A Statistical Explanation to Markowitz Optimization Enigma

Ming Yuan

Morgridge Institute for Research
and
Department of Statistics
University of Wisconsin-Madison

myuan@stat.wisc.edu
http://www.stat.wisc.edu/~myuan

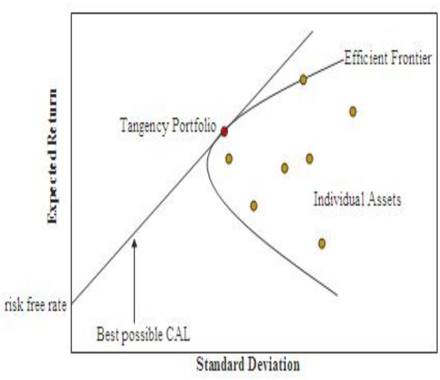
Outline

- What – Markowitz optimization enigma (Michaud, 1998)
- How – Subspace mean-variance analysis
- Why – Asymptotic efficiency

Markowitz Optimization Enigma

Mean-Variance Analysis

$$\min_{\boldsymbol{w} \in \boldsymbol{R}^N} \boldsymbol{w}^\top \Sigma \boldsymbol{w} \qquad \text{subject to } \boldsymbol{w}^\top E = \mu, \quad \boldsymbol{w}^\top \mathbb{1} = 1$$



- Investment universe $\boldsymbol{r} \in \boldsymbol{R}^N$

$$\boldsymbol{E}(\boldsymbol{r}) = E \qquad \text{var}(\boldsymbol{r}) = \Sigma$$

- Tangency portfolio

$$\boldsymbol{w}_{\text{tp}} = \frac{\Sigma^{-1} E}{\mathbb{1}^\top \Sigma^{-1} E}$$

- Sharpe ratio

$$s(\boldsymbol{w}_{\text{tp}}) = (E^\top \Sigma^{-1} E)^{1/2}$$

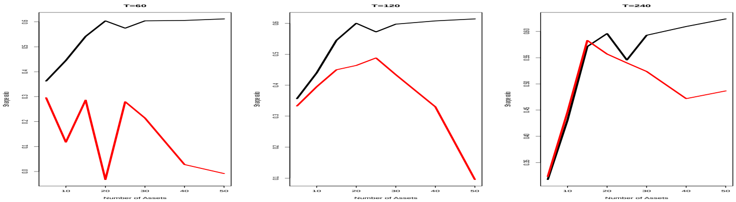
The Markowitz Optimization Enigma

$$\boldsymbol{w}_{\mathrm{tp}} = \frac{\Sigma^{-1} E}{\mathbb{1}^\top \Sigma^{-1} E} \qquad \Longrightarrow \qquad \hat{\boldsymbol{w}}_{\mathrm{tp}} = \frac{\hat{\Sigma}^{-1} \hat{E}}{\mathbb{1}^\top \hat{\Sigma}^{-1} \hat{E}}$$

- Classical econometric theory

$$\text{small } N \Longrightarrow \left.\begin{array}{c} \hat{\Sigma} \approx_T \Sigma \\ \hat{E} \approx_T E \end{array}\right\} \Longrightarrow \hat{\boldsymbol{w}}_{\mathrm{tp}} \approx_T \boldsymbol{w}_{\mathrm{tp}} \Longrightarrow s(\hat{\boldsymbol{w}}_{\mathrm{tp}}) \approx_T s(\boldsymbol{w}_{\mathrm{tp}})$$

- Lackluster performance in practice



Is Optimized Optimal

- Improved moment estimators
  - Factor models – e.g., McKinlay and Pastor (2000)
  - Shrinkage – e.g., Jorion, 1986; Ledoit and Wolf (2004)
- Estimated minimum variance portfolio
  - "*Minimum-variance portfolio usually performs better out of sample than mean-variance portfolios – even when performance measure depends on variance and mean*"
    – DeMiguel, Garlappi and Uppal (2009)
- Impose short-sale constraints
  - "*Sample covariance matrix (with shortsale constraints) performs almost as well as those constructed using factor models, shrinkage estimators or daily returns*"
    – Jagannathan and Ma (2003)
- "Naïve" diversification – DeMiguel, Garlappi and Uppal (2009)......

