第三章 金融分析中的几个重要模型

Lianghai Xiao

https://github.com/styluck/mlb

作业邮箱：alswfx@126.com

- CAPM模型
- Fama-French三因子模型
- Barra多因子模型
- 因子对冲模型（alpha模型）
- 投资组合模型

CAPM模型

CAPM模型

- CAPM模型，即资本资产定价模型（Capital Asset Pricing Model），是金融学中的一个基本理论，用于衡量证券的系统性风险和预期收益之间的关系。CAPM模型的核心公式如下：
    - $E(R_i) = R_f + \beta_i\left(E(R_m) - R_f\right)$
- $E(R_i)$: 资产 i 的预期收益率
- R_f : 无风险利率（通常以国债利率为代表）
- β_i : 资产 i 的贝塔系数，衡量该资产相对于市场的系统性风险
- $E(R_m)$: 市场的预期收益率
- $E(R_m) - R_f$: 市场风险溢价，即投资于市场组合而非无风险资产获得的额外收益

模型构成要素

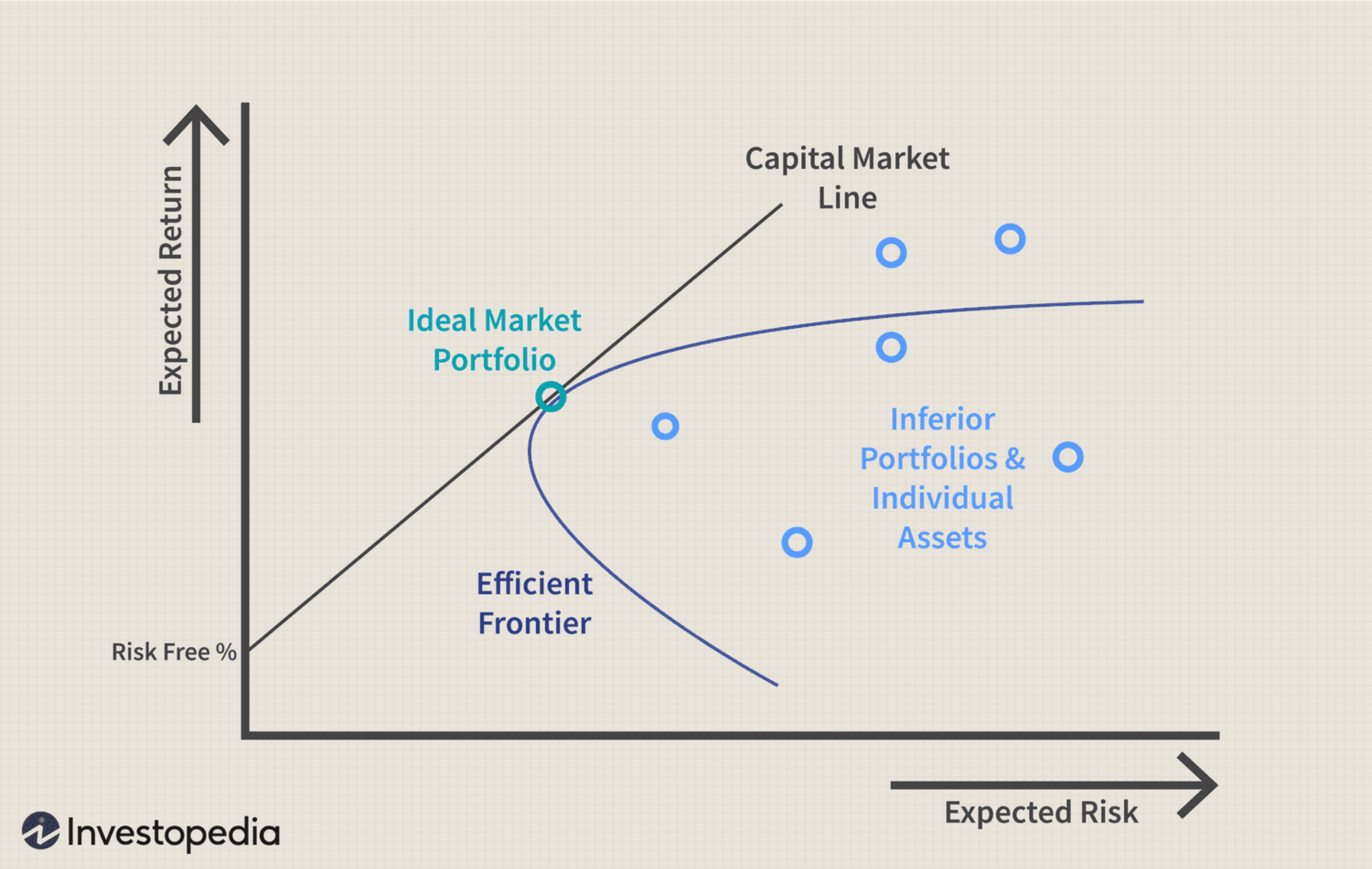
- **无风险资产（$\boldsymbol{R_f}$）**
  - 指在持有期间内，能够确定地获得固定收益，且不存在任何违约风险的资产。
  - 在实际应用中，通常用短期国债的收益率来近似代表无风险利率。
  - 无风险资产在 CAPM 模型中起到了基准的作用。
  - 投资者可以将资金在无风险资产和风险资产之间分配，通过调整这种分配比例来控制投资组合的风险和收益。
  - 例如，投资者如果将全部资金投资于无风险资产，那么其投资组合的收益率就是无风险利率，风险为零。

