilon > 0$, there exists deterministic sequences of matrices $\widetilde{\mathbf{Q}} \in \mathbb{R}^{p\times p}$ such that with probability 1,

$$1 - \varepsilon \leq \lim_{n\to+\infty} SR(\widetilde{\mathbf{Q}})/SR_{\max} \leq 1.$$

Answers of Q1

Theorem

Suppose that Assumptions 1-4 hold. Then for any given $\varepsilon > 0$, there exists deterministic sequences of matrices $\widetilde{\mathbf{Q}} \in \mathbb{R}^{p\times p}$ such that with probability 1,

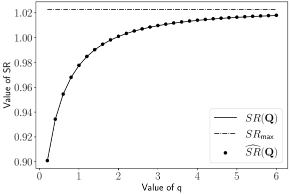
$$1 - \varepsilon \leq \lim_{n\to+\infty} SR(\widetilde{\mathbf{Q}})/SR_{\max} \leq 1.$$

Key of Proof: The existence of $\widetilde{\mathbf{Q}}$ is proved by letting $\widetilde{\mathbf{Q}} = C\boldsymbol{\Sigma}$ for some constant C large enough.

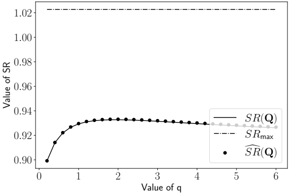
Simulations: Different Structure of $\mathbf{Q}$

1. Fix $n = 1500$, consider $p = 750$ (ratio $c = 1/2$) and $p = 2250$ (ratio $c = 3/2$).
2. $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \ldots, \lambda_p) + 2 \cdot \mathbf{1}\mathbf{1}^\top$, where $\{\lambda_i\}_{i=1}^p$ are generated from a truncated $\Gamma^{-1}(1, 1)$ distribution, truncated with the interval $[0.01, 9]$, and then ranked in decreasing order.
3. $r_0 = 0$, $\boldsymbol{\mu} = \sqrt{5/p} \cdot (\mathbf{1}(S_+) - \mathbf{1}(S_-)) \in \mathbb{R}^p$. S_+ and S_- are randomly selected subsets of $[p]$ with $|S_+| = |S_-| = p/10$ and $S_+ \cup S_- = \emptyset$.
4. Let $\mathbf{Q}_0 = \text{diag}(3, ..., 3, 1, ..., 1)$, where the numbers of 3 and 1 entries are both $p/2$. Define $\mathbf{Q}_1 = 0.1\mathbf{Q}_0 + q \cdot \text{diag}(\lambda_1, ..., \lambda_p)$; $\mathbf{Q}_2 = 0.5\mathbf{I}_p + q\mathbf{Q}_0$ and $\mathbf{Q}_3 = q\boldsymbol{\Sigma}$. We will vary q.
5. Repeat 1000 times.

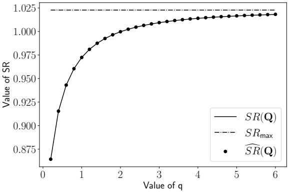
Simulations: Different Structure of $\mathbf{Q}$



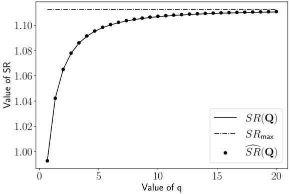
(a) $\mathbf{Q} = \mathbf{Q}_1, c = 1/2$



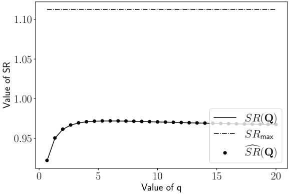
(b) $\mathbf{Q} = \mathbf{Q}_2, c = 1/2$



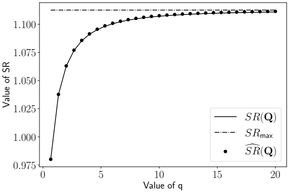
(c) $\mathbf{Q} = \mathbf{Q}_3, c = 1/2$



(d) $\mathbf{Q} = \mathbf{Q}_1, c = 3/2$



(e) $\mathbf{Q} = \mathbf{Q}_2, c = 3/2$



(f) $\mathbf{Q} = \mathbf{Q}_3, c = 3/2$

Figure 5: Simulation results with different $\mathbf{Q}$'s.

