Asset Pricing Review Session 2

Daojing Zhai

Yale University

November 15, 2024

Outlines for review sessions

First Part (Static Asset Pricing)

1. Choice under uncertainty
2. Static portfolio choice
3. Static asset pricing
4. Stochastic discount factor

Second Part (Intertemporal Asset Pricing)

1. Present value relations
2. Long run risk (BY)
3. Intertemporal CAPM (CV)
4. Rare Disaster (Martin)
5. Stochastic volatility (BKY and CGPT)
6. Intertemporal portfolio choice
7. Term structure & bond pricing

Bernoulli Utility (1/2)

Problem 6: An individual has Bernoulli utility function $u(\cdot)$ and initial wealth w. Let lottery L offer a payoff of G with probability p and a payoff B with probability $1-p$.

- If the individual owns the lottery, what is the minimum price he would sell it for?

$$\underbrace{pu(w+G)+(1-p)u(w+B)}_{\text{if hold}} = \underbrace{u\left(w+P_s\right)}_{\text{if sell}}.$$

- If he does not own it, what is the maximum price he would be willing to pay for it?

$$\underbrace{pu(w+G-P_b)+(1-p)u(w+B-P_b)}_{\text{if buy}} = \underbrace{u\left(w\right)}_{\text{don't buy}} \quad .$$

Bernoulli Utility (2/2)

Problem 6: $\{(G, p), (B, 1-p)\}$

- Are buying and selling prices equal?
  - Denote $w^* = w - P_b$,

$$pu(w+G) + (1-p)u(w+B) = u\left(w + P_s\right)$$
$$pu(w^*+G) + (1-p)u(w^*+B) = u\left(w^* + P_b\right)$$

  - Bernoulli utility function: decreasing absolute risk aversion → more comfortable taking risks when you are richer → less costly (low risk premium) holding the lottery when you are richer (at w) → $P_s > P_b$
  - CARA utility (exponential utility) → constant risk attitude to risk no matter you are rich or poor → $P_s = P_b$

Probability Premium

"Risk premium" in the probability dimension.
For any fixed amount of money x and a positive number ϵ, the probability premium $\pi(x,\epsilon)$, is the excess in wining probability over fair odds that makes the individual indifferent between the certain outcome x and a gamble between the two outcomes $x+\epsilon$ and $x-\epsilon$. That is,

$$u(x) = \left(\frac{1}{2} + \pi(x,\epsilon)\right) u(x+\epsilon) + \left(\frac{1}{2} - \pi(x,\epsilon)\right) u(x-\epsilon)$$

For example, in question 7b: $x = 10, \epsilon = 6, u(x) = \sqrt{x}$, probability premium π subject to

$$u(10) = \left(\frac{1}{2} + \pi\right) u(16) + \left(\frac{1}{2} - \pi\right) u(4)$$

Mean Variance Efficiency

Settings

- N risky assets, w_i (weight), R_i, mean return $\bar{R}_i$
- Return vector $R = \{R_i\}_{i \in N}$, covariance matrix Σ, mean return vector $\bar{R} = \{\bar{R}_i\}_{i \in N}$
- A risk-free asset, R_f
- Want to form a portfolio $w = \{w_i\}_{i \in N}$ with lowest variance, targeting mean return $\bar{R}_p$

Solution

- solve $\min_w \frac{1}{2} w' \Sigma w$ s.t. $\bar{R}_p - R_f = \left(\bar{R} - R_f 1\right)' w$
- FOC gives [Campbell book section 2.2.6.]

$$w = \frac{(\bar{R}_p - R_f)' \Sigma^{-1} (\bar{R} - R_f 1)}{(\bar{R} - R_f 1)' \Sigma^{-1} (\bar{R} - R_f 1)}$$

- Apply mean-variance efficiency in practice? Practical difficulties:
  - N means, $N(N-1)/2$ (co)variances to estimate (not constant over long periods of time)
  - 2000 stocks $\rightarrow$ 2 million (co)variances
  - $N \geq T \rightarrow$ sample covariance-covariance matrix can be non-singular.
- need a shortcut! $\rightarrow$ CAPM & APT

CAPM Assumptions

Key: Find the mean-variance efficient portfolio under equilibrium argument

Preference and belief

- All investors are mean variance optimizers (preference)
- All investors care about returns measured over one period (static)
- All investors perceive the same means, variances, and covariances for returns (no asymmetric info)

Technology

- Investors can borrow or lend at a given risk-free interest rate
  - If both, Sharpe-Lintner version (here)
  - If cannot borrow, Black version
- There are no non-traded assets, taxes, or transaction costs (frictionless)

Market structure

- All investors are price takers (competitive market)

