第2讲 资产定价和机器学习

南京大学金融与保险学系

杨 念

2025年2月24日

- **Machine learning for factor investing**. G. Coqueret, T. Guida, Chapman & Hall, 2020.
- **Machine learning in asset pricing.** Stefan Nagel, Princeton University Press, 2021.
- **Financial machine learning**. B. Kelly, D. Xiu, Foundations and Trends® in Finance, vol 13(3-4), pages 205-363. 2023
- **Machine Learning in Finance: From Theory to Practice**. Dixon, Matthew F., Igor Halperin, and Paul Bilokon. Springer. 2020.
- 因子投资方法与实践. 石川 等著， 2020

收益率(return rate)

➢单期简单收益率

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

➢单期对数收益率(Continuously-compounded return)

$$r_t = \log \frac{P_t}{P_{t-1}} = \log(1 + R_t) \approx R_T$$

➢收益率优点：衡量投资表现(正、负)，便于资产比较(股票、债券、投资策略)，考虑时间因素(年化收益率)，风险调整分析(Sharpe ratio)，投资组合管理

现代投资组合理论(Modern Portfolio Theory, MPT)

- Harry Markowitz 哈里·马科维茨
  - 在1952年他的博士论文中提出现代投资组合理论(MPT)
  - 并于**1990年**因其对金融经济学的贡献获得了**诺贝尔经济学奖**
- MPT简介
  - 通过分散化投资降低投资组合的风险
  - 关注组合整体的风险与收益
  - 资产之间的相关性影响组合风险
  - 提出“**有效前沿**”概念，寻找最优投资组合（最优的风险收益组合）
- 问题：找出一个投资组合w，给定投资组合均值的情况下，最小化投资组合的方差

$$\min_{w} w'\Sigma \mathrm{w} \quad s.t.\ w'R = \mu;\quad w'1 = 1$$

资本资产定价模型(Capital Asset Pricing Model, CAPM)

- **威廉·夏普(William Sharpe)**、约翰·林特纳（John Lintner）和简·莫辛（Jan Mossin）1960年代提出

$$E[R_i] = R_f + \beta_i\,(E[R_M] - R_f)$$

  - R_i 资产 i 收益率
  - R_f 无风险收益率
  - R_M 市场组合(或**市场因子**)的收益率, 市场组合因子
  - $\beta_i = \mathrm{cov}(R_i, R_M)/\mathrm{var}(R_M)$ 资产i对市场风险的暴露程度(因子载荷)
- 资产预期收益由**无风险收益**(R_f)和**市场风险溢价**($E[R_M] - R_f$)组成
- 基于MPT的假设（投资者如何优化组合），CAPM关注市场均衡状态下资产如何定价

因子定价模型 (Multi-Factor Pricing Model)

- 套利定价理论(Arbitrage Pricing Theory, APT)
  - **斯蒂芬·罗斯(Stephen Ross)**1976 年提出的静态统计模型，是对 CAPM 的扩展
  - 资产的预期收益由**多个系统性风险因子**决定，而不是CAPM单一的市场 Beta 因子
  - 用于描述**没有套利机会**（Arbitrage-Free）的资产定价关系
  - $E[R_i] - R_f = \beta_{i1}\lambda_1 + \beta_{i2}\lambda_2 + \cdots + \beta_{iK}\lambda_K, \quad i = 1, \cdots, N$
- 因子定价模型
  - $\boldsymbol{R_i = \alpha_i + \beta_{i1} f_1 + \beta_{i2} f_2 + \cdots + \beta_{iK} f_K + \epsilon_i}$
  - 如果有定价误差，则超额收益$R_i^e = R_i - R_f$

$$E\left[R_i^e\right] = \boldsymbol{\alpha_i} + \beta_{i1}\lambda_1 + \beta_{i2}\lambda_2 + \cdots + \beta_{iK}\lambda_K$$

  - 资产i预期超额收益分解**系统性风险因子预期超额收益(因子溢价)部分**和**定价误差**
  - Fama-French三因子模型

$$E\left[R_i^e\right] = \boldsymbol{\alpha_i} + \beta_i E[R_M^e] + \beta_S E[R_{SMB}^e] + \beta_H E[R_{HML}^e]$$