Answers of Q2 with Numerical Results

Table 1: Comparison for $SR(\widehat{\mathbf{Q}})$ and $\widehat{SR}(\widehat{\mathbf{Q}})$. The mean gives the average value and the range gives the minimum and maximum values over the 20 independent trials.

(n,p)	$SR_{\max}$	Optimization over full $\mathbf{Q}$			
		mean of $SR(\widehat{\mathbf{Q}})$	Range of $SR(\widehat{\mathbf{Q}})$	mean of $\widehat{SR}(\widehat{\mathbf{Q}})$	Range of $\widehat{SR}(\widehat{\mathbf{Q}})$
$(500, 250)$	0.923	0.643	[0.604,0.694]	1.299	[1.178,1.454]
$(1000, 500)$	1.123	0.791	[0.738,0.824]	1.513	[1.406,1.578]
(n,p)	$SR_{\max}$	Optimization over diagonal $\mathbf{Q}$			
		mean of $SR(\widehat{\mathbf{Q}})$	Range of $SR(\widehat{\mathbf{Q}})$	mean of $\widehat{SR}(\widehat{\mathbf{Q}})$	Range of $\widehat{SR}(\widehat{\mathbf{Q}})$
$(500, 250)$	0.923	0.770	[0.715,0.818]	0.967	[0.909,1.056]
$(1000, 500)$	1.123	0.944	[0.912,0.979]	1.146	[1.081,1.224]

Answers of Q2 with Numerical Results

Table 1: Comparison for $SR(\widehat{\mathbf{Q}})$ and $\widehat{SR}(\widehat{\mathbf{Q}})$. The mean gives the average value and the range gives the minimum and maximum values over the 20 independent trials.

(n,p)	$SR_{\max}$	Optimization over full $\mathbf{Q}$			
		mean of $SR(\widehat{\mathbf{Q}})$	Range of $SR(\widehat{\mathbf{Q}})$	mean of $\widehat{SR}(\widehat{\mathbf{Q}})$	Range of $\widehat{SR}(\widehat{\mathbf{Q}})$
$(500, 250)$	0.923	0.643	[0.604,0.694]	1.299	[1.178,1.454]
$(1000, 500)$	1.123	0.791	[0.738,0.824]	1.513	[1.406,1.578]
(n,p)	$SR_{\max}$	Optimization over diagonal $\mathbf{Q}$			
		mean of $SR(\widehat{\mathbf{Q}})$	Range of $SR(\widehat{\mathbf{Q}})$	mean of $\widehat{SR}(\widehat{\mathbf{Q}})$	Range of $\widehat{SR}(\widehat{\mathbf{Q}})$
$(500, 250)$	0.923	0.770	[0.715,0.818]	0.967	[0.909,1.056]
$(1000, 500)$	1.123	0.944	[0.912,0.979]	1.146	[1.081,1.224]

Primary Reason: $\widehat{\mathbf{Q}}$ is overfitted to in-sample data, breaking down the independence between $\widehat{\boldsymbol{\Sigma}}$ and $\mathbf{Q}$.

Assumptions

6. There exists universal constants $l, L > 0$ such that for all $\mathbf{Q} \in \mathcal{Q}$, both $SR(\mathbf{Q})$ and $\widehat{SR}(\mathbf{Q})$ satisfy $l \leq SR(\mathbf{Q}),\ \widehat{SR}(\mathbf{Q}) \leq L$ almost surely for all n large enough.
7. There exists a sequence of bijections $\phi_n : \mathcal{B} \to \mathcal{Q}$, where $\mathcal{B} \subset \mathbb{R}^k$ is a fixed compact set (independent of n) for some constant $k > 0$. Furthermore, the sequence $\{\phi_n\}$ is equicontinuous with respect to the operator norm: for any $\varepsilon > 0$, there exists $\delta > 0$ (independent of n) such that for all n and all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{B}$,

$$\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|_2 \leq \delta \quad \Longrightarrow \quad \|\phi_n(\boldsymbol{\alpha}) - \phi_n(\boldsymbol{\alpha}')\|_{\text{op}} \leq \varepsilon.$$