CAPM Derivation

- Mean variance efficiency: fix return $\bar{R}_p$, minimize variance $\mathrm{Var}(R_p)$.
- For a mean variance efficient portfolio $w = \{w_i\}_{i \in N}$, adjust w_i and w_j

$$\mathrm{d}\bar{R}_p = \left(\bar{R}_i - R_f\right)\mathrm{d}w_i + \left(\bar{R}_j - R_f\right)\mathrm{d}w_j$$

$$\mathrm{d}\,\mathrm{Var}(R_p) = 2\,\mathrm{Cov}(R_i, R_p)\,\mathrm{d}w_i + 2\,\mathrm{Cov}(R_j, R_p)\,\mathrm{d}w_j$$

- Set $d\bar{R}_p = 0$. By efficiency, we have $\mathrm{d}\,\mathrm{Var} = 0$

$$\mathrm{d}\bar{R}_p = 0 \Rightarrow \mathrm{d}w_j = -\frac{\left(\bar{R}_i - R_f\right)}{\left(\bar{R}_j - R_f\right)}\mathrm{d}w_i$$

$$\Rightarrow \frac{\bar{R}_i - R_f}{\mathrm{Cov}(R_i, R_p)} = \frac{\bar{R}_j - R_f}{\mathrm{Cov}(R_j, R_p)}$$

- In equilibrium, the demand for assets must equal the supply of assets. It requires the *market portfolio* (or *value-weighted index*) must be mean-variance efficiency

$$\bar{R}_i - R_f = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)}\left(\bar{R}_m - R_f\right) = \beta_{im}\left(\bar{R}_m - R_f\right)$$

Arbitrage Pricing Theory

- **Key**:
  - N is too large for portfolio choice?
  - Assume a **factor structure** in the variance-covariance matrix of risky returns
  - reducing the dimension of portfolio choice problem from N to K, where K is the number of factors.
- We start with a single-factor model, use APT theory to show that single-factor model comes to the CAPM world.
- Then generalize it to a multifactor model.

A Single-Factor Model

- Suppose there is a single-factor model (or **market model** as asset returns only comove with the market returns). That is, you run the regression:

$$R_{it} - R_f = \alpha_i + \beta_{im}(R_{mt} - R_f) + \epsilon_{it}.$$

- Suppose further the residual risk in any stock is idiosyncratic

$$E[\epsilon_{it}\epsilon_{jt}] = 0$$

- Then consider forming a portfolio of N assets i. Its excess return

$$R_{pt} - R_f = \alpha_p + \beta_{pm}(R_{mt} - R_f) + \epsilon_{pt}$$

  where $\alpha_p = \sum_{i=1}^N w_i\alpha_i, \beta_{pm} = \sum_{i=1}^N w_i\beta_{im}, \epsilon_{pt} = \sum_{i=1}^N w_i^2\epsilon_{it}$
- Note: the residual risk $\text{Var}(\epsilon_{pt})$ is negligible when N is large

$$\text{Var}(\epsilon_{pt}) = \sum_{i=1}^N w_i^2 \, \text{Var}(\epsilon_{it}) \to 0$$

- Also: it must $\alpha_p = 0$ or else there is an arbitrage opportunity:
  - hold 1 unit of portfolio p and short β_{pm} unit of market portfolio $\to$ riskless profit α_p
- That is, it must be CAPM!

$$R_{pt} - R_f = \alpha_p + \beta_{pm}(R_{mt} - R_f)$$

Multifactor model

- Suppose there are K "factor portfolios" capturing the common influence of K underlying sources of risk

$$R_{it} - R_f = \alpha_i + \sum_{k=1}^{K} \beta_{ik}(R_{kt} - R_f) + \epsilon_{it}.$$

- Follow the same logic as in the single-factor model, it must be $\alpha_i = 0$ for almost all stocks.
- A more common representation: measure the factors directly as mean-zero shocks (for example, innovations to macroeconomic variables), then we have

$$R_{it} - R_f = \mu_i + \sum_{k=1}^{K} \beta_{ik} F_{kt} + \epsilon_{it}, \quad \mu_i = \sum_{k=1}^{K} \beta_{ik}\lambda_k.$$

where $\lambda_k = E[R_{kt} - R_f]$ and $F_{kt} = R_{kt} - R_f - \lambda_k$

- How to find a factor? Fama and French returns-based model!
    - suppose find some stocks that **systematically** give a high return on average (e.g., small stocks)
    - the portfolio must be highly exposed to some risk factor
- Most recently proposed reduced-form factor SDFs:
    - Hou et al. (2015): 4 factors
    - Fama and French (2016): 5 factors

