Lecture 1. Mean-Variance Optimization Theory: An Overview

Outline

- Chapter 3 of *Statistical Models and Methods for Financial Markets*.
- The mean-variance portfolio optimization theory of Markowitz (1952, 1959) is widely regarded as one of the major theories in financial economics.
- It is a single-period theory on the choice of portfolio weights that provide optimal tradeoff between the mean and the variance of the portfolio return for a future period.

Framework

- A portfolio consisting of p assets

$$\begin{aligned} P_t : &= \text{the value of the portfolio at time } t, \\ w_i : &= \text{the weight of the portfolio value invested in asset } i, \\ P_{it} : &= w_i P_t = \text{the value of asset } i, \\ r_{it} : &= (P_{it} - P_{i,t-1})/P_{i,t-1}, \\ \mathbf{r} : &= (r_{1t}, \ldots, r_{pt})^T, \\ \boldsymbol{\mu} : &= E(\mathbf{r}_t), \quad \boldsymbol{\Sigma} = \text{Cov}(\mathbf{r}_t), \\ \mathbf{w} : &= (w_1, \ldots, w_p)^T, \quad \mathbf{1} = (1, \ldots, 1)^T, \end{aligned}$$

- The mean and the variance of the portfolio return:

$$(\mathbf{w}^T \boldsymbol{\mu}, \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}).$$

Markowitz portfolio optimization

Given a target value μ_* for the mean return of a portfolio, Markowitz characterizes an efficient portfolio by

$$\mathbf{w}_{\text{eff}} = \arg\min_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \quad \text{subject to } \mathbf{w}^T \boldsymbol{\mu} = \mu_*, \ \mathbf{w}^T \mathbf{1} = 1.$$

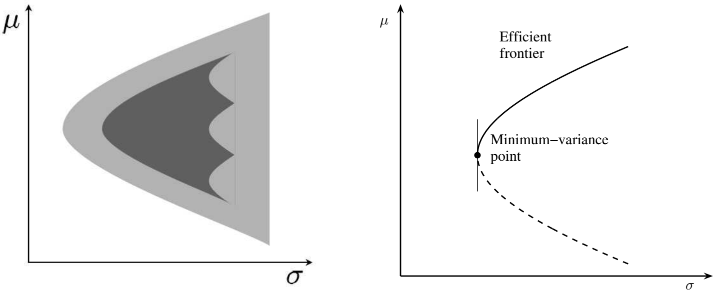
where μ_* is the target return.

- When short selling is allowed,

  $$\mathbf{w}_{\text{eff}} = \left\{ B\boldsymbol{\Sigma}^{-1}\mathbf{1} - A\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mu_* \left(C\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - A\boldsymbol{\Sigma}^{-1}\mathbf{1} \right) \right\} \Big/ D,$$

  where $A = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, $B = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, $C = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$, and $D = BC - A^2$.
- When short selling is not allowed ($\mathbf{w} \geq \mathbf{0}$), $\mathbf{w}_{\text{eff}}$ is solved via quadratic programming.

Feasible region and efficient frontier



The "plug-in" principle

- A natural idea is to replace the mean and covariance matrix by their sample estimates.
- However, Different studies have documented that sample estimates are not as effective as other estimates.
  - The efficient portfolio based on sample estimates may not be as effective as an equally weighted portfolio. (Frankfurther et al, 1971; Korkie, 1980)
  - The mean-variance portfolio based on sample esimates has serious deficiencies, in practice, often called “Markowitz optimization enigma” (Michaud, 1989; Best & Brauer, 1991; Chopra et al, 1993; Canner et al, 1997; Simann, 1997; Britten-Jones, 1999).

APT and multifactor pricing models

Multifactor pricing models relate the p asset returns r_i to k factors $f_1, \ldots, f_k$ in a regression model of the form

$$r_i = \alpha_i + (f_1, \ldots, f_k)^T \boldsymbol{\beta}_i + \epsilon_i,$$

in which α_i and $\boldsymbol{\beta}_i$ are unknown regression parameters and ϵ_i is an unobserved random disturbance that has mean 0 and is uncorrelated with $\mathbf{f} := (f_1, \ldots, f_k)^T$.

- Arbitrage pricing theory (APT), introduced by Ross (1976), relates the expected return μ_i of the ith asset to the risk-free return, or to a more general parameter λ_0 without assuming the existence of a risk-free asset, and to a $k \times 1$ vector $\boldsymbol{\lambda}$ of risk premiums:

$$\mu_i \approx \lambda_0 + \boldsymbol{\beta}_i^T \boldsymbol{\lambda}, \qquad i = 1, \ldots, p, \tag{1}$$

Captial asset pricing model (CAPM)

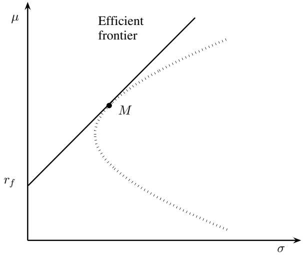


Figure: Minimum-variance portfolios of risky assets and a risk-free asset. Left panel: short selling is allowed. Right panel: short selling is not allowed.

- The one-fund theorem: There is a single fund M of risky assets such that any efficient portfolio can be constructed as a linear combination of the fund M and the risk-free asset.

