常用公式

随机变量性质

- 如果$E(X)=0$,则$Var(X)=E(X^2)$
- 偏度$S=\frac{\mu_3}{\sigma^3}=\frac{E(X-E(X))^3}{\mathrm{Var}(X)^{3/2}}$
- 峰度$K=\frac{\mu_4}{\sigma^4}=\frac{E(X^4)}{(\mathrm{Var}X)^2}$

Beta

- portfolio可以直接把asset的beta按照weight相加
- 收益率相除可以计算 $\mu_R-\mu_f=\beta_R(\mu_M-\mu_f)$
- 风险由两部分组成$\sigma_i^2=\beta_i^2\sigma_M^2+\sigma_\epsilon^2$

AR(1)

- 可以从ARCH模型获得,关注α

$$\gamma(h)=\frac{\sigma^2\alpha^{|h|}}{1-\alpha^2}$$
$$\rho(h)=\alpha^{|h|}$$

MA(q)

- 从0开始

$$\gamma(h)=\sigma^2\sum_{i=0}^{q-h}\psi_i\psi_{i+h}$$

ARMA(1,1)

- 可以从GARCH(1,1)模型获得,关注α_1,β_1

$$X_n^2=\sigma_n^2\epsilon_n^2=\sigma_n^2+V_n=\alpha_0+\alpha_1\sigma_{n-1}^2+\beta_1X_{n-1}^2+V_n=\alpha_0+(\alpha_1+\beta_1)\sigma_n^2-\beta_1V_{n-1}+V_n$$
$$\sigma_{X_n^2}=\frac{1+\beta_1^2-2(\alpha_1+\beta_1)\beta_1}{1-(\alpha_1+\beta_1)^2}\sigma_\epsilon^2$$
$$\rho(h)=(\alpha_1+\beta_1)^{|h|-1}\frac{\alpha_1[1-(\alpha_1+\beta_1)\beta_1]}{1+\beta_1^2-2(\alpha_1+\beta_1)\beta_1}$$

Linear Predictor

- 通常先计算自方差,然后带入Cov

$$\hat{\alpha}=\Sigma_{YX}\Sigma_X^{-1}=YX^T(XX^T)^{-1}=\frac{\mathrm{Cov}(Y,X)}{\mathrm{Var(X)}}$$
$$MSE=E[(Y-\hat{Y})^2]=\sigma_Y^2-\frac{\mathrm{Cov}(Y,X)^2}{\mathrm{Var}(X)}$$

1. Time Series

1.1 基本概念

不同时间序列定义

- First TS model: t_n对应$n \triangle t$的时间点
- Second-order: 所有t_n时刻的随机变量X_n的均值和方差都是有限的

平稳性

- 强平稳性: 任意h个连续是时间序列变量分布相同
- 弱平稳性: 均值相同,h阶自方差相同

特征估计 (只对平稳时间序列有效)

- 均值: $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$
- h阶自方差: $\hat{\gamma}(h) = \frac{1}{N}\sum_{i=1}^{h-1}(X_i - \hat{\mu})(X_{i+h} - \hat{\mu})$
- h阶自相关系数: $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\hat{\gamma}(h)}{\hat{\sigma}}$

偏自相关系数

- 使用X_m作为应变量,$X_{m-1}, \cdots, X_{m-h}$作为自变量,进行线性回归
- 线性回归得到的参数,就是PACF (可以有效去除其他变量的作用)

Ergodicity

- 当平稳的时间序列足够长,均值和自方差的估计值会趋向于真实值
- $\lim_{n\to\infty} \hat{\mu} = \mu, \lim_{n\to\infty} \hat{\rho}(h) = \rho(h)$

1.2 时间序列检验

平稳性检验Phillips-Perron Test

- 使用ARMA拟合数据来进行检验
- H_0:单位圆内存在单位根,数据不平稳
- H_1 :数据平稳,因此P-value越小越好

Jarque-Bera检验正态分布

- 通过样本的偏度S和峰度K来构造参数检验

$$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4}) \sim \chi_2^2$$

- H_0:数据服从正态分布,JB值越偏离χ_2^2,说明样本正态分布的概率越小

Shapiro-Wilk Test

- 检验样本是否服从假设的正态分布,分布参数决定α_i
- $W = \frac{(\sum_{i=1}^{n}\alpha_i x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