Generalized Method of Moments

Modern Formulation

Hansen (1982) generalizes, derives asymptotic properties and optimal W

Data	Moment function	Weighting matrix	True value
$\underbrace{y_t}_{h\times 1}, \underbrace{Y_T}_{Th\times 1}$	$\underbrace{h(\theta, y_t)}_{r\times 1}$	$\underbrace{W_t}_{r\times r}$	$\underbrace{\theta_0}_{a\times 1}$

- Sample average of moment function:

$$g_T(\theta, Y_T) = \frac{1}{T}\sum_t h(\theta, y_t)$$

- The GMM principle: Make sample moment g as small as possible
- GMM estimator $\hat{\theta}_T$ defined as minimizer of

$$g_T(\theta, Y_T)' W_T g_T(\theta, Y_T) \equiv Q(\theta, Y_t)$$

- $r > a$: “overidentified” case

Generalized Method of Moments

Simple Example

Returning to Student's t example,

- estimator ν,
- weighting matrix $W_t = I_{2\times 2}$
- moment function

$$h(\theta, y_t) = \begin{bmatrix} y_t^2 - \frac{\nu}{\nu-2} \\ y_t^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix}$$

Optimal weight: inverse asymptotic covariance matrix of moments

$$S \equiv \lim_{T\to\infty} \mathrm{E}\left[g_T(\theta_0, Y_T)\, g_T(\theta_0, Y_T)'\right]$$

$$Q(\theta, Y_t) = g_T(\theta, Y_T)'\, S^{-1} g_T(\theta, Y_T)$$

- Intuition: Overweight precise moments, underweight noisy moments
- Challenge: Want weight S^{-1} estimate θ, but need θ_0 to estimate S!
- Solution: Estimate S with $\hat{S}_T = \frac{1}{T}\sum_t h(\theta, y_t)\, h(\theta, y_t)'$

Generalized Method of Moment

Procedure: two-step estimation

1. Estimate $\hat{\theta}$ using $W_T = I$
2. Find optimal weighting matrix $\hat{S}_T$ with $\hat{\theta}$ estimated in step 1:

$$\hat{S}_T = \frac{1}{T} \sum_t h(\theta, y_t) \, h(\theta, y_t)' \xrightarrow{p} S$$

Then re-estimate $\hat{\theta}$ using $W_T = \hat{S}_T^{-1}$

Generalized Method of Moment

Hints on Question 3 (follow the notation of Cochrane *Asset Pricing*)

- Parameters to be estimate $\mathbf{b} = (a, b, c)$
- 10 moment conditions $E[M_t(1 + R_{it}) - 1] = 0 \rightarrow$ FOC in close-form
- The standard errors of the parameter estimates using prespecified weighting matrices (Cochrane Section 11.5):

$$\mathrm{var}(\hat{b}) = \frac{1}{T}\left(d'Wd\right)^{-1} d'WSWd \left(d'Wd\right)^{-1}$$

- Test whether a coefficient is zero (Cochrane Section 10.1):

$$\frac{\widehat{\mathbf{b}}_i}{\sqrt{\mathbf{V}(\widehat{\mathbf{b}})_{ii}}} \sim N(0, 1)$$

- Standard error of γ by delta method(Cochrane Section 11.3)

$$\sqrt{T}\left(\gamma(\widehat{\mathbf{b}}) - \gamma\left(\mathbf{b}_0\right)\right) \rightarrow N\left(\mathbf{0}, \left(\frac{\partial \gamma\left(\mathbf{b}_0\right)}{\partial \mathbf{b}'}\right) \mathbf{V}(\widehat{\mathbf{b}}) \left(\frac{\partial \gamma\left(\mathbf{b}_0\right)}{\partial \mathbf{b}'}\right)'\right)$$

- $J-$test of over-identifying restrictions (Cochrane Section 10.2)

$$J_{\widehat{S}^{-1}}\left(\widehat{\mathbf{b}}_{\widehat{S}^{-1}}\right) = T g_T(\widehat{\mathbf{b}}_{\widehat{S}^{-1}})' \widehat{S}^{-\mathbf{1}} g_T(\widehat{\mathbf{b}}_{\widehat{S}^{-1}}) \xrightarrow{d} \chi^2_{\#\text{moments } - \#\text{parameters}}$$

GMM vs MLE

Motivating Principle

- MLE principle: Choose probability model (i.e., estimate parameter values) from which data is most likely to have been generated
- GMM principle: Set parameters to match certain attributes (sample moments) of the data

Trade-offs

1. MLE specifies full likelihood — and therefore all moments — of the data → generally much more structured than GMM
2. **Benefit of structure**: Variance efficiency
   - Correct structure → preciser estimation
3. **Cost of structure**: Misspecification.
   - Incorrect structure introduces bias