模型构成要素

- **市场组合（M）**
  - 市场组合是包含了所有可交易资产的投资组合，其中每个资产的权重是该资产的市场价值占总市场价值的比例。
  - 在实际操作中，市场组合通常用一个广泛的市场指数（如美国的标准普尔 500 指数、中国的沪深 300 指数等）来近似表示。
  - 市场组合代表了整个市场的风险和收益状况。
  - 投资者的投资决策往往是在无风险资产和市场组合之间进行权衡，因为根据 CAPM 模型，市场组合是最优风险资产组合，所有理性投资者都应该持有市场组合（在模型假设下）。

模型构成要素

- **贝塔系数（β）**
  - 贝塔系数是衡量单个资产或投资组合系统性风险的指标。它反映了资产或投资组合的收益率对市场组合收益率变动的敏感度。
  - 通过贝塔系数，投资者可以评估资产的风险特性。
  - 如果$\beta = 1$，表示资产的系统性风险与市场组合相同，其收益率随市场组合收益率同比例变动；
  - 如果$\beta > 1$，资产的系统性风险高于市场组合，当市场上涨或下跌时，资产收益率的变动幅度大于市场组合收益率的变动幅度；
  - 如果$\beta < 1$，资产的系统性风险低于市场组合，资产收益率的变动幅度小于市场组合收益率的变动幅度。


Expected Return
Capital Market
Line
Ideal Market
Portfolio
Inferior
Portfolios &
Individual
Assets
Efficient
Frontier
Risk Free %
Expected Risk
Investopedia

Fama-French三因子模型

Fama-French三因子模型

- CAPM 存在局限性，它假设市场只存在一个系统性风险因子，即市场组合的超额收益，但在实际市场中，许多股票的收益表现无法仅用市场风险来解释。
- 尤金·法玛（Eugene F. Fama）和肯尼斯·弗兰奇（Kenneth R. French）通过大量实证研究发现，除了市场风险外，还有其他因素显著影响股票收益，从而提出了三因子模型。