Assumptions

6. There exists universal constants $l, L > 0$ such that for all $\mathbf{Q} \in \mathcal{Q}$, both $SR(\mathbf{Q})$ and $\widehat{SR}(\mathbf{Q})$ satisfy $l \leq SR(\mathbf{Q}),\ \widehat{SR}(\mathbf{Q}) \leq L$ almost surely for all n large enough.
7. There exists a sequence of bijections $\phi_n : \mathcal{B} \to \mathcal{Q}$, where $\mathcal{B} \subset \mathbb{R}^k$ is a fixed compact set (independent of n) for some constant $k > 0$. Furthermore, the sequence $\{\phi_n\}$ is equicontinuous with respect to the operator norm: for any $\varepsilon > 0$, there exists $\delta > 0$ (independent of n) such that for all n and all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{B}$,

$$\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|_2 \leq \delta \quad \Longrightarrow \quad \|\phi_n(\boldsymbol{\alpha}) - \phi_n(\boldsymbol{\alpha}')\|_{\mathrm{op}} \leq \varepsilon.$$

Remark: $\mathcal{Q}$ is the candidate set for $\mathbf{Q}$. A specific example for $\mathcal{Q}$ is $\phi_n(\boldsymbol{\alpha}) = \alpha_1 \mathbf{Q}_1 + \alpha_2 \mathbf{Q}_2 + \cdots + \alpha_k \mathbf{Q}_k$ where $\mathbf{Q}_1, \ldots, \mathbf{Q}_k$ are predetermined matrices with $\|\mathbf{Q}_j\|_{\mathrm{op}}$ bounded for all j, $\mathbf{Q}_j$ are linearly independent, and the coefficients $\boldsymbol{\alpha}$ vary over a compact set in $\mathbb{R}^k$.

Consistency Results for Optimal $\mathbf{Q}$

Theorem

Suppose that Assumptions 1-4 and 6-7 hold. Define

$$\widehat{\mathbf{Q}} = \operatorname{argmax}_{\mathbf{Q} \in \mathcal{Q}} \widehat{SR}(\mathbf{Q}).$$

It holds that

$$\widehat{SR}(\widehat{\mathbf{Q}})/SR(\widehat{\mathbf{Q}}) \overset{a.s}{\to} 1, \quad \widehat{SR}(\widehat{\mathbf{Q}}) - SR(\widehat{\mathbf{Q}}) \overset{a.s}{\to} 0.$$

Consistency Results for Optimal $\mathbf{Q}$

Theorem

Suppose that Assumptions 1-4 and 6-7 hold. Define

$$\widehat{\mathbf{Q}} = \operatorname{argmax}_{\mathbf{Q}\in\mathcal{Q}} \widehat{SR}(\mathbf{Q}).$$

It holds that

$$\widehat{SR}(\widehat{\mathbf{Q}})/SR(\widehat{\mathbf{Q}}) \overset{a.s}{\to} 1, \quad \widehat{SR}(\widehat{\mathbf{Q}}) - SR(\widehat{\mathbf{Q}}) \overset{a.s}{\to} 0.$$

Control of Overfitting Issues: When the search space for $\mathbf{Q}$ ($\mathcal{Q}$) is well behaved and restricted to a finite dimensional family, the overfitting issue can be controlled, and the optimized candidate achieves consistent performance in the large-sample limit.

Estimating with Sample Mean $\widehat{\boldsymbol{\mu}}$

OOS Sharpe with Sample Mean $\widehat{\boldsymbol{\mu}}$

ℓ_2-Regularized-MV: Consider the optimization with regularization $\mathbf{Q}$:

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w} \in \mathbb{R}^p} \mathbf{w}^\top (\widehat{\boldsymbol{\Sigma}} + \mathbf{Q}) \mathbf{w} \quad \text{s. t.} \quad \mathbf{w}^\top \widehat{\boldsymbol{\mu}} = \mu_0,$$

where $\mathbf{Q}$ is positive definite. The optimal $\mathbf{w}^*$ satisfies

$$\mathbf{w}^* \propto (\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1} \widehat{\boldsymbol{\mu}}.$$

OOS Sharpe with Sample Mean $\widehat{\boldsymbol{\mu}}$

ℓ_2-Regularized-MV: Consider the optimization with regularization $\mathbf{Q}$:

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w}\in\mathbb{R}^p} \mathbf{w}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})\mathbf{w} \quad \text{s. t.} \quad \mathbf{w}^\top\widehat{\boldsymbol{\mu}} = \mu_0,$$

where $\mathbf{Q}$ is positive definite. The optimal $\mathbf{w}^*$ satisfies

$$\mathbf{w}^* \propto (\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\widehat{\boldsymbol{\mu}}.$$

