Mean-variance portfolio optimization when means and covariances are estimated

Zehao Chen

June 1, 2007

Joint work with Tze Leung Lai (Stanford Univ.) and Haipeng Xing (Columbia Univ.)
Outline

1. Introduction and review
   - The Markowitz framework
   - Different efficient frontier definitions under stochastic setting

2. A high dimensional plug-in covariance matrix estimator
   - The $L^2$ boosting estimator
   - Simulation and empirical study

3. A modified Markowitz framework
   - The framework
   - An example and simulation

4. Conclusion
The basic formulation

We denote the returns of \( p \) risky assets (e.g. stock returns) by a \( p \times 1 \) vector \( R \), and an unobserved future return by \( r \),

\[
E(R) = \mu, \quad \text{Cov}(R) = \Sigma
\]

The mean-variance optimization solves for an asset allocation \( w \), which minimizes the portfolio risk \( \sigma_w^2 \), while achieving a certain target return \( \mu^* \), i.e.,

\[
\min_w w^T \Sigma w = \min_w \sigma_w^2
\]

subject to

\[
w^T \mu \geq \mu^*, \quad w_1 + w_2 + \ldots + w_p = 1
\]
The basic formulation

This formulation has a closed form solution for $w$,

$$w^* = f(\mu, \mu^*, \Sigma^{-1})$$

The weights sometimes have additional constraints, i.e.

$$l_i \leq w_i \leq u_i, \quad i = 1, \ldots, p$$

If $l_i \geq 0$, the constraint is also called no short selling constraint. The solution under this additional constraint requires quadratic programming.
Efficient frontier
For $w = w(X, \mu^*)$ and reasonable $\mu$ and $\Sigma$ estimates $\hat{\mu}(X)$ and $\hat{\Sigma}(X)$,

$$\min_w w^T \hat{\Sigma} w$$

subject to

$$w^T \hat{\mu} \geq \mu^*$$

Define the "plug-in" efficient frontier as parametrized by $\mu^*$

$$C(\mu^*) = (w^T \Sigma w, w^T \mu)$$
"Plug-in" covariance estimates

- Factor model (with domain knowledge)

\[ R = \alpha + BF + \epsilon \]

\[ \Sigma = W_1 + W_2 = B\Omega B^T + \text{Cov}(\epsilon) \]


**"Plug-in" covariance estimates**

- **Factor model (with domain knowledge)**
  \[ R = \alpha + BF + \epsilon \]
  \[ \Sigma = W_1 + W_2 = B\Omega B^T + \text{Cov}(\epsilon) \]

- **Shrinkage estimator**
  \[ \alpha \hat{F} + (1 - \alpha) \hat{\Sigma} \]
Random efficient frontier?

Figure: $n = 50, p = 5$ i.i.d multivariate normal
Random efficient frontier?

- The data dependent efficient frontier is not well defined, and is a random curve.
Random efficient frontier?

- The data dependent efficient frontier is not well defined, and is a random curve.
- By plugging the estimates of mean $\hat{\mu}$, it’s essentially constraining on

$$w^T \hat{\mu} \geq \mu^*$$
Random efficient frontier?

- The data dependent efficient frontier is not well defined, and is a random curve.
- By plugging the estimates of mean $\hat{\mu}$, it’s essentially constraining on
  \[ w^T \hat{\mu} \geq \mu^* \]
- Conceptually, one should constrain on
  \[ \mathbb{E}(w^T r) \geq \mu^* \]
For bootstrapped samples $X^*_1, \ldots, X^*_B$ of $X$, 

$$w_M(X, \mu^*) = \frac{1}{B} \sum_{i=1}^{B} w(X^*_i, \mu^*)$$

Define the resampled efficient frontier as 

$$C_M(\mu^*) = (w_M^T \Sigma w_M, w_M^T \mu)$$
Previous definitions of efficient frontiers