Fama-French三因子模型

- $E\left[R_i^e\right] = \boldsymbol{\alpha_i} + \beta_i E[R_M^e] + \beta_S E[R_{SMB}^e] + \beta_H E[R_{HML}^e]$
- $E[R_M^e]$ 市场组合的预期超额收益率
  - 市场组合代理变量
- $E[R_{SMB}^e]$ 规模因子的预期超额收益率
  - 规模因子代理变量：市值
  - 市值(size)：一家上市公司所有流通股的市场总价值=股价*流动股数
- $E[R_{HML}^e]$ 价值因子的预期超额收益率
  - 价值因子代理变量：账面市值比
  - 账面市值比(book-to-market ratio)= 账面价值/市场价值
  - 解释: 高账面市值比公司被低估; 低账面市值比公司被高估

Fama-MacBeth回归估计风险溢价

$$R_{i,t} = \alpha_i + \beta_{i1} f_{t,1} + \beta_{i2} f_{t,2} + \cdots + \beta_{iK} f_{t,K} + \epsilon_{t,i} \qquad (*)$$

- 第一步：在时间序列上，对所有股票逐个进行回归
  - 固定$\boldsymbol{i}$，时序回归 $R_{i,t} = \alpha_i + \beta_{i1} f_{t,1} + \beta_{i2} f_{t,2} + \cdots + \beta_{iK} f_{t,K} + \epsilon_{t,i}$，$\mathbf{1} \le \boldsymbol{t} \le \boldsymbol{T}$
  - 得到估计值 $\hat{\beta}_{ik}$， $i = 1, \cdots N$， $k = 1 \cdots, K$
- 第二步：将上述估计值 $\hat{\beta}_{ik}$代入$(*)$, 在每个时间点，做截面上回归
  - 固定$\boldsymbol{t}$，截面回归 $R_{i,t} = \lambda_{t,0} + \hat{\beta}_{i1} \lambda_{t,1} + \hat{\beta}_{i2} \lambda_{t,2} + \cdots + \hat{\beta}_{iK} \lambda_{t,K} + \epsilon_{t,i}$，$\mathbf{1} \le \boldsymbol{i} \le \boldsymbol{N}$
  - 得到估计值： $\hat{\lambda}_{t,k}$， $t = 1, \cdots, T$， $k = \mathbf{0}, 1, \cdots, K$
  - 时间加权平均 **风险溢价** $\hat{\lambda}_k = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_{t,k}$
  - 可以检验该系数是否显著

$$E_T[R_i] = \hat{\lambda}_0 + \hat{\beta}_{i1} \hat{\lambda}_1 + \hat{\beta}_{i2} \hat{\lambda}_2 + \cdots + \hat{\beta}_{iK} \hat{\lambda}_K$$

因子和机器学习

- 目前学术界已经挖出400+因子，大多数因子是数据窥探的产物
- 2011,John Cochrane用“因子动物园”(factor zoo)描述因子研究状况

- 机器学习
  - 预测
  - 特征选择（feature selection）
  - 非线性
  - 降维
  - Instrumented principal components analysis (IPCA) Kelly, B.T. , Pruitt, S. , Su, Y. , 2019. Characteristics are covariances: a unified model of risk and return. J. Financ. Econ. 134, 501–524 .
  - Factor model, machine learning, and asset pricing

收益率预测

[illegible]. The general formulation is the following. At time T, the agent or investor seeks to solve the following program:

$$\max_{\boldsymbol{\theta}_T} \mathbb{E}_T \left[u(r_{p,T+1}) \right] = \max_{\boldsymbol{\theta}_T} \mathbb{E}_T \left[u\left((\bar{\mathbf{w}}_T + \mathbf{x}_T \boldsymbol{\theta}_T)' \mathbf{r}_{T+1} \right) \right],$$

where u is some utility function and $r_{p,T+1} = (\bar{\mathbf{w}}_T + \mathbf{x}_T \boldsymbol{\theta}_T)' \mathbf{r}_{T+1}$ is the return of the portfolio, which is defined as a benchmark $\bar{\mathbf{w}}_T$ plus some deviations from this benchmark that are a linear function of features $\mathbf{x}_T \boldsymbol{\theta}_T$. The above program may be subject to some external constraints (e.g., to limit leverage).