2. ARIMA模型

2.1 AR(1)

定义

- 给定均值μ和模型参数α,时间序列在去除均值之后符合递推式

$$Y_n - \mu = \alpha(X_{n-1} - \mu) + \epsilon_n$$

- ϵ_n服从均值$N(0,\sigma^2)$的分布
- $|\alpha| < 1$,否则不平稳

时序性质

- $\gamma(h) = \frac{\sigma^2\alpha^{|h|}}{1-\alpha^2}$
- $\rho(h) = \alpha^{|h|}$

参数估计

- 当变量之间服从多元正态分布时,线性回归就是最小化MSE的分布
- 去除均值后给定自变量矩阵$X \in \mathbb{R}^{d\times m}$和应变量矩阵$Y \in \mathbb{R}^{1\times m}$时,求解$\alpha^T X = Y$
- $\alpha^T = \Sigma_{YX}\Sigma_X^{-1} = YX^T(XX^T)^{-1}$
- 使用$\alpha^T X$带入$\hat{Y}$来分析$MSE = E[(Y-\hat{Y})^2] = \sigma_Y^2 - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}$
- 变量相关性越大,线性回归拟合越准

2.2 AR(p)

定义

$$Y_n = \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + \cdots \phi_p Y_{n-p} + \epsilon_n$$

- 系数参数绝对值小于1,则模型平稳

参数拟合

- 假定Y_t和$Y_{t-1},\cdots,Y_{t-p}$服从多元正态分布,因此可以直接使用线性回归来最小化MSE并估计参数
- 构造自方差矩阵

$$\Sigma_Y = \begin{bmatrix} \gamma(0) & \cdots & \gamma(n-1) \\ \cdots & \cdots & \cdots \\ \gamma(n-1) & \cdots & \gamma(0) \end{bmatrix}$$

- 构造协方差向量

$$\Sigma_{Y_n,Y_{n-p:n-1}} = [\gamma(p),\cdots,\gamma(1)]$$

- 可以生成预测模型$\hat{Y}_t = \Sigma_{Y_n,Y_{n-p:n-1}}\Sigma_Y^{-1}(Y_{n-p:n-1} - \mu) + \mu$

矩量法估计参数(Yule-Walker)

- 基于自方差定义,可以得到方程组

$$\gamma(0)=\sum_{i=1}^{p}\phi_i\gamma(|i|)+\sigma^2$$
$$\gamma(1)=\sum_{i=1}^{p}\phi_i\gamma(|i-1|)$$
$$\cdots$$
$$\gamma(p)=\sum_{i=1}^{p}\phi_i\gamma(|i-p|)$$

- 带入自方差估计值$\hat{\gamma}$求得σ和ϕ_i

线性拟合求参数

- 假定Y_t和$Y_{t-1},\cdots,Y_{t-p}$服从多元正态分布,因此可以直接使用线性回归来最小化MSE并估计参数
- 构造自方差矩阵

$$\Sigma_Y=\begin{bmatrix}\gamma(0) & \cdots & \gamma(n-1)\\ \cdots & \cdots & \cdots\\ \gamma(n-1) & \cdots & \gamma(0)\end{bmatrix}$$

- 构造协方差向量

$$\Sigma_{Y_n,Y_{n-p:n-1}}=[\gamma(p),\cdots,\gamma(1)]$$

- 可以生成预测模型$\hat{Y}_t=\Sigma_{Y_n,Y_{n-p:n-1}}\Sigma_Y^{-1}(Y_{n-p:n-1}-\mu)+\mu$

Conditional MLE估计参数

- 对于$X_p,X_{p-1},\cdots X_0$, 有$X_p^2-\alpha_0-\alpha_1X_{p-1}^2-\cdots\alpha_pX_0^2=\epsilon_n\sim N(0,\sigma^2)$,因为残差服从独立正态分布,因此可以计算联合似然函数

$$f(\alpha,\sigma^2|X)=\prod_{i=p+1}^{n}\frac{1}{\sqrt{2\pi\sigma}}\exp[-\frac{1}{2}(\frac{X_i-\sum_{j=1}^{p}\alpha_jX_{i-j}}{\sigma})^2]$$