"Plug-in" efficient frontier

\[
\min_w w^T \hat{\Sigma} w, \quad w^T \hat{\mu} \geq \mu^*
\]

\[C(\mu^*) = (w^T \Sigma w, w^T \mu)\]

Resampled efficient frontier

\[w_M(X, \mu^*) = \frac{1}{B} \sum_{i=1}^{B} w(X_i^*, \mu^*)\]

\[C_M(\mu^*) = (w_M^T \Sigma w_M, w_M^T \mu)\]
If one really wants to use the plug-in

- Estimate $\Sigma$ or $\Sigma^{-1}$?
If one really wants to use the plug-in

- Estimate $\Sigma$ or $\Sigma^{-1}$?
- Employ a sparsity assumption (in practice, residuals from some factor model): reduce the number of parameters to estimate.
If one really wants to use the plug-in

- Estimate $\Sigma$ or $\Sigma^{-1}$?
- Employ a sparsity assumption (in practice, residuals from some factor model): reduce the number of parameters to estimate.
- Impose proper weight constraints.
A high dimensional covariance estimator

- Use Modified Cholesky decomposition $\Sigma^{-1} = T'D^{-1}T$ to reparametrize covariance matrix, and enforce sparsity on $T$. 
A high dimensional covariance estimator

- Use Modified Cholesky decomposition $\Sigma^{-1} = T’D^{-1}T$ to reparametrize covariance matrix, and enforce sparsity on $T$.
- Use coordinate-wise greedy (component-wise $L^2$ boosting) with a modified BIC stopping criterion. Can be shown to be a consistent estimator.
A high dimensional covariance estimator

- Use Modified Cholesky decomposition $\Sigma^{-1} = T' D^{-1} T$ to reparametrize covariance matrix, and enforce sparsity on $T$.
- Use coordinate-wise greedy (component-wise $L^2$ boosting) with a modified BIC stopping criterion. Can be shown to be a consistent estimator.
- Spectral norm convergence for both $\Sigma$ and $\Sigma^{-1}$.
Modified Cholesky decomposition

For a random vector $Y = (y_1, y_2, \ldots, y_p)$, one can write them in the "auto-regressive" form, for $k \geq 2$

$$y_k = \mu_k + \phi_{k,1}y_1 + \phi_{k,2}y_2 + \ldots + \phi_{k,k-1}y_{k-1} + \epsilon_k$$

Let $T$ be a $p \times p$ unit lower triangular matrix, with $-\phi_{i,j}$ ($i < j$) on the lower triangular. And let $D$ be the $p \times p$ diagonal matrix

$$D = \text{diag}(\text{var}(y_1), \text{var}(\epsilon_2), \text{var}(\epsilon_3), \ldots, \text{var}(\epsilon_p))$$

The Modified Cholesky decomposition is

$$\Sigma^{-1} = T'D^{-1}T$$
The main ideas is to iteratively look for the predictor which is mostly correlated with the current model residual. Include this predictor and re-calculate the residual.
The coordinate-wise greedy algorithm

- The main idea is to iteratively look for the predictor which is mostly correlated with the current model residual. Include this predictor and re-calculate the residual.
- This process will generate a sequence of nested models $M_1 \subseteq M_2 \subseteq \ldots \subseteq M_{m_n}$. 
The main ideas is to iteratively look for the predictor which is mostly correlated with the current model residual. Include this predictor and re-calculate the residual.

This process will generate a sequence of nested models $M_1 \subseteq M_2 \subseteq \ldots \subseteq M_{m_n}$.