Are we back to square one?

- Difficulty

$$\text{estimate } E, \Sigma \gg \text{estimate } \boldsymbol{w}_{\text{tp}} \gg \text{achieve optimal Sharpe ratio}$$

- Occam's razor – when N is large
  - "Impossibility" in accurate estimation of E and Σ
  - "Extreme difficulty" in accurate estimation of $\boldsymbol{w}_{\text{tp}}$
  - But it is possible to achieve optimal Sharpe ratio
- "Asymptotic" efficient portfolio selection

$$s(\hat{\boldsymbol{w}}) \approx_{T,N} s(\boldsymbol{w}_{\text{tp}}) =: s_{\text{opt}}$$

Subspace Mean-Variance Analysis

Subspace Mean-Variance Analysis

$$\min_{\boldsymbol{w} \in \mathcal{P}} \boldsymbol{w}^\top \Sigma \boldsymbol{w} \qquad \text{subject to } \boldsymbol{w}^\top E = \mu, \quad \boldsymbol{w}^\top \mathbb{1} = 1$$

- Subspace "Tangency portfolio"

$$\boldsymbol{w}_{\text{tp}}^{\mathcal{P}} = \frac{P_{\mathcal{P}} \left(P_{\mathcal{P}}^\top \Sigma P_{\mathcal{P}} \right)^{-1} P_{\mathcal{P}}^\top E}{\mathbb{1}^\top P_{\mathcal{P}} \left(P_{\mathcal{P}}^\top \Sigma P_{\mathcal{P}} \right)^{-1} P_{\mathcal{P}}^\top E}$$

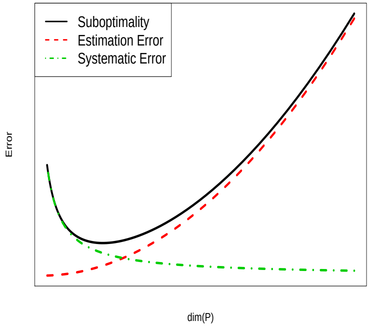
- "Optimal" Sharpe ratio

$$s(\boldsymbol{w}_{\text{tp}}^{\mathcal{P}}) = \left\{ E^\top P_{\mathcal{P}} \left(P_{\mathcal{P}}^\top \Sigma P_{\mathcal{P}} \right)^{-1} P_{\mathcal{P}}^\top E \right\}^{-1}$$

- Examples
  - $\dim(\mathcal{P}) = 1$ – Naïve diversification, value weighted porfolio, ...
  - $\dim(\mathcal{P}) = N$ – Tangency portfolio

A Tale of Two Errors

$$s(\widehat{\boldsymbol{w}_{\text{tp}}^{\mathcal{P}}}) - s_{\text{opt}} = \underbrace{\left[s(\widehat{\boldsymbol{w}_{\text{tp}}^{\mathcal{P}}}) - s(\boldsymbol{w}_{\text{tp}}^{\mathcal{P}})\right]}_{\text{Estimation Error}} + \underbrace{\left[s(\boldsymbol{w}_{\text{tp}}^{\mathcal{P}}) - s_{\text{opt}}\right]}_{\text{Systematic Error}}$$



- When $\dim(\mathcal{P})$ is small
  - Easy to estimate – Estimation Error is small
  - Choice may be suboptimal – Systematic Error could be large
- When $\dim(\mathcal{P})$ is large
  - Difficult to estimate – Curse of dimensionality
  - Systematic Error is small

Constructing a sequence of $\mathcal{P}$

- Construction
  - Compute the sample covariance matrix
  - Construct the linear subspace $\hat{\mathcal{P}}$

$$\hat{\mathcal{P}} = \text{l.s.}\{\mathbb{1}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{K-1}\}$$