Fama-French三因子模型

- Fama-French三因子模型增加了两个因子，使得对股票回报率的解释更全面。其模型公式如下：
    - $E(R_i) = R_f + \beta_i\big(E(R_m) - R_f\big) + s_i \cdot SMB + h_i \cdot HML$
- $E(R_i)$: 资产 i 的预期收益率
- R_f : 无风险利率
- $E(R_m) - R_f$: 市场风险溢价
- SMB : 小盘股风险因子（Small Minus Big），
- HML : 价值风险因子（High Minus Low），
- β_i : 资产对市场风险的敏感度（与CAPM中的定义一致）
- s_i : 资产对规模因子的敏感度
- h_i : 资产对价值因子的敏感度

原有的要素

- **市场风险溢价因子（Market Risk Premium Factor，$R_M - R_f$）**
  - 市场组合通常用一个广泛的市场指数来代表，如美国市场的标准普尔500 指数，中国市场的沪深 300 指数等。
  - 无风险利率一般可以用国债收益率、长期政策无风险利率等来近似表示。
  - **市场风险溢价因子反映整个市场的系统性风险。**它等于市场组合的预期收益率（R_M）减去无风险利率（R_f）。
  - 如果市场预期经济形势向好，市场组合的预期收益率会上升，这意味着投资者承担市场风险能获得更高的回报。

新的要素

- **规模因子（Small Minus Big，SMB）**
  - SMB 是**小市值公司组合**的平均收益率减去**大市值公司组合**的平均收益率。这里通过将市场中的股票按照市值大小进行分组，分别构建小市值公司组合和大市值公司组合来计算。
  - **规模因子衡量公司规模对股票收益的影响。**一般来说，小市值公司的风险和成长潜力相对较大。它们可能在市场竞争中面临更大的不确定性，但如果发展良好，也可能获得更高的收益。相反，大市值公司通常较为成熟稳定，增长速度相对较慢。
  - 例如，在新兴行业中，许多小公司有创新的业务模式，但面临资金、市场份额等多方面挑战。如果行业发展顺利，这些小公司的股票可能会有较高的涨幅，导致 SMB 值反映的规模效应显著。

新的要素

- **价值因子（High Minus Low，HML）**
  - HML 是**高账面市值比**（B/M，Book - to - Market Ratio）公司组合的平均收益率减去**低账面市值比**公司组合的平均收益率。账面市值比是公司的账面价值与市场价值的比值，反映了公司的估值水平。
  - **价值因子衡量了公司的价值特性对股票收益的影响。**高账面市值比的公司通常被市场认为是价值型公司，它们的市场价格相对账面价值较低，可能存在被低估的情况。这类公司可能处于行业周期的低谷或者面临短期困境，但从长期来看，有回归合理价值的潜力。
  - 例如，一些传统制造业企业，由于市场竞争、产能过剩等原因，股价被压低，账面市值比升高。如果企业通过改革、行业环境改善等因素实现价值回归，其股票收益可能会增加，使 HML 因子在收益解释中发挥作用。

Barra 模型

Barra 模型

- Barra 模型由美国 Barra 公司（后被明晟公司 MSCI 收购）开发。20 世纪 70 年代，随着金融市场的发展和机构投资者资产规模的不断扩大，投资者对资产组合的风险管理需求日益增长。传统的风险管理方法难以全面、准确地衡量资产组合的风险，在这样的背景下，Barra 模型应运而生。
- 目前是对冲基金公司实际**运用最多**的模型之一。

Barra 模型

- Barra模型是一种多因子风险模型，利用了一系列股票特征（因子）来解释资产组合的波动性。其基本公式为：
  - $R_i = \alpha + \sum_{k=1}^{K} f_k \cdot \beta_{ik} + \epsilon_i$
- R_i: 资产i的收益率
- α : 无解释部分的收益（即超额收益）
- f_k : 因子k的收益
- β_{ik} : 资产i对因子k的暴露度（即敏感度）
- ϵ_i : 非系统性风险或特质性风险，代表该资产中无法通过因子解释的部分