BIC: $\log \hat{\sigma}^2 + \frac{\log n}{n} \cdot \#\{M_k\}$
The coordinate-wise greedy algorithm

- The main ideas is to iteratively look for the predictor which is mostly correlated with the current model residual. Include this predictor and re-calculate the residual.
- This process will generate a sequence of nested models $M_1 \subseteq M_2 \subseteq \ldots \subseteq M_{m_n}$.
- BIC: $\log \hat{\sigma}^2 + \frac{\log n}{n} \cdot \#\{M_k\}$
- Proposed Modified BIC: $n\hat{\sigma}^2 + \hat{\sigma}^2 \log p \log n \cdot \#\{M_k\}$
Convergence results

Definition of spectral norm:

\[ \|A\|_2 = \sqrt{\lambda_{\text{max}}(A'A)} \]

Convergence:

\[ \|\hat{\Sigma} - \Sigma\|_2 \to 0 \]
\[ \|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_2 \to 0 \]
An example

**Setting**: $n = 300, p = 600$, number of nonzero parameters in $T$ is 900 (i.e. 3 per row), and $T$ elements uniformly distributed in $[0.5, 1]$. 
MSE to true inverse covariance

![MSE to true inverse covariance graph](image)

Zehao Chen

M.V. optimization when means and covariances are estimated
The $L^2$ boosting estimator

Simulation and empirical study

MSE to true covariance

![Graph showing MSE to true covariance for different algorithms.](image)
An empirical example: NASDAQ-100, 1990-2006

Figure: Empirical performance curve with S&P500 index as market factor
Efficient frontier when means and covariances are unknown

For \( w(X, \mu^*) \), solve

\[
\min_w \text{var}(w^T r)
\]

\[
E(w^T r) \geq \mu^*
\]

Define the expected efficient frontier as

\[
C(\mu^*) = E_X \{(w^T \Sigma w, w^T \mu)\}
\]
A comparison of efficient frontiers

"Plug-in" efficient frontier

$$\min_w w^T \hat{\Sigma} w, \quad w^T \hat{\mu} \geq \mu^*, \quad C(\mu^*) = (w^T \Sigma w, w^T \mu)$$

Resampled efficient frontier

$$w_M(X, \mu^*) = \frac{1}{B} \sum_{i=1}^{B} w(X_i^*, \mu^*), \quad C_M(\mu^*) = (w_M^T \Sigma w_M, w_M^T \mu)$$

Expected efficient frontier

$$\min_w \text{var}(w^T r), \quad \mathbb{E}(w^T r) \geq \mu^*, \quad C(\mu^*) = \mathbb{E}_X \{ (w^T \Sigma w, w^T \mu) \}$$
Solving for the expected efficient frontier

Introduce a risk-aversion parameter $\lambda$ and reformulate the problem as

$$\min_w -\mathbb{E}(w^T r) + \lambda \text{var}(w^T r)$$
Solve for a portfolio weight $w(X, \lambda)$

$$-E(w^T r) + \lambda \text{var}(w^T r) = -Ew^T r + \lambda E(w^T r)^2 - \lambda E^2(w^T r)$$
A not-so-bayesian stochastic control approach

- Solve for a portfolio weight $w(X, \lambda)$

$$-\mathbb{E}(w^T r) + \lambda \text{var}(w^T r) = -\mathbb{E}w^T r + \lambda \mathbb{E}(w^T r)^2 - \lambda \mathbb{E}^2(w^T r)$$

- However, one needs the following conditional form,

$$\mathbb{E}_X[-w^T \mathbb{E}(r|X) + \lambda w^T \mathbb{E}(rr^T|X)w - \lambda w^T \mathbb{E}(??|X)w]$$
This conditional form entitles us to plug in reasonable guesses for $E(r|X)$, $E(rr^T|X)$ and $E(??|X)$. 
This conditional form entitles us to plug in reasonable guesses for $E(r|X)$, $E(rr^T|X)$ and $E(??|X)$.

Use bayesian way to develop a procedure and approximate it in frequentist setting.
Embedding technique:
This bayesian formulation can be shown to be equivalent to

$$E_X[-\eta w^T E(r|X) + \lambda w^T E(r r^T|X) w]$$

where $$\eta = 1 + 2\lambda E((w^*_\lambda)^T r)$$ and $$w^*_\lambda$$ is the solution to the original optimization problem with risk aversion $$\lambda$$.