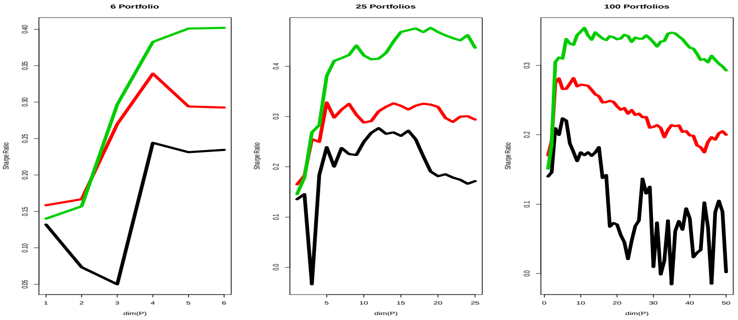
  where $\boldsymbol{v}_k$ is the kth eigenvector of $\hat{\Sigma}$
  - Construct the subspace tangency portfolio in $\hat{\mathcal{P}}$:

$$\hat{\boldsymbol{w}}_{\text{tp}}^{\hat{\mathcal{P}}} = \frac{P_{\hat{\mathcal{P}}} \left(P_{\hat{\mathcal{P}}}^\top \hat{\Sigma} P_{\hat{\mathcal{P}}} \right)^{-1} P_{\hat{\mathcal{P}}}^\top \hat{E}}{\mathbb{1}^\top P_{\hat{\mathcal{P}}} \left(P_{\hat{\mathcal{P}}}^\top \hat{\Sigma} P_{\hat{\mathcal{P}}} \right)^{-1} P_{\hat{\mathcal{P}}}^\top \hat{E}}.$$

- $\hat{\mathcal{P}}_1 \subset \hat{\mathcal{P}}_2 \subset \ldots \subset \hat{\mathcal{P}}_N$
  - $\hat{\mathcal{P}}_1$ leads to naïve diversification
  - $\hat{\mathcal{P}}_N$ leads to sample tangency portfolio

Can we do better than 1/N?

- Fama-French portfolios formed on Size and Book-to-Market – 2×3, 5×5, 10×10
- Period: 50 years (01/01/1961 – 12/31/2010)
- Black – $T = 120$ months Red – $T = 180$ Green – $T = 240$



Asymptotic Efficiency

Structure of Asset Returns

- Approximate factor model

$$r_{jt} = E_j + \beta_{j1} f_{1t} + \ldots + \beta_{jK} f_{Kt} + \varepsilon_{jt}, \qquad j = 1, \ldots, N; \quad t = 1, \ldots, T$$

- Beta pricing relationship

$$E_j = \alpha_j + \beta_{j1}\mu_1^f + \ldots + \beta_{jK}\mu_K^f, \qquad j = 1, 2, \ldots, N$$

  - Introduced by Chamberlain and Rothchild (1983)
  - Detailed discussion in Connor, Goldberg and Korajczyk (2010)
  - No arbitrage $\Longleftrightarrow \boldsymbol{\alpha}^\top \boldsymbol{\alpha}$ is bounded – Huberman (1982)
  - Examples – CAPM, APT

- Idiosyncratic noise
  - Weakly temporal dependence – e.g., α-mixing
  - Allow cross-section dependence – Σ_ε has eigenvalues bounded away from 0 and ∞

How to Choose $\mathcal{P}$

- Best choice – $\mathcal{P}^* = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K]$
- Second best choice – $\mathcal{P}$ = leading eigenspace of $\Sigma := \mathrm{cov}(\boldsymbol{r})$
- Feasible choice – $\hat{\mathcal{P}}$ = leading eigenspace of $\hat{\Sigma}$
- Estimated subspace tangency portfolio

$$\hat{\boldsymbol{w}}_{\mathrm{tp}}^{\hat{\mathcal{P}}} = \frac{P_{\hat{\mathcal{P}}} \left(P_{\hat{\mathcal{P}}}^\top \hat{\Sigma} P_{\hat{\mathcal{P}}} \right)^{-1} P_{\hat{\mathcal{P}}}^\top \hat{E}}{\mathbb{1}^\top P_{\hat{\mathcal{P}}} \left(P_{\hat{\mathcal{P}}}^\top \hat{\Sigma} P_{\hat{\mathcal{P}}} \right)^{-1} P_{\hat{\mathcal{P}}}^\top \hat{E}}$$