OOS Sharpe Ratio of $\mathbf{w}^*$:

$$SR(\mathbf{Q}) = \frac{\mathbb{E}_{\widetilde{\mathbf{R}}}[\mathbf{w}^{*\top}(\widetilde{\mathbf{R}} - r_0\mathbf{1})]}{\sqrt{\operatorname{Var}_{\widetilde{\mathbf{R}}}[\mathbf{w}^{*\top}(\widetilde{\mathbf{R}} - r_0\mathbf{1})]}} = \frac{\widehat{\boldsymbol{\mu}}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\mu}}{\sqrt{\widehat{\boldsymbol{\mu}}^\top(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\boldsymbol{\Sigma}(\widehat{\boldsymbol{\Sigma}} + \mathbf{Q})^{-1}\widehat{\boldsymbol{\mu}}}},$$

where $\widetilde{\mathbf{R}}$ is an out-of-sample point with mean $\mathbf{r} = \boldsymbol{\mu} + r_0\mathbf{1}$ and cov $\boldsymbol{\Sigma}$.

Assumptions

8. Observed sample data $\mathbf{R} \in \mathbb{R}^{n \times p}$ satisfies

$$\mathbf{R} = \mathbf{1}_n \mathbf{r}^\top + \mathbf{X},$$

where $\mathbf{X} = \mathbf{Z}\boldsymbol{\Sigma}^{\frac{1}{2}} \in \mathbb{R}^{n \times p}$. The elements in $\mathbf{Z} \in \mathbb{R}^{n \times p}$ are i.i.d zero mean, variance 1 Gaussian random variables.

Sharpe Estimation with $\widehat{\boldsymbol{\mu}}$

Theorem

Suppose Assumptions 2-4 and 8 hold. For any given $\mathbf{Q}$, *a good estimator* $\widehat{SR}(\mathbf{Q})$ *for* $SR(\mathbf{Q})$ *is as follows.*

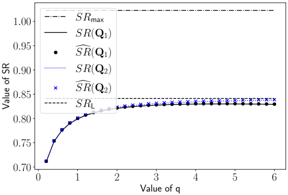
$$\widehat{SR}(\mathbf{Q}) = \frac{\widehat{\boldsymbol{\mu}}^\top(\widehat{\boldsymbol{\Sigma}}+\mathbf{Q})^{-1}\widehat{\boldsymbol{\mu}} - \frac{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}+\mathbf{Q})^{-1}\widehat{\boldsymbol{\Sigma}}}{n-\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}+\mathbf{Q})^{-1}\widehat{\boldsymbol{\Sigma}}}}{\sqrt{\widehat{\boldsymbol{\mu}}^\top(\widehat{\boldsymbol{\Sigma}}+\mathbf{Q})^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\Sigma}}+\mathbf{Q})^{-1}\widehat{\boldsymbol{\mu}}}} \cdot \left(1 - \frac{c}{p}\mathrm{tr}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\Sigma}}+\mathbf{Q})^{-1}\right).$$

If $\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\|_2$ *is bounded and* $\boldsymbol{\mu}^\top(\frac{\boldsymbol{\Sigma}}{1+s_0}+\mathbf{Q})^{-1}\boldsymbol{\mu}$ *is lower bounded by some constant, it holds that*

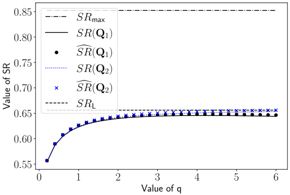
$$\widehat{SR}(\mathbf{Q})/SR(\mathbf{Q}) \overset{a.s}{\to} 1.$$