模型选择

- 画残差的QQplot,是否符合正态分布
- 画残差的ACF,是否在bound之内
- 对ACF进行Box检验,确认是否stationary
- 在确保残差正态分布的情况下,选择AIC最小的模型

2.3 MA

定义

$$Y_n=\epsilon_n+\psi_1\epsilon_{n-1}+\cdots+\psi_q\epsilon_{n-q}$$

- 使用MA项可以有效拟合时序数据中的自方差

时序性质

- $\sigma_Y^2=\sigma^2\sum_{i=0}^{q}\psi_i^2$
- $\gamma(h)=\sigma^2\sum_{i=0}^{q-h}\psi_i\psi_{i+h}$

2.4 ARMA模型

模型定义

- AR和MA模型的组合

$$Y_n = \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + \cdots \phi_p Y_{n-p} + \epsilon_n + \psi_1 \epsilon_{n-1} + \cdots + \psi_q \epsilon_{n-q}$$

- 使用算符简化,如果两边可以消元,这说明可以化简为更简单的ARMA模型: $\phi(B)Y_n = \psi(B)\epsilon_n$
- 拟合效果往往比单独使用AR或者MA更准确,但是有时候需要去除平均数

ARMA模型性质

- 方差

$$\sigma_Y^2 = \frac{1 + \sum_{i=1}^q \psi_i^2 + 2\sum_{i=1}^{\min(p,q)} \psi_i \phi_i}{1 - \sum_{i=1}^p \phi_i^2}$$

3. Garch

3.1 ARCH模型

定义

- X_n是误差大小σ_n和白噪音ϵ_n的乘积 $X_n = \sigma_n \epsilon_n$
- $\epsilon_n \sim N(0,1)$,因此$E(\epsilon^2) = 1$
- σ_n是时间序列数据,决定了误差的绝对值 $\sigma_n^2 = \alpha_0 + \sum_{j=1}^p \alpha_j X_{n-j}^2$ $\alpha_0 > 0, \alpha_j \geq 0$
- ϵ_n是均值为0,方差为1的白噪声
- α_0决定了X_n的下界

性质

- 自激发:当前X_n较大时,之后的X_{n+p}都容易较大
- 由于$\mathrm{Var}(\epsilon_n) = 1$, $E(X_n^2) = E(\sigma_n^2)E(\epsilon_n^2) = [\alpha_0 + \sum_{j=1}^p \alpha_j E(X_{n-j}^2)]$,因此期望符合$p$阶的AR模型

参数估计

- MLE估计太困难,因此使用condition MLE来估计α_j的值
- 对于$X_n = \sqrt{\alpha_0 + \sum_{i=1}^p \alpha_i X_{n-i}^2}\epsilon_n$,服从均值为0,方差为$\alpha_0 + \sum_{i=1}^p \alpha_i X_{n-i}^2$的正态分布,因此可以用来计算似然函数,并对参数进行估计

$$L(\alpha_0, \alpha_1 | x) = -\frac{1}{2}\sum_{j=p+1}^n \log[2\pi(\alpha_0 + \sum_{i=1}^p \alpha_i X_{j-i}^2)] - \frac{1}{2}\sum_{j=p+1}^n \frac{X_j^2}{\alpha_0 + \sum_{i=1}^p \alpha_i X_{j-i}^2}$$

- 通过最大化所有项的f可以得到对α的估计值

3.2 GARCH模型

定义

- 给X_n添加了均值,$X_n = \mu_n + \sigma_n \epsilon_n$
- 误差绝对值由之前的两项组成

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \beta_i \sigma_{n-i}^2 + \sum_{j=1}^{q} \alpha_j (X_{n-j} - \mu_{n-j})^2$$

Garch(1,1)

- 实际应用最常见,且通常先去除均值μ
- Garch(1,1)中X_n^2服从ARMA(1,1)模型,且残差是$V_n = \sigma_n^2(\epsilon_n^2 - 1)$的白噪声,$\alpha_1 + \beta_1$是会决定波动率性质

$$X_n^2 = \sigma_n^2 \epsilon_n^2 = \sigma_n^2 + V_n = \alpha_0 + \alpha_1 \sigma_{n-1}^2 + \beta_1 X_{n-1}^2 + V_n = \alpha_0 + (\alpha_1 + \beta_1)\sigma_n^2 - \beta_1 V_{n-1} + V_n$$

$$\sigma_{X_n^2} = \frac{1 + \beta_1^2 - 2(\alpha_1 + \beta_1)\beta_1}{1 - (\alpha_1 + \beta_1)^2} \sigma_\epsilon^2$$

$$\rho(h) = (\alpha_1 + \beta_1)^{|h|-1} \frac{\alpha_1[1 - (\alpha_1 + \beta_1)\beta_1]}{1 + \beta_1^2 - 2(\alpha_1 + \beta_1)\beta_1}$$