---

$^1$X. Y. Zhou and D. Li (2000)
How to proceed? A simple example

Take prior $\Sigma \sim IW(\Phi, \nu)$, so

$$E(\Sigma|X) = \alpha \Phi + (1 - \alpha) \hat{\Sigma}$$

$$E(r|X) = E(\mu|X) \approx \bar{X}$$

$$E(rr^T|X) = E(\Sigma|X) + \text{var}(\mu|X, \Sigma) + E(r|X)E(r|X)^T$$

$$\approx \frac{n+1}{n} [\alpha \Phi + (1 - \alpha) \hat{\Sigma}] + \bar{X}\bar{X}^T$$

"Semi-Empirical Bayes": estimate $\Phi$ by $\text{diag}(\hat{\Sigma})$
How to proceed? A simple example

Define a parametrized function $w(X, \lambda, \alpha, \eta)$ that solves

$$-\eta w^T E(r|X) + \lambda w^T E(rr^T|X) w$$

with $E(r|X)$ and $E(rr^T|X)$ replaced by approximations.
How to proceed? A simple example

Having a parametrized procedure \( w(X, \lambda, \alpha, \eta) \), we wish to evaluate

\[
R(\lambda, \alpha, \eta) = -\mathbb{E}w^T r + \lambda \text{var}(w^T r) \\
= \mathbb{E}_X[-\mathbb{E}(w^T r|X) + \lambda \text{var}(w^T r|X)] + \lambda \text{var}[\mathbb{E}(w^T r|X)]
\]
How to proceed? A simple example

Having a parametrized procedure \( w(X, \lambda, \alpha, \eta) \), we wish to evaluate

\[
R(\lambda, \alpha, \eta) = -Ew^T r + \lambda \text{var}(w^T r) \\
= EX [ -E(w^T r|X) + \lambda \text{var}(w^T r|X) ] + \lambda \text{var}[E(w^T r|X)]
\]

Bootstrap

Bootstrap: \( \{X_1^*, X_2^*, ..., X_B^*\} \)
How to proceed? A simple example

The above can be approximated by the bootstrapped empirical risk,

\[
\hat{R}(\lambda, \alpha, \eta) = \tilde{E}_X[-E(w^T r|X)] + \lambda \tilde{E}_X[var(w^T r|X)] + \lambda \tilde{var}[E(w^T r|X)]
\]

\[
\tilde{E}_X[-E(w^T r|X)] = \frac{1}{B} \sum_{i=1}^{B} [-w^T (X_i^*, \lambda, \alpha, \eta) \hat{\mu}(X_i^*)]
\]

\[
\tilde{E}_X[var(w^T r|X)] = \frac{1}{B} \sum_{i=1}^{B} [w^T (X_i^*, \lambda, \alpha, \eta) \hat{\Sigma}(X_i^*) w(X_i^*, \lambda, \alpha, \eta)]
\]

\[
\tilde{var}[E(w^T r|X)] = var[w^T (X_i^*, \lambda, \alpha, \eta) \hat{\mu}(X_i^*)]
\]

where \( \hat{\mu}(X_i^*) \) and \( \hat{\Sigma}(X_i^*) \) are reasonable estimates for mean and covariance matrix of \( X_i^* \).
How to proceed? A simple example

Solve for

$$\min_{\alpha, \eta} \hat{R}(\lambda, \alpha, \eta)$$

with reasonable initial values.
Example: $n = 50, p = 5$
Example: \( n = 50, \ p = 30 \)
Example: \( n = 50, p = 50 \)

![Graph showing return vs. standard deviation with different methods: SBF, Shrinkage, HCD.](image)
The proposed high dimensional covariance estimator is appropriate in sparse setting, and has better performance.
Conclusion

- The proposed high dimensional covariance estimator is appropriate in sparse setting, and has better performance.
- The modified Markowitz framework tries to optimize on the expected efficient frontier. It’s a general framework and has better performance in many practical scenarios.