价值因子->基本面因子

- **基本面因子**
  - 基于公司的**财务报表和基本经营数据**构建。**基本面因子**的变化直接反映了公司内在价值的改变，从而影响股票价格。
  - **盈利因子**可能包括净资产收益率（ROE）、每股收益（EPS）等指标，这些指标反映了公司的盈利能力。再如，**杠杆因子**可以通过资产负债率等指标衡量，体现公司的债务负担和财务风险。
  - 一家公司的 ROE 持续上升，意味着其盈利能力增强，在市场有效的假设下，该公司股票价格往往会上涨，因此对投资组合的风险和收益产生影响。

规模因子->风格因子

- **风格因子**
  - 风格因子用于描述股票的风格特征。不同风格的股票在市场上往往有不同的表现。
  - 常见的风格因子包括**市值因子**（如小市值股和大市值股）、**价值成长因子**（通过账面市值比等指标衡量，高账面市值比的股票可能被视为价值股，低账面市值比的股票可能被视为成长股）、**动量因子**（衡量股票价格的短期走势）等。
  - 在市场上涨阶段，成长风格的股票可能表现更好，而在市场波动较大或经济下行阶段，价值风格的股票可能相对更稳定。投资者对不同风格股票的偏好变化会影响股票价格，进而影响投资组合的风险和收益。

新的因子

- **宏观经济因子**
  - 宏观经济因子考虑了宏观经济环境对资产价格的影响。宏观经济的变化会对企业的经营环境和盈利能力产生广泛影响。
  - 国内生产总值（GDP）增长率、通货膨胀率、利率水平等宏观经济变量被纳入模型。这些因子反映了宏观经济的整体状况。
  - 当 GDP 增长率上升时，企业的市场需求可能增加，盈利有望改善，从而推动股票价格上涨。相反，高通货膨胀率可能增加企业成本，降低利润，对股票价格产生负面影响。
- **国家或地区因子**：
  - 主要用于全球投资组合，捕捉地理区域或国家特定的经济、政治风险。例如，美股和欧洲股市可能受到不同经济条件的影响。

因子对冲模型

什么是做空

- **做空**（Short Selling）是金融市场中的一种交易策略和操作方式。是指投资者预期某种资产（如股票、债券、商品、外汇等）的价格将会下跌，通过**借入该资产**并在市场上卖出，待价格下跌后再以**低价买回该资产**归还出借方，从而**获取差价利润**的交易行为。
- **做多**：投资者以每股 10 元的价格买入股票，当股价上涨到每股 15 元时卖出，每股获利 5 元。
- **做空**：投资者借入股票以每股 100 元的价格卖出，当股价下跌到每股 80 元时，投资者以每股 80 元的价格买回股票归还出借方，每股获利 20 元。