Simulations: Sharpe Estimation with $\widehat{\boldsymbol{\mu}}$

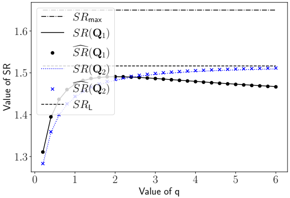
1. Fix $n = 1500$, consider $p = 750$ (ratio $c = 1/2$) and $p = 2250$ (ratio $c = 3/2$).
2. $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) + 2 \cdot \mathbf{1}\mathbf{1}^\top$, where $\{\lambda_i\}_{i=1}^p$ are generated from a truncated $\Gamma^{-1}(1, 1)$ distribution, truncated with the interval $[0.01, 9]$, and then ranked in decreasing order.
3. $r_0 = 0$, $\boldsymbol{\mu}_0 = \sqrt{5/p} \cdot (\mathbf{1}(S_+) - \mathbf{1}(S_-)) \in \mathbb{R}^p$. S_+ and S_- are randomly selected subsets of $[p]$ with $|S_+| = |S_-| = p/10$ and $S_+ \cup S_- = \emptyset$. For $\boldsymbol{\mu}_3$, we assume that each element follows an independent uniform distribution, $\mathrm{Unif}(-\sqrt{2/p}, \sqrt{2/p})$, $\boldsymbol{\mu}_4 = \boldsymbol{\mu}_3 + 2 \cdot \mathbf{1}_p$.
4. $\mathbf{Q}_1 = q \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $\mathbf{Q}_2 = q\boldsymbol{\Sigma}$. We will vary q.
5. Repeat 1000 times.



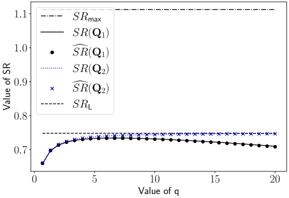
(a) $\boldsymbol{\mu} = \boldsymbol{\mu}_0, c = 1/2$



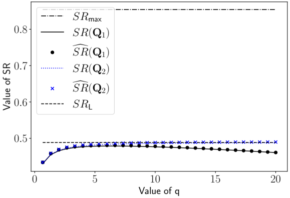
(b) $\boldsymbol{\mu} = \boldsymbol{\mu}_3, c = 1/2$



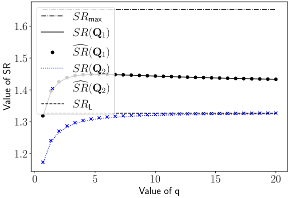
(c) $\boldsymbol{\mu} = \boldsymbol{\mu}_4, c = 1/2$



(d) $\boldsymbol{\mu} = \boldsymbol{\mu}_0, c = 3/2$



(e) $\boldsymbol{\mu} = \boldsymbol{\mu}_3, c = 3/2$



(f) $\boldsymbol{\mu} = \boldsymbol{\mu}_4, c = 3/2$

Figure 6: $SR_{\max} = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}$, and $SR_L = SR^2_{\max} / \sqrt{SR^2_{\max} + c}$.

Real Data Experiments

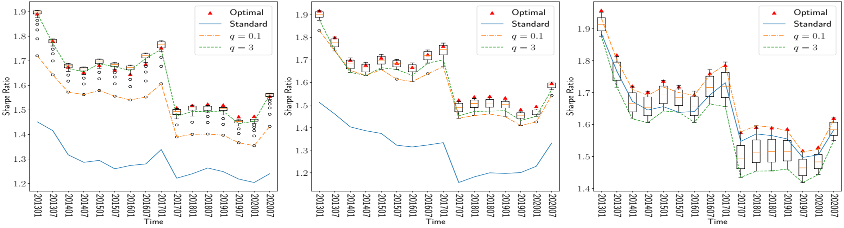
Real Data: Mean-Variance Portfolio

- After deleting stocks with missing values, we have $p = 365$ stocks.
- Portfolios are built using historical data spanning 1, 2, 4 years, and rebalanced monthly.
- Each allocation vector $\mathbf{w}^*$ is held for the entire future testing month. We then have returns of the portfolio $\mathbf{w}^*$ in each trading day of the month.
- For now, we use OOS average return as $\boldsymbol{\mu}$ in each testing month for portfolio construction (known $\boldsymbol{\mu}$).
- Repeat the procedure in a rolling fashion for all testing months from Jan 2013 to Jun 2023 and record daily returns for each trading day.