In practice, the vector $\boldsymbol{\theta}_T$ must be estimated using past data (from $T-\tau$ to $T-1$): the agent seeks the solution of

$$\max_{\boldsymbol{\theta}_T} \frac{1}{\tau} \sum_{t=T-\tau}^{T-1} u\left(\sum_{i=1}^{N_T} \left(\bar{w}_{i,t} + \boldsymbol{\theta}'_T \mathbf{x}_{i,t} \right) r_{i,t+1} \right) \tag{3.5}$$

on a sample of size τ where N_T is the number of asset in the universe. The above formulation can be viewed as a learning task in which the parameters are chosen such that the reward (average return) is maximized.

收益率预测

➢正则化预测性回归

- Chinco, Alexander, Adam D Clark-Joseph, and Mao Ye. 2019. “Sparse Signals in the Cross-Section of Returns.” *Journal of Finance* 74 (1): 449–92.

➢非线性预测

- Gu, Shihao, Bryan T Kelly, and Dacheng Xiu. 2020b. “Empirical Asset Pricing via Machine Learning.” *Review of Financial Studies* 33 (5): 2223–73.

➢工具变量(Instrumented principal components analysis, IPCA)

- Kelly, Pruitt, Su. 2019. Characteristics are covariances A unified model of risk and return. *Journal of Financial Economics* 134: 501-524

线性回归

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}.$$

Standard assumptions are the following:

- $\mathbb{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta$: **linear shape for the regression function**;
- $\mathbb{E}[\boldsymbol{\epsilon}|\mathbf{X}] = \mathbf{0}$: errors are **independent of predictors**;
- $\mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}] = \sigma^2\mathbf{I}$: **homoscedasticity** - errors are uncorrelated and have identical variance;
- the ϵ_i are normally distributed.

$$L = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \sum_{i=1}^{I} \epsilon_i^2$$

$$\begin{aligned}\nabla_\beta L &= \frac{\partial}{\partial\beta}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) = \frac{\partial}{\partial\beta}\beta'\mathbf{X}'\mathbf{X}\beta - 2\mathbf{y}'\mathbf{X}\beta \\ &= 2\mathbf{X}'\mathbf{X}\beta - 2\mathbf{X}'\mathbf{y}\end{aligned}$$

$$\beta^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

惩罚线性回归-LASSO

- The **L**east **A**bsolute **S**hrinkage and **S**election **O**perator (LASSO) (Tibshirani, 1996)
- L1-regularization

$$y_i = \sum_{j=1}^{J} \beta_j x_{i,j} + \epsilon_i, \quad i = 1, \ldots, I, \quad \text{s.t.} \quad \sum_{j=1}^{J} |\beta_j| < \delta,$$

- 等价于求解下列Lagrangian优化

$$\min_{\beta} \left\{ \sum_{i=1}^{I} \left(y_i - \sum_{j=1}^{J} \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^{J} |\beta_j| \right\}$$

惩罚线性回归-岭回归 (ridge regression)

- L2-regularization

$$\min_{\beta}\left\{\sum_{i=1}^{I}\left(y_i-\sum_{j=1}^{J}\beta_j x_{i,j}\right)^2+\lambda\sum_{j=1}^{J}\beta_j^2\right\}$$

- 等价于求解下列Lagrangian优化

$$y_i=\sum_{j=1}^{J}\beta_j x_{i,j}+\epsilon_i,\quad i=1,\ldots,I,\quad \text{s.t.}\quad \sum_{j=1}^{J}\beta_j^2<\delta$$

- 当变量数量多于观测值数量，回归系数会变小（收缩）

$$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X}+\lambda\mathbf{I}_N)^{-1}\mathbf{X}'\mathbf{Y}$$

惩罚线性回归-弹性网络(Elastic net)

- 弹性网络 (Zou and Hastie, 2005)

$$y_i = \sum_{j=1}^{J} \beta_j x_{i,j} + \epsilon_i, \quad \text{s.t.} \quad \alpha \sum_{j=1}^{J} |\beta_j| + (1-\alpha) \sum_{j=1}^{J} \beta_j^2 < \delta, \quad i = 1, \ldots, N,$$

- 等价于求解下列Lagrangian优化

$$\min_{\beta} \left\{ \sum_{i=1}^{I} \left(y_i - \sum_{j=1}^{J} \beta_j x_{i,j} \right)^2 + \lambda \left(\alpha \sum_{j=1}^{J} |\beta_j| + (1-\alpha) \sum_{j=1}^{J} \beta_j^2 \right) \right\}$$

- Rapach, David, and Guofu Zhou. 2019. “Time-Series and Cross-Sectional Stock Return Forecasting: New Machine Learning Methods.” *SSRN Working Paper* 3428095.