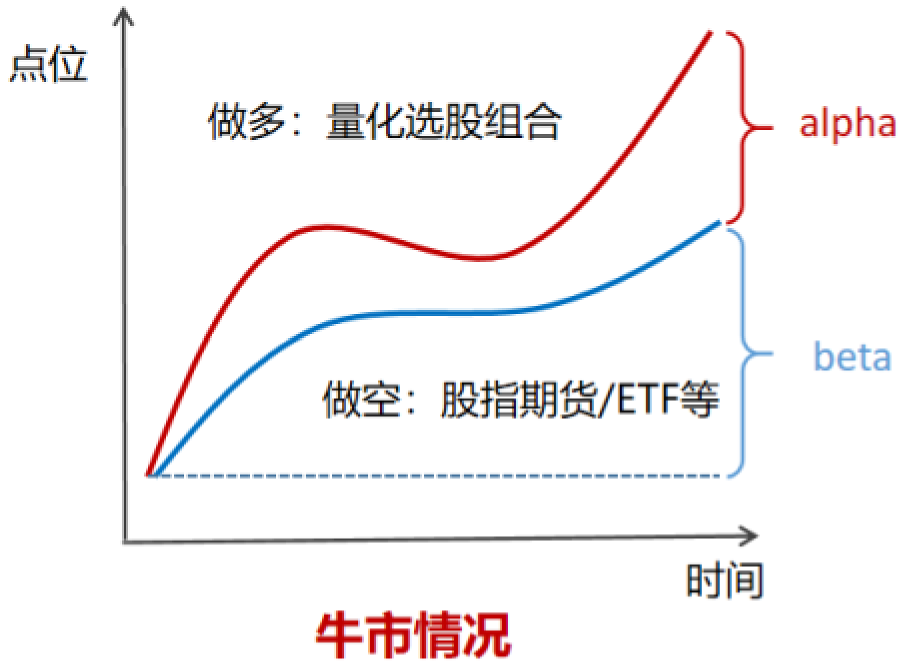
因子对冲模型

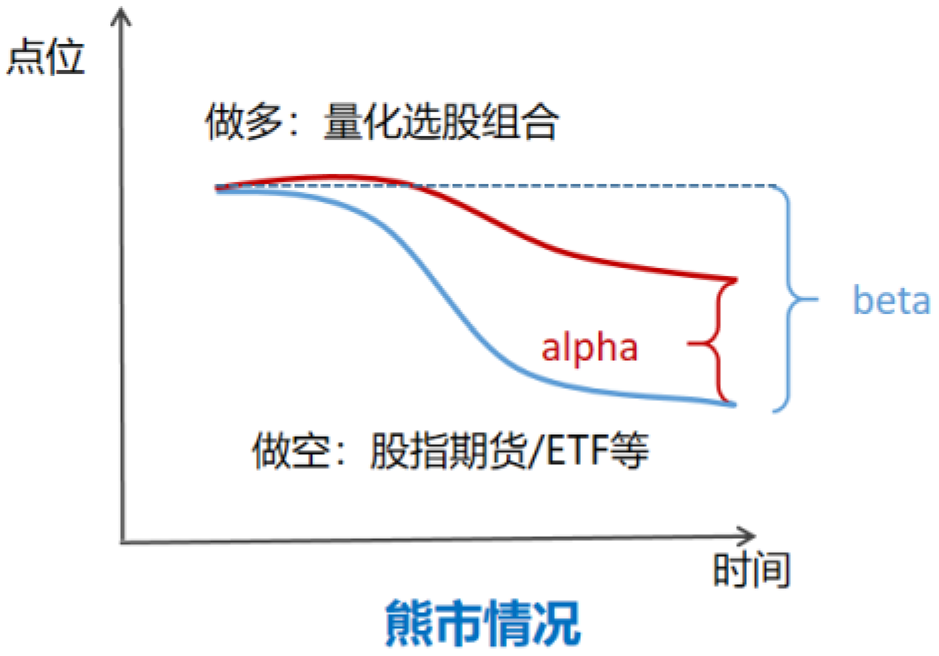
- **从原理上说，因子对冲模型：**
- 一方面选出具备较高超额收益的股票**进行最多**投资，
- 一方面通过股指期货等工具对市场**进行做空**，帮助资产组合抵消市场整体波动带给个股的风险（即Beta）
- 从而力争获得绝对超额收益（即Alpha）。

买入股票 (Alpha+Beta) + 卖出股指期货 (-Beta) = 获得超额收益 (Alpha)

对冲的原理

- 对冲的基本原理就是通过建立反向的市场头寸，抵消市场波动对投资组合收益的影响，使投资组合的收益更加稳定。





投资组合模型

等权模型和加权模型

- **均匀等权：** 在构建投资组合时，对组合中的每一种资产分配相同的权重。一个投资组合包含n种资产，那么每种资产的权重:
    - $w_i = \frac{1}{n}$.
- **市值加权：** 根据资产的市值来分配权重。第i种资产的权重：

$$w_i = \frac{\text{市值}_i}{\sum_{j=1}^{n} \text{市值}_j}$$

- 例：标普 500 指数、沪深300指数
- **价格加权：** 基于成分股的价格来分配权重。例：道琼斯工业平均指数。

均值方差模型

- 均值方差模型是**现代投资组合理论**的**核心内容**，由哈里・马科维茨（Harry Markowitz）提出。基本思想是通过权衡投资组合的预期收益和风险来构建最优投资组合。该模型可以描述为，在给定预期收益率μ_p的情况下，最小化投资组合方差:

$$\min_{w_1,w_2,\cdots,w_n} \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} w_i = 1 \text{且} \sum_{i=1}^{n} w_i \mu_i = \mu_p$$

