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

\[
\min_{\mathbf{w}} \mathbf{w}^\top \Sigma \mathbf{w} \quad \text{subject to } \mathbf{w}^\top \mathbf{E} = \mu, \quad \mathbf{w}^\top \mathbf{1} = 1
\]

- Investment universe
  \( \mathbf{r} \in \mathbb{R}^N \)
  \[
  \mathbf{E}(\mathbf{r}) = \mathbf{E} \quad \text{var}(\mathbf{r}) = \Sigma
  \]

- Tangency portfolio
  \[
  \mathbf{w}_{\text{tp}} = \frac{\Sigma^{-1} \mathbf{E}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{E}}
  \]

- Sharpe ratio
  \[
  s(\mathbf{w}_{\text{tp}}) = \left( \mathbf{E}^\top \Sigma^{-1} \mathbf{E} \right)^{1/2}
  \]
The Markowitz Optimization Enigma

\[ \mathbf{w}_{tp} = \frac{\mathbf{E}}{\mathbf{1}^\top \mathbf{\Sigma}^{-1} \mathbf{E}} \quad \implies \quad \hat{\mathbf{w}}_{tp} = \frac{\hat{\mathbf{E}}}{\mathbf{1}^\top \hat{\mathbf{\Sigma}}^{-1} \hat{\mathbf{E}}} \]

- Classical econometric theory

small \( N \) \( \implies \quad \hat{\mathbf{\Sigma}} \approx T \sum \quad \hat{\mathbf{E}} \approx T \mathbf{E} \)

\( \implies \quad \hat{\mathbf{w}}_{tp} \approx T \mathbf{w}_{tp} \implies s(\hat{\mathbf{w}}_{tp}) \approx T s(\mathbf{w}_{tp}) \)

- 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”

- “Naïve” diversification – DeMiguel, Garlappi and Uppal (2009)
Are we back to square one?

- Difficulty

  estimate $E, \Sigma \gg$ estimate $w_{tp} \gg$ achieve optimal Sharpe ratio

- Occam’s razor – when $N$ is large
  - “Impossibility” in accurate estimation of $E$ and $\Sigma$
  - “Extreme difficulty” in accurate estimation of $w_{tp}$
  - But it is possible to achieve optimal Sharpe ratio

- “Asymptotic” efficient portfolio selection

  $s(\hat{w}) \approx T,N \ s(w_{tp}) =: s_{opt}$
Subspace Mean-Variance Analysis
Subspace Mean-Variance Analysis

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

- Subspace “Tangency portfolio”

\[ w_{\text{tp}} = \frac{P_{\mathcal{P}} \left( P_{\mathcal{P}}^\top \Sigma P_{\mathcal{P}} \right)^{-1} P_{\mathcal{P}}^\top E}{\mathbb{1}^\top P_{\mathcal{P}} \left( P_{\mathcal{P}}^\top \Sigma P_{\mathcal{P}} \right)^{-1} P_{\mathcal{P}}^\top E} \]

- “Optimal” Sharpe ratio

\[ s(w_{\text{tp}}) = \left\{ E^\top P_{\mathcal{P}} \left( P_{\mathcal{P}}^\top \Sigma P_{\mathcal{P}} \right)^{-1} P_{\mathcal{P}}^\top E \right\}^{-1} \]

- Examples
  - \( \dim(\mathcal{P}) = 1 \) – Naïve diversification, value weighted portfolio, …
  - \( \dim(\mathcal{P}) = N \) – Tangency portfolio
**A Tale of Two Errors**

\[ s(\hat{\mathbf{w}}_{tp}^P) - s_{opt} = \underbrace{s(\hat{\mathbf{w}}_{tp}^P) - s(\mathbf{w}_{tp}^P)}_{\text{Estimation Error}} + \underbrace{s(\mathbf{w}_{tp}^P) - s_{opt}}_{\text{Systematic Error}} \]