Captial asset pricing model (CAPM)

- Sharpe ratio: For a portfolio whose return has mean μ and variance σ^2, its *Sharpe ratio* is $(\mu - r_f)/\sigma$, which is the expected excess return per unit of risk.
- The beta, denoted by β_i, of risky asset i that has return r_i is defined by $\beta_i = \text{Cov}(r_i, r_M)/\sigma^2$.
- CAPM: The CAPM relates the expected excess return $\mu_i - r_f$ of asset i to its beta via

$$\mu_i - r_f = \beta_i(\mu_M - r_f).$$

- The alpha (or Jensen index) is the α in the generalization of CAPM to $\mu - r_f = \alpha + \beta(\mu_M - r_f)$.

Choice of factors

- The economic approach specifies
  - macroeconomic and financial market variables that are thought to capture the systematic risks of the economy (Chen, Roll and Ross, 1986).
  - characteristics of firms that are likely to explain differential sensitivity to the systematic risks, forming factors from portfolios of stocks based on the characteristics (Fama & French, 1992, 1993).
- The statistical approach includes principal component analysis and factor analysis. (Section 3.4.2 of Chapter 3)

Bootstrapping and resampled frontier

- To correct for the bias of $\widehat{\mathbf{w}}_{\text{eff}}$, Michaud (1989) proposed to use the average of the bootstrap weights

$$\bar{\mathbf{w}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\mathbf{w}}_b^*, \qquad (*)$$

where $\widehat{\mathbf{w}}_b^*$ is the estimated optimal portfolio weight vector based on the bth bootstrap sample $\{\mathbf{r}_{b1}^*, \ldots, \mathbf{r}_{bn}^*\}$ drawn with replacement from the observed sample $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$.

- Michaud's resampled efficiency corresponds to the plot $\sqrt{\bar{\mathbf{w}}^T\widehat{\boldsymbol{\Sigma}}\bar{\mathbf{w}}}$ versus $\bar{\mathbf{w}}^T\bar{\mathbf{r}} = \mu_*$ for a fine grid of μ_* values.
- Although Michaud claims that (*) provides an improvement over $\widehat{\mathbf{w}}_{\text{eff}}$, there have been no convincing theoretical developments and simulation studies to support the claim.

Bayes Procedure

In statistical decision theory, the problem of estimation and hypothesis testing have the following ingredients:

- a parameter space Θ and a family of distribution $\{P_\theta, \theta \in \Theta\}$;
- data $(X_1, \ldots, X_n) \in \mathcal{X}$ sampled from the distribution P_θ when θ is the true parameter, where $\mathcal{X}$ is called the "sample space";
- an action space $\mathcal{A} = \{$ all available actions to be chosen $\}$;
- a loss function $L : \Theta \times \mathcal{A} \to [0, \infty)$ representing the loss $L(\theta, a)$ when θ is the parameter value and action a is taken.

A *statistical decision rule* is a function $d : \mathcal{X} \to \mathcal{A}$ that takes action $d(\mathbf{X})$ when $\mathbf{X} = (X_1, \ldots, X_n)$ is observed. Its performance is evaluated by the *risk function*

$$R(\theta, d) = E_\theta L\big(\theta, d(\mathbf{X})\big), \qquad \theta \in \Theta.$$

Given a prior distribution π on Θ, the *Bayes risk* of a statistical decision rule d is $B(d) = \int R(\theta, d) d\pi(\theta)$. A *Bayes rule* is a statistical decision rule that minimizes the Bayes risk.

Bayes and shrinkage estimators

$$\mathbf{r}_t \text{ i.i.d. } N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

- Bayes estimate: Denote m the number of assets and IW the inverted Wishart distribution. Consider the prior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\nu}, \boldsymbol{\Sigma}/\kappa), \qquad \boldsymbol{\Sigma} \sim IW_m(\boldsymbol{\Psi}, n_0),$$

the posterior distribution of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ given $\mathbf{r}_1, \ldots, \mathbf{r}_n$ is given by

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N\left(\frac{n\bar{\mathbf{r}} + \kappa\boldsymbol{\nu}}{n+\kappa}, \frac{\boldsymbol{\Sigma}}{n+\kappa}\right),$$

$$\boldsymbol{\Sigma} \sim IW_m\left(\boldsymbol{\Psi} + \sum_{i=1}^{n}(\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})^T + \frac{n\kappa}{n+\kappa}(\bar{\mathbf{r}} - \nu)(\bar{\mathbf{r}} - \nu)^T, n + n_0\right).$$

Bayes and shrinkage estimators

- Shrinkage estimators:

$$\widehat{\boldsymbol{\Sigma}} = \delta \mathbf{F} + (1 - \delta)\mathbf{S},$$

  where **S** is the MLE of $\boldsymbol{\Sigma}$, **F** is a structured covariance matrix, and δ is the shrinkage parameter.
    - Ledoit and Wolf (2003, 2004): How to choose an *optimal shrinkage constant* δ^*.
    - **F** can be estimated via multifactor models, or assigned certain structure (e.g., constant correlation).
- The estimators have relatively small "estimation error" in comparison with the MLE of $\boldsymbol{\Sigma}$.