均值方差模型

- 构建拉格朗日函数

$$L(\mathbf{w},\lambda_1,\lambda_2)=\mathbf{w}^T\mathbf{\Sigma}\mathbf{w}+\lambda_1(\sum_{i=1}^{n}w_i-1)+\lambda_2(\sum_{i=1}^{n}w_i\mu_i-\mu_p)$$

- 求偏导数

$$\frac{\partial L}{\partial \mathbf{w}}=2\mathbf{\Sigma}\mathbf{w}+\lambda_1\mathbf{1}+\lambda_2\boldsymbol{\mu}=0$$

$$\frac{\partial L}{\partial \lambda_1}=\sum_{i=1}^{n}w_i-1=0$$

$$\frac{\partial L}{\partial \lambda_2}=\sum_{i=1}^{n}w_i\mu_i-\mu_p=0$$

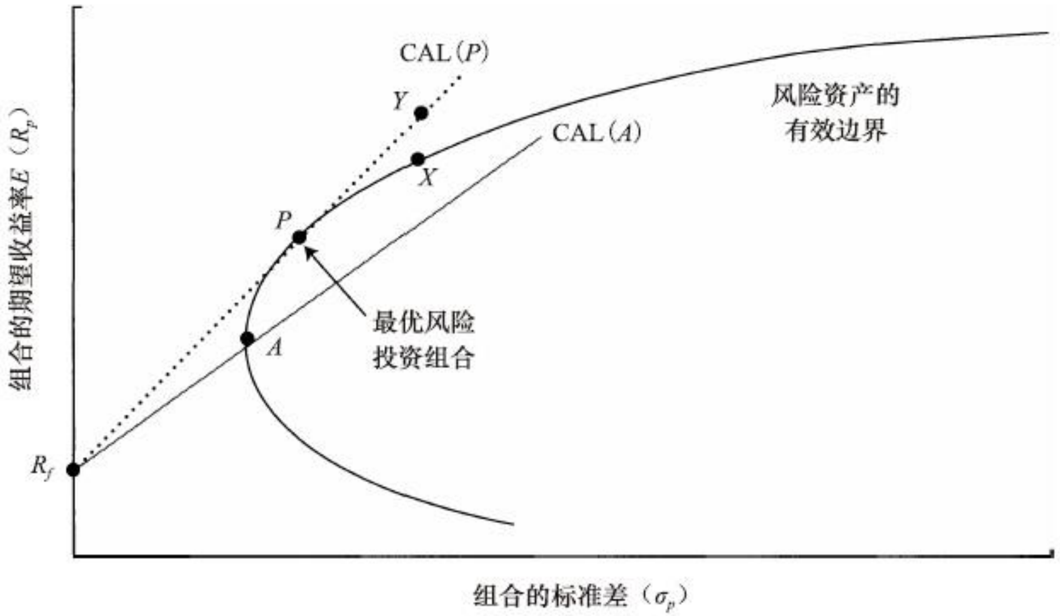
- 得到w的闭式解： $\mathbf{w}^*=\frac{1}{A}[C(\mathbf{\Sigma}^{-1}\mathbf{1})-B(\mathbf{\Sigma}^{-1}\boldsymbol{\mu})]$

$$A=BC-D^2,\ B=\mathbf{1}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}\quad C=\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1},\ D=\boldsymbol{\mu}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$$