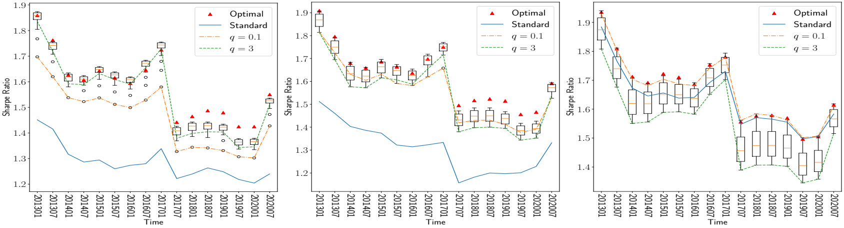
Real Data: Mean-Variance Portfolio

1. Consider two candidate sets. $\mathcal{Q}_1 = \{q \cdot \widehat{\boldsymbol{\Sigma}}_{pre}, q \in [1:30]/10\}$, where $\widehat{\boldsymbol{\Sigma}}_{pre}$ represents the sample covariance pre-trained from 2004 to 2008, not overlapping with data for portfolio construction and evaluation. $\mathcal{Q}_2 = \{q \cdot \mathbf{I}_p, q \in [1:30]/10\}$, where $\mathbf{I}_p$ is the identity matrix.
2. Calculate sample cov $\widehat{\boldsymbol{\Sigma}}$ with 1,2,4-year historical data and construct the regularized MV portfolio.
3. For each testing month, we run experiments for all candidate q values and also consider no regularization, i.e. $q = 0$, where we have $\mathbf{w} \propto \widehat{\boldsymbol{\Sigma}}^+ \boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}}^+$ is the pseudo inverse, and the optimized $q^* \in \mathcal{Q}$ using our estimator. Note that q^* changes from month to month.
4. We report the average Sharpe ratio of daily portfolio returns over the future three years.

Real Data: Mean-Variance Portfolio



(a) 1 year, $\mathcal{Q} = \mathcal{Q}_1, c > 1$ (b) 2 years, $\mathcal{Q} = \mathcal{Q}_1, c < 1$ (c) 4 years, $\mathcal{Q} = \mathcal{Q}_1, c < 1$



(d) 1 year, $\mathcal{Q} = \mathcal{Q}_2, c > 1$ (e) 2 years, $\mathcal{Q} = \mathcal{Q}_2, c < 1$ (f) 4 years, $\mathcal{Q} = \mathcal{Q}_2, c < 1$

Figure 7: SR of mean-variance portfolios. The x-axis labels the rolling period, while the

Real Data: Global Minimum Variance Portfolio

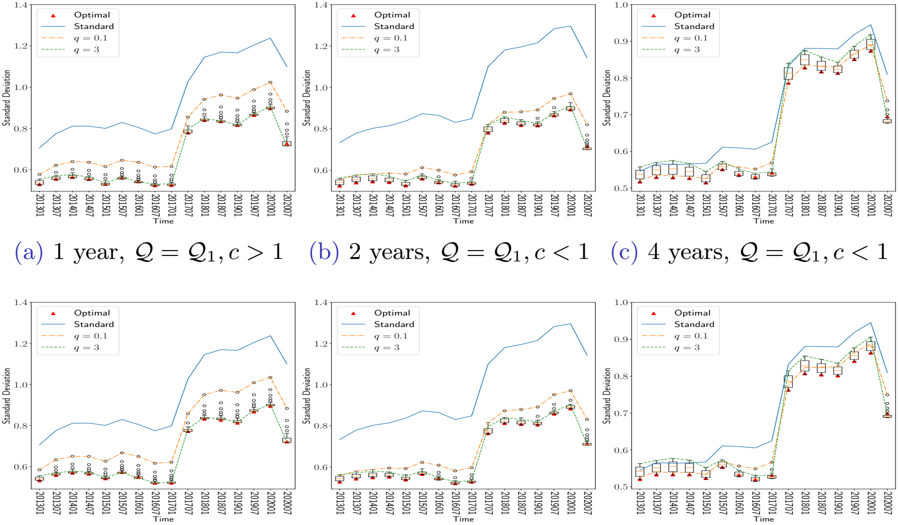
Using future average returns as $\boldsymbol{\mu}$ is not feasible in practical portfolio construction. One remedy approach is:

- Consider GMV portfolio, which does not require the knowledge of $\boldsymbol{\mu}$:

$$\mathbf{w}^* = \frac{(\widehat{\boldsymbol{\Sigma}} + q\mathbf{I})^{-1}\mathbf{1}}{\mathbf{1}^\top(\widehat{\boldsymbol{\Sigma}} + q\mathbf{I})^{-1}\mathbf{1}}.$$

  Then we check which GMV portfolio attains the minimum OOS empirical variance.