- 当时间序列趋向于无穷时,方差会收敛于特定值

$$\sigma_\infty^2 \to \frac{\alpha_0}{1 - \alpha_1 - \beta_1} (\alpha_1 + \beta_1 < 1)$$

- 给定n时刻的σ,则可以估计h时刻后的σ_{n+h}期望 $(\alpha_1 + \beta_1)$决定了偏向σ_n^2还是σ_∞^2

$$E(\sigma_{n+h}^2 | \sigma_{1:n}) = \sigma_\infty^2 + (\alpha_1 + \beta_1)[E(\sigma_{n+h-1}^2) - \sigma_\infty^2] = (\alpha_1 + \beta_1)^{h-1} \sigma_n^2 + [1 - (\alpha_1 + \beta_1)^{h-1}] \sigma_\infty^2$$

Half-Life of Volatility

- 波动率之差恢复到之前一半的耗时k

$$|E(\sigma_{n+k}^2 | \sigma_{1:n}) - \sigma_\infty^2| \le \frac{1}{2} |\sigma_{n+1}^2 - \sigma_\infty^2|$$

- 对于Gacrh(1,1),可以通过带入预测式求解$(\alpha_1 + \beta_1)^{k-1} \le \frac{1}{2}$
- $\alpha_1 + \beta_1$越大,说明作用持续时间越久,衰减期越长

3.3 GARCH模型变种

结合ARMA模型生成的综合模型

- 观测值服从ARMA模型$X_n = \sum_{i=1}^{p} \psi_i X_{n-i} + \sum_{j=1}^{q} \phi_j \epsilon_{n-j} + \epsilon_n$ (通常只有AR部分)
- 残差波动服从Garch模型$\epsilon_n = \sigma_n \delta_n, \sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \sigma_{n-i}^2 + \sum_{j=1}^{q} \beta_j \epsilon_{n-j}^2$
- 分步估计法: 先用ARMA模型拟合,再用Garch模型拟合残差,结果可能不准确,但泛化能力更强
- 联合估计法: 直接估计所有参数,理论上更准确,但结果鲁棒性不强

APARCH Model

- 模型定义

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta$$

- δ: power parameter,决定了波动率的幂,对于GARCH来说是2
- γ: leverage parameter,决定了模型的不对称性,对于GARCH来说是0

4. Stochastic Volatility

4.1 模型定义

特点

- 波动率是随机的,和表现出来的收益率无关,即σ不会再由X决定
- 通常使用马隐模型中的隐状态和显状态来描述模型
- 使用MCMC采样来获得模型的似然函数

优点

- 可以描述更多更复杂的波动率表现
- 估计准确度高,似然函数大

缺点

- 参数更多,模型更复杂,稳定性可能不好
- 计算困难,没有直接计算似然函数的方式

4.2 Teh Volatility Model

定义

- X_n符合AR过程

$$X_n = \alpha_0 + \sum_{i=1}^{p} \alpha_1 X_{n-i} + \epsilon_n$$

- 残差(白噪音)的波动率服从SV模型

$$\epsilon = \sqrt{h_n}\delta_n \qquad \delta_n \sim N(0,1)$$
$$\log\left(h_n\right) = \beta_0 + \sum_{i=1}^{p} \phi_i \log\left(h_{n-i}\right) + \sum_{j=1}^{q} \psi_j \nu_{n-j} + \nu_n$$

性质

- h_n的对数服从ARMA(p,q)模型
- h_n不再由X_{n-1}或者ϵ_{n-1}决定,而是使用新的独立随机项

5. Arbitrage

5.1 Basic Idea

Assumption

- 找到价格平稳的资产组合
- 当价格过高或者过低时,根据均值回归,存在套利机会
- 有限时间内,总可以找到做空或者做多的机会

风险

- 基本面风险: 当公司发生基本面变化时,价格不再平稳
- 流动性风险: 当发现套利机会时,却无法成交,或者成交成本很高
- 噪音风险: 可能因为散户等因素影响,导致套利逻辑失效
- 执行风险: 套利可能无法完美执行,受到交易时间差,滑点等因素影响