- **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 $\hat{P}$

Construction

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

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

where $\mathbf{v}_k$ is the $k$th eigenvector of $\hat{\Sigma}$
- Construct the subspace tangency portfolio in $\hat{P}$:

$$\hat{w}_{\text{tp}} = \frac{P_{\hat{\mathbf{p}}} \left(P_{\hat{\mathbf{p}}}^T \hat{\Sigma} P_{\hat{\mathbf{p}}} \right)^{-1} P_{\hat{\mathbf{p}}}^T \hat{E}}{\mathbf{1}^T P_{\hat{\mathbf{p}}} \left(P_{\hat{\mathbf{p}}}^T \hat{\Sigma} P_{\hat{\mathbf{p}}} \right)^{-1} P_{\hat{\mathbf{p}}}^T \hat{E}}.$$

$\hat{P}_1 \subset \hat{P}_2 \subset \ldots \subset \hat{P}_N$
- $\hat{P}_1$ leads to naïve diversification
- $\hat{P}_N$ leads to sample tangency portfolio
Can we do better than 1/N?

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

**Structure of Asset Returns**

- **Approximate factor model**
  \[ r_{jt} = E_j + \beta_j f_1 t + \ldots + \beta_j K f_K t + \varepsilon_{jt}, \quad j = 1, \ldots, N; \quad t = 1, \ldots, T \]

- **Beta pricing relationship**
  \[ E_j = \alpha_j + \beta_j 1 \mu_1^f + \ldots + \beta_j K \mu_K^f, \quad j = 1, 2, \ldots, N \]

- Introduced by Chamberlain and Rothchild (1983)
- Detailed discussion in Connor, Goldberg and Korajczyk (2010)
- No arbitrage \( \iff \alpha^\top \alpha \) is bounded – Huberman (1982)
- Examples – CAPM, APT
- **Idiosyncratic noise**
  - Weakly temporal dependence – e.g., \( \alpha \)-mixing
  - Allow cross-section dependence – \( \Sigma_\varepsilon \) has eigenvalues bounded away from 0 and \( \infty \)
How to Choose $\mathcal{P}$

- **Best** choice – $\mathcal{P}^* = [\beta_1, \ldots, \beta_K]$
- **Second best** choice – $\mathcal{P} = \text{leading eigenspace of } \Sigma := \text{cov}(\mathbf{r})$
- **Feasible** choice – $\hat{\mathcal{P}} = \text{leading eigenspace of } \hat{\Sigma}$
- Estimated subspace tangency portfolio

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

- **Error analysis**

$$
s(\hat{\mathbf{w}}_{tp}) - s(\mathbf{w}_{tp}) = \left[ s(\hat{\mathbf{w}}_{tp}) - s(\mathbf{w}_{tp}) \right] + \left[ s(\mathbf{w}_{tp}) - s(\mathbf{w}_{tp}) \right] + \left[ s(\mathbf{w}_{tp}) - s(\mathbf{w}_{tp}) \right]
$$

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Error I</td>
<td>$s(\hat{\mathbf{w}}<em>{tp}) - s(\mathbf{w}</em>{tp})$</td>
</tr>
<tr>
<td>Estimation Error II</td>
<td>$s(\mathbf{w}<em>{tp}) - s(\mathbf{w}</em>{tp})$</td>
</tr>
<tr>
<td>Systematic Error</td>
<td>$s(\mathbf{w}<em>{tp}) - s(\mathbf{w}</em>{tp})$</td>
</tr>
</tbody>
</table>
3. Asymptotic efficiency

**Asymptotic Efficiency**

\[ s(\hat{w}_{tp}) \approx_{N,T} s(w_{tp}) \]

- **Estimation Error I**
  - Suffices to estimate \( P_\hat{\mu}^\top \Sigma P_\hat{\mu} \) and \( P_\hat{\mu}^\top E \)
  - Classical large sample analysis – \( s(\hat{w}_{tp}^\hat{\mu}) - s(w_{tp}^\hat{\mu}) \approx_T 0 \)

- **Estimation Error II**
  - High dimensional data analysis – Large \( T \) and \( N \)

\[ s(w_{tp}^\hat{\mu}) \approx_{N,T} s(w_{tp}^\mu) \]

- **Systematic Error – Effect of large market**

\[ s(w_{tp}^\mu) \approx_N s(w_{tp}) \]
3. Asymptotic efficiency

Numerical Experiments

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

**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.