- Error analysis

$$s(\hat{\boldsymbol{w}}_{\mathrm{tp}}^{\hat{\mathcal{P}}}) - s(\boldsymbol{w}_{\mathrm{tp}}) = \underbrace{\left[s(\hat{\boldsymbol{w}}_{\mathrm{tp}}^{\hat{\mathcal{P}}}) - s(\boldsymbol{w}_{\mathrm{tp}}^{\hat{\mathcal{P}}}) \right]}_{\text{Estimation Error I}} + \underbrace{\left[s(\boldsymbol{w}_{\mathrm{tp}}^{\hat{\mathcal{P}}}) - s(\boldsymbol{w}_{\mathrm{tp}}^{\mathcal{P}}) \right]}_{\text{Estimation Error II}} + \underbrace{\left[s(\boldsymbol{w}_{\mathrm{tp}}^{\mathcal{P}}) - s(\boldsymbol{w}_{\mathrm{tp}}) \right]}_{\text{Systematic Error}}$$

Asymptotic Efficiency

$$s(\hat{\boldsymbol{w}}_{\text{tp}}^{\hat{\mathcal{P}}}) \approx_{N,T} s(\boldsymbol{w}_{\text{tp}})$$

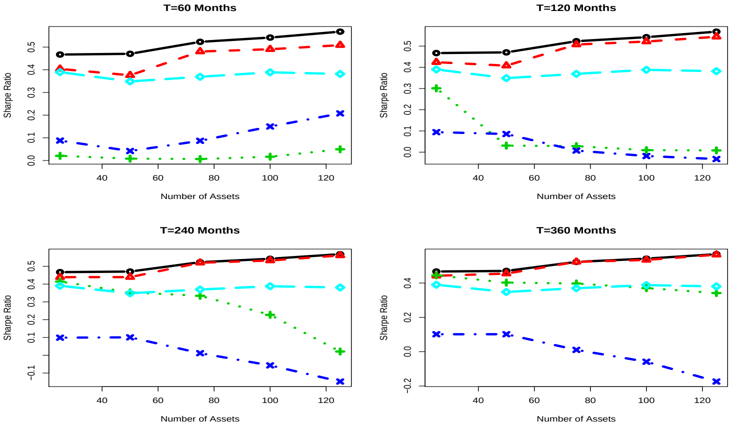
- Estimation Error I
  - Suffices to estimate $P_{\hat{\mathcal{P}}}^{\top} \Sigma P_{\hat{\mathcal{P}}}$ and $P_{\hat{\mathcal{P}}}^{\top} E$
  - Classical large sample analysis – $s(\hat{\boldsymbol{w}}_{\text{tp}}^{\hat{\mathcal{P}}}) - s(\boldsymbol{w}_{\text{tp}}^{\hat{\mathcal{P}}}) \approx_T 0$
- Estimation Error II
  - High dimensional data analysis – Large T and N

$$s(\boldsymbol{w}_{\text{tp}}^{\hat{\mathcal{P}}}) \approx_{N,T} s(\boldsymbol{w}_{\text{tp}}^{\mathcal{P}})$$

- Systematic Error – Effect of large market

$$s(\boldsymbol{w}_{\text{tp}}^{\mathcal{P}}) \approx_N s(\boldsymbol{w}_{\text{tp}})$$

Numerical Experiments



- 3 factor model calibrated
- Black – Tangency portfolio
- Red – Estimated subspace TP
- Green – Estimated TP
- Light blue – Naïve
- Blue – Estimated MV

Summary

- As the size of the investment universe increases, plug-in strategy becomes less efficient for implementing mean-variance portfolio
- Estimation error and systematic error trade-off can be balanced through subspace mean-variance analysis
- Under approximate factor model, it is possible to achieve asymptotic efficient portfolio selection even with a large number of assets