Real Data: Global Minimum Variance Portfolio



(a) 1 year, $\mathcal{Q} = \mathcal{Q}_1, c > 1$ (b) 2 years, $\mathcal{Q} = \mathcal{Q}_1, c < 1$ (c) 4 years, $\mathcal{Q} = \mathcal{Q}_1, c < 1$

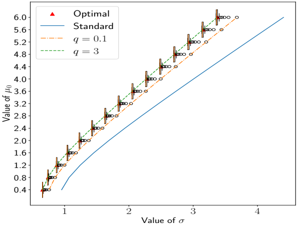
(d) 1 year, $\mathcal{Q} = \mathcal{Q}_2, c > 1$ (e) 2 years, $\mathcal{Q} = \mathcal{Q}_2, c < 1$ (f) 4 years, $\mathcal{Q} = \mathcal{Q}_2, c < 1$

Figure 8: Standard deviation (volatility) of global minimum variance

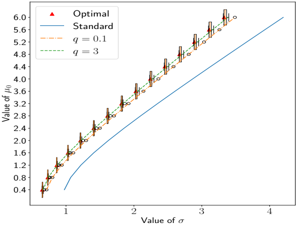
Real Data: Efficient Frontier

1. Set $\mathbf{Q} \in \mathcal{Q}_1$ or $\mathbf{Q} \in \mathcal{Q}_2$ as before. We vary μ_0 from 0.4 to 6 with an increment of 0.4. For each μ_0, carry out 2 to 5 below.
2. We build portfolio assuming no risk-free asset.
3. Let $\mathbf{r}$ be the average return vector in the testing month. The optimal portfolio is given by $\mathbf{w}^* = \mathbf{g} + \mu_0 \mathbf{h}$. We run experiments for all q values in the candidate sets, the case of $q = 0$ and the optimized q^*, which is obtained by minimizing $\frac{(\mathbf{g}+\mu_0\mathbf{h})^\top \widehat{\mathbf{\Sigma}}(\mathbf{g}+\mu_0\mathbf{h})}{(1-c/p \cdot \mathrm{tr}\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{\Sigma}}+q\mathbf{I})^{-1})^2}$ over all q's.
4. We monthly roll the procedure from Jan 2013 to Jun 2023 and collect daily portfolio returns for each q value.
5. We calculate the standard deviation of the daily returns for each q value, including $q = 0$ and $q = q^*$, over the ten-year period.

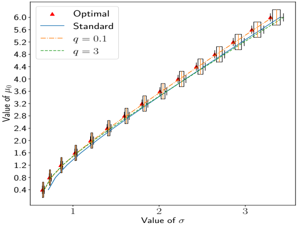
Real Data: Efficient Frontier



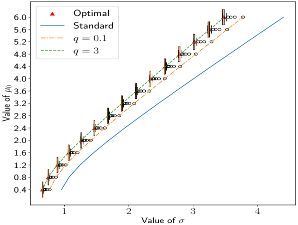
(a) 1 year, $\mathcal{Q} = \mathcal{Q}_1$



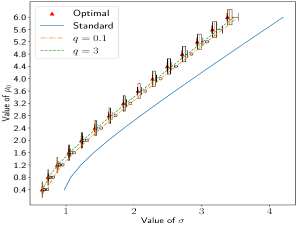
(b) 2 years, $\mathcal{Q} = \mathcal{Q}_1$



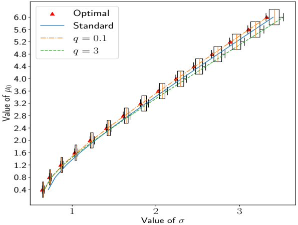
(c) 4 years, $\mathcal{Q} = \mathcal{Q}_1$



(d) 1 year, $\mathcal{Q} = \mathcal{Q}_2$



(e) 2 years, $\mathcal{Q} = \mathcal{Q}_2$



(f) 4 years, $\mathcal{Q} = \mathcal{Q}_2$

Figure 9: The corrected efficient frontier. The x-axis represents the value of σ, while the

Conclusion

Concluding Remarks

- ★ Introduced a novel in-sample approach to estimate the out-of-sample Sharpe ratio in high-dimensional portfolio optimization.
- ★ Relaxed conditions allowing arbitrary diverging spikes when $c < 1$ and K diverging spikes when $c \geq 1$.
- ★ Extended to the estimation of efficient frontier when no risk-free asset.
- ★ Used the OOS Sharpe estimator as objective to optimize the Ridge tuning parameter cycle by cycle.
- ★ Verified the performance of the estimator via extensive numerical experiments.

Thank you!