均值方差模型

- 含有**无风险资产时**的闭式解:
- 投资者将资金分配在无风险资产和风险资产组合上，最优风险资产组合的权重向量的闭式解为:

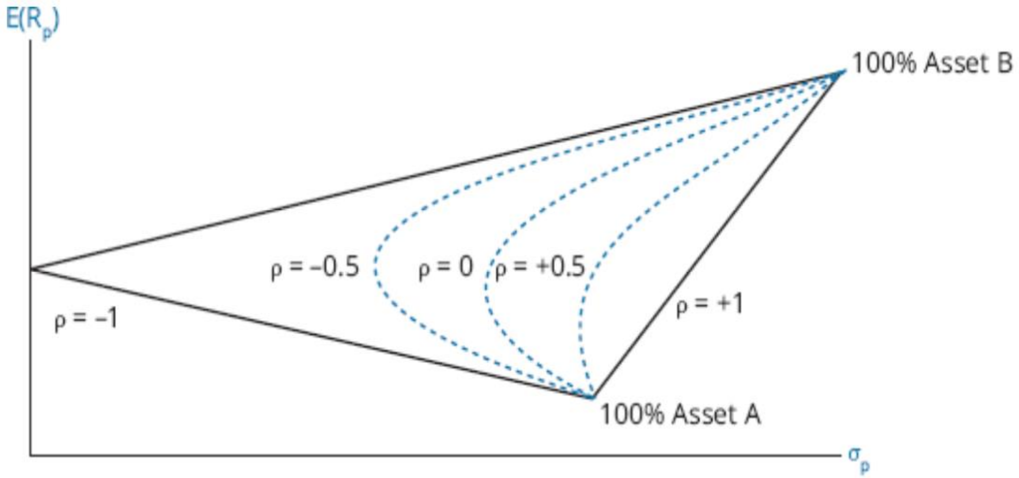
$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1})}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1})}。$$



均值方差模型

- 含有无风险资产时的闭式解:
- 投资者将资金分配在无风险资产和风险资产组合上，最优风险资产组合的权重向量的闭式解为:

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1})}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1})}。$$



风险平价模型

- 风险平价（Risk parity）投资组合核心目标是让投资组合中不同资产类别对组合整体风险的贡献相等。
- 设投资组合的权重向量为w，资产收益率的协方差矩阵为Σ，则资产i对组合方差（风险的一种度量）的贡献为:

$$RC_i = w_i \times \frac{\sum_{j=1}^{n} w_j \sigma_{ij}}{\sqrt{\sum_{k=1}^{n} \sum_{l=1}^{n} w_k w_l \sigma_{kl}}}$$

- 在风险平价策略中，目标是使所有资产的风险贡献相等，通过求解这个方程组来确定投资组合中各个资产的权重。即

$$RC_1 = RC_2 = \cdots = RC_n。$$

风险平价模型

- 基于这样的假设：资产的风险可以用标准差来衡量，并且标准差倒数的比例能够平衡各资产的风险贡献。那么我们可以构建一个简化版的风险平价模型，利用标准差的倒数来计算权重：
- 首先计算资产i标准差的倒数

$$w_i' = \frac{1}{\sigma_i}。$$

- 然后对标准差倒数进行归一化处理得到各资产在投资组合中的权重：

$$w_i = \frac{w_i'}{\sum_{j=1}^{n} w_j'}。$$

实验：Fama-French因子模型

- **目标**
  - 构造市场、规模（SMB）、价值（HML≈由BP构造）三因子；
  - 估计个股的三因子 β 并预测未来收益与价格路径。

- 步骤1：
  - 将所有的数据分成2份：2023-12-31前的数据作为训练集，2024-01-01后的数据作为测试集
  - 处理停牌/缺失值，前复权（若使用日频）
  - 将变量换算到所用收益频率
  - 剔除新上市股票的影响：剔除上市一年内的股票数据
  - 生成对数收益：（ $r_{i,t} = \ln\left(P_{i,t}/P_{i,t-1}\right)$ ），市场超额收益：（ $r_{m,t} - r_{f,t}$ ）