5.2 资产间套利(Co-Integration)

双资产套利

- 对于时间序列Y_1和Y_2,可以找到λ使得$Y_1 - \lambda Y_2$是平稳的
- 使用线性回归找到λ,再通过Phillips-Ouliaris检验平稳性

多资产套利

- 给定d个资产$Y_t = [Y_{1,t} \quad Y_{2,t} \quad \cdots \quad , Y_{d,t}]$
- 找到分配矩阵$A \in \mathbb{R}^{r \times d}$使得$X_t = AY_t^T$是平稳的
- 即矩阵自方差(Cross Auto Covariance) $\text{Cov}(X_t, X_{t+h})$只由h决定

6. Portfolio Management

6.1 Portfolio基本性质

双资产管理

给定双资产组成的Portfolio $V = n_1 P_1 + n_2 P_2$

- 权重向量$w = \begin{bmatrix} \frac{n_1 P_1}{n_1 P_1' + n_2 P_2'} & \frac{n_2 P_2}{n_1 P_1' + n_2 P_2'} \end{bmatrix}$
- 收益率$R = \frac{V'}{V} = \frac{n_1 P_1' + n_2 P_2'}{n_1 P_1 + n_2 P_2}$,通过$R$来计算$VaR$和Expected Shortfall
- 对于双资产Portfolio $V = wR_1 + (1-w)R_2$,则通过分散投资不相关资产来最小化风险

$$w = \frac{\sigma_2^2 - \rho_{12}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2}$$

Copula模型应用

1. 对每个asset估计marginal distribution $\hat{F}_P$
2. 使用QQ plot判断$\hat{F}_p$的估计准确度,注意tail
3. 计算asset之间的spearman相关系数
4. 用$\hat{F}_p$生成概率并带入copula模型
5. 绘制copula的等高线,分析估计质量
6. 从copula生成模拟数据,并获得不同权重时的VaR

6.2 风险资产模型

定义

资产收益率向量	资产收益率期望	收益率协方差矩阵
$R = [R_1 \cdots R_N]^T$	$E[R] = [\mu_1 \cdots \mu_N]^T$	$\text{Cov}(R) = \Sigma$

已知上述条件,找到最合适的权重向量$W = [w_1 \cdots w_N]^T$在收益率最大的情况下最小化风险

- 最小化标准差: $\min w^T \Sigma w$
- 最大化收益率: $w^T \mu = \mu^*$
- 权重和为1: $w^T I = 1$
- 可能还有分散投资,不能做空等附加条件
- 生成的结果为有效边界: 在固定资产的方差σ的情况下,可以获得的最大收益μ

风险资产收益率/方差估计

- 使用经验数据对资产间的收益率和协方差进行估计,但是会引入bias
- 使用Bootstrap对样本收益率,方差和最优夏普率进行估计,置信区间是$[2\hat{\theta}-q_U, 2\hat{\theta}-q_L]$

6.3 模型最优化

无风险资产混合投资

- 在资产组合中引入了收益率为μ_f的无风险资产
- 需要最大化夏普比率:描述了额外收益$E(R)-\mu_f$和风险σ的比值,最大化夏普比率等于寻找切线

$$\frac{E(R)-\mu_f}{\sigma}$$

- 令风险资产权重为w则有
  - 收益率$\mu=\mu_f+w(\mu_p-\mu_f)$
  - 方差$\sigma=w\sigma_p$

最大化Portfolio夏普比率流程

1. 根据无风险资产收益率μ_f,最大化风险资产的夏普比率

$$w_T=\frac{v_1\sigma_2^2-v_2\rho_{12}\sigma_1\sigma_2}{v_1\sigma_2^2+v_2\sigma_1^2-(v_1+v_2)\rho_{12}\sigma_1\sigma_2} \quad (v_1=\mu_1-\mu_f, v_2=\mu_2-\mu_f)$$

2. 根据w_T获得风险资产标准差σ_T,μ_T
3. 根据给定的风险/收益率约束,获得风险资产占比
   - 给定风险约束,则风险资产总占比权重 $w=\frac{\sigma_{target}}{\sigma_T}$
   - 给定收益率目标,则获得风险资产总比权重$w=\frac{\mu_{target}-\mu_f}{\mu_T-\mu_f}$
4. 每个风险资产的比重为$w\cdot w_T$

6.4 混合资产定价模型 (CAPM)

Capital Market Line(CML)

- 通过无风险收益率和波动率描述资产收益率
- 说明了夏普率相等,即增加风险资产配比时,超额收益永远和风险成正比

$$\mu_R=\mu_f+\frac{\mu_M-\mu_f}{\sigma_M}\sigma_R \leftrightarrow \frac{\mu_M-\mu_f}{\sigma_M}=\frac{\mu_R-\mu_f}{\sigma_R}$$

- 给定VaR_q和收益率μ_m的q分位数获得配比资产w

$$P((1-w)\mu_f+w\mu_m\le -V\tilde{a}R_q)=q$$
$$P(\mu_m\le\frac{(w-1)\mu_f-V\tilde{a}R_q}{w})=q$$
$$w=\frac{V\tilde{a}R_q+\mu_f}{\mu_f-\Phi_q}$$

β和Security Market Line (SML)

- 对于任意资产,β_R的定义和估计值分别有

$$\beta_R = \frac{\mathrm{Cov}(R, R_M)}{\sigma_M^2}$$

- 通过β可以给任意资产定义Security Market Line: $\mu_R - \mu_f = \beta_R(\mu_M - \mu_f)$
- 通过对数据做线性回归,可以估计$\hat{\beta}_R$,并带入到SML中估计收益率

$$\hat{\beta}_R = \frac{\sum_{t=1}^{T}(R_t - \bar{R})(R_{Mt} - \bar{R}_M)}{\sum_{t=1}^{T}(R_{Mt} - \bar{R}_M)^2}$$

- 和1做对比β_R越大,则风险越高

资产收益率拟合

- 通过线性回归,可以对i资产收益率进行拟合

$$\mu_i - \mu_f = \alpha_i + \beta_i(\mu_m - \mu_f) + \epsilon_i$$

- 在CAPM假设成立的情况下,α_i应该为0,如果α_i大于0,则说明μ_i被低估了
- 通过分析β_i的R^2,可以验证收益率的来源是市场还是误差

6.5 风险分析

风险计算

- 对于资产i,则有$\sigma_i^2 = \beta_i^2\sigma_M^2 + \sigma_\epsilon^2$
  - 市场系统性风险: $\beta_i^2\sigma_M^2$,可以通过对冲减少
  - 非市场风险: σ_ϵ^2,可以通过分散投资减少
- 对于市场协方差,则有$\sigma_{iM} = \beta_i\sigma_M^2$
- 对于任意资产i, j则有$\sigma_{ij} = \beta_i\beta_j\sigma_M^2$

Portfolio风险

- 给定权重为w的portfolio,则可以重新计算风险
- $\beta_P = \sum_{i=1}^{N} w_i\beta_i$
- 通过分散投资减少非市场风险: $\epsilon_P = \sum_{i=1}^{N} w_i\epsilon_i$

6.6 三因子模型

模型定义

- 通过添加其他参数作为回归项,来解释标的的收益率

$$\mu_i - \mu_f = \alpha_i + \beta_{i1}(\mu_m - \mu_f) + \beta_{i2}\cdot\mathrm{SMB} + \beta_{i3}\cdot\mathrm{HML} + \epsilon_i$$

- SMB: 小市值公司收益率更高
- HML: 低市净率的公司收益率更高

计算流程

1. 根据流通市值,将标的分成1:1的大市值(B)和小市值(S)组
2. 根据BM(市净率倒数)数据将标的按照3:4:3分成(H/M/L)三组
3. 通过市值加权计算每组中的平均收益率

4. 通过收益率计算因子
    - $\mathrm{SMB} = \frac{1}{3}(\mu_{SL} + \mu_{SM} + \mu_{SH}) - \frac{1}{3}(\mu_{BL} + \mu_{BM} + \mu_{BH})$
    - $\mathrm{HML} = \frac{1}{2}(\mu_{SH} + \mu_{BH}) - \frac{1}{2}(\mu_{SL} + \mu_{BL})$