实验：Fama-French因子模型

- 步骤2：
- 因子数据的预处理：
  - 分布分析：选取最近一期的的bp因子和mv因子值，分析数据的分布类型
  - 去空值：将数据缺失的地方设为上一期的数据值
  - 改变分布：将每期bp因子和mv因子，采用box-cox方法改变分布，使其接近正态分布（scipy包可以自动估计lambda值）
  - 去极值：将每期bp因子的数据采用winsor方法去极值，区间为 $[\mu-3\sigma,\mu+3\sigma]$。
  - 标准化：将去极值处理后的因子序列进行z-score标准化处理。即，减去其现在的均值、除以其标准差，得到一个新的近似服从$N(0,1)$分布的序列

实验：Fama-French因子模型

- 步骤3：
- 市场层面的 SMB_t、HML_t计算：
  - 用每期的市值中位数把股票分成 Small (S) / Big (B)。
  - 用 BP（Book-to-Price）或 BM（Book-to-Market）分位点 30%、70% 分成 Low (L) / Neutral (N) / High (H)。
  - 得到 6 个**等权**（或市值加权）组合：S/L、S/N、S/H、B/L、B/N、B/H。
  - 计算各组合当月收益，记为 $(S/L)_t \dots (B/H)_t$，然后合成因子收益：

$$\text{SMB}_t = \frac{1}{3}\big[(S/L)_t + (S/N)_t + (S/H)_t\big] - \frac{1}{3}\big[(B/L)_t + (B/N)_t + (B/H)_t\big],$$

$$\text{HML}_t = \frac{1}{2}\big[(S/H)_t + (B/H)_t\big] - \frac{1}{2}\big[(S/L)_t + (B/L)_t\big].$$

实验：Fama-French因子模型

- 步骤4：
  - 估计个股三因子 β：
  - 对目标股票 (i) 做20天滚动时序回归：
  - $r_{i,t} - r_{f,t} = \alpha_i + \beta_{i,M}, \mathrm{MKT}_t + \beta_{i,S}, \mathrm{SMB}_t + \beta_{i,H}, \mathrm{HML}_t + \varepsilon_{i,t}$.
  - 记录 $\widehat{\alpha}_i$, $\widehat{\beta_{i,M}}$，$\widehat{\beta_{i,S}}$，$\widehat{\beta_{i,H}}$ 与残差方差 $\widehat{\sigma^2_{\varepsilon,i}}$。

实验：Fama-French因子模型

- 步骤5：基于训练集的数据
- 因子评价：
  - 以预处理后的t期$\widehat{\beta_{i,M}}$，$\widehat{\beta_{i,S}}$，$\widehat{\beta_{i,H}}$ 因子暴露作为解释变量，t+1期的超额收益率数据作为被解释变量，采用最小二乘法进行**全样本**线性回归（OLS）。
  - 判断采用最小二乘法进行线性回归的各项假设是否成立。(取最近一期的数据进行判断即可)
- 计算以下评价指标：
  - t 值序列的绝对值的平均值
  - R方值的序列
  - t 值序列绝对值大于 2 的占比
  - t 值序列均值的绝对值除以 t 值序列的标准差
- 计算以下评价指标并绘制图表进行展示：
  - bp因子和mv因子的累计回归系数
  - t 值的绝对值序列
  - R方值的序列

实验：Fama-French因子模型

- 步骤6：
- 基于因子权重配置的投资组合构建：
  - 利用三因子模型和市场数据，计算资产的预期收益率，从中选出预期收益率最高的30个资产。
  - 从以下3种权重构成方案中，选择一种，进行投资组合的构建：
    - **等权配置（推荐）**
    - **市值等权配置**
    - **风险平价模型**
    - **均值方差模型**
  - 使用上述模型，基于测试集数据，构建投资组合的模拟回测图。