

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，量邦科技与小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



量化投资组合管理

量化投资与机器学习系列课程

目录

- 阿尔法是什么东西
- A型阿尔法策略
- B型阿尔法策略
- X型阿尔法策略



量化投资组合管理 (一)

阿尔法是什么东西

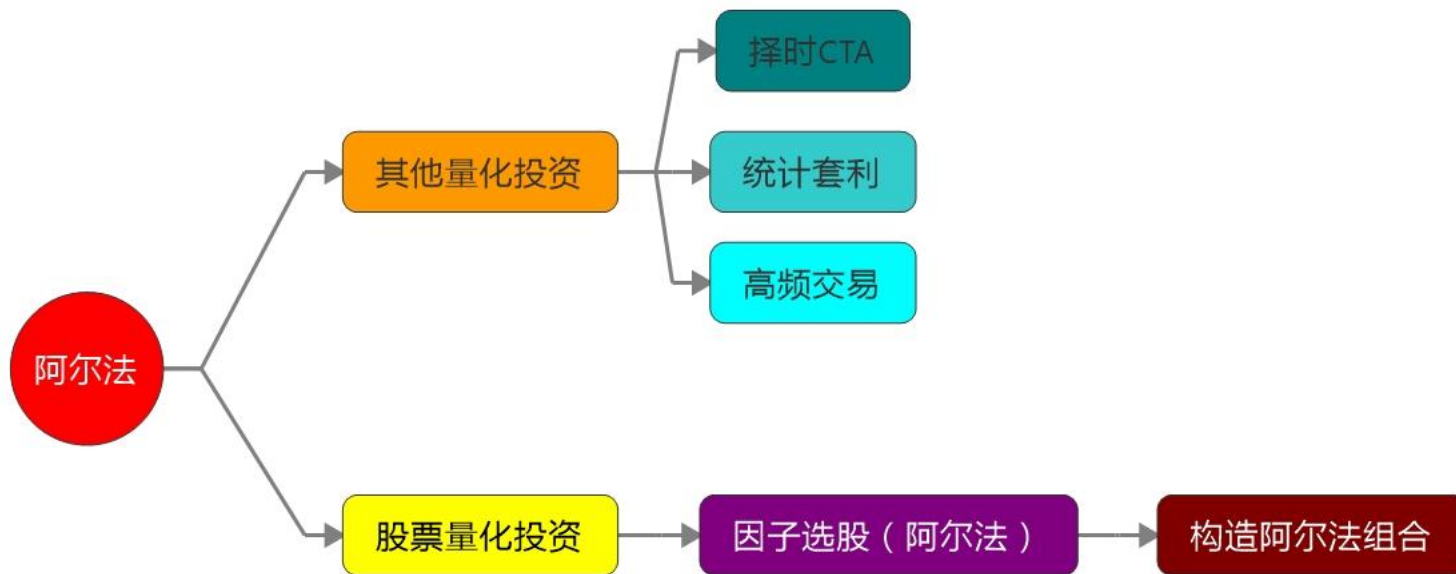


阿尔法到底是什么

- 股票量化投资是一个庞杂纷繁的问题，从来源于民间的技术分析到现代金融学的基石投资组合理论，从服务股民的大V战法和选股器到纵横华尔街的火箭科学博士和BARRA风险模型，似乎所有人都在追寻一只上帝之手，在瞬息万变的股票市场获得笑傲江湖的收益，究其本质，大家追求的这种东西叫做“阿尔法”，也就是一种所谓的“绝对收益能力”。绝对收益能力是真实的投资管理能力，这区别于主动承担大盘风险（泛指系统性风险、基准风险）的投资。
- 事后来看来，任何跑赢大盘或者基准的能力都叫做阿尔法能力，这也是所有量化投资策略追求的目标，不管是择时型CTA，高频交易，还是嵌入机器学习的智能算法，我们期待的收益就是绝对收益，我们获取的绝对收益就是我们的阿尔法。但是，在股票量化投资领域，追求阿尔法的方法比较特别，因为每一只股票的涨跌自然产生了相对于基准或者大盘的阿尔法，我们获取的阿尔法的主要手段是用某种指标或者方法找到这些股票，然后通过对冲大盘或者基准的方式，获取这些阿尔法，形成策略事实上的阿尔法能力。
- 因为收益率自带时间属性，谈论阿尔法时也必然涉及时间属性，在什么时间刻度上获取阿尔法非常重要，这也是我们后验（回测）时选择窗宽的关键。
- 股票量化最特别的地方在于，我们要用指标（选股因子）找到在一个时间区间内有阿尔法的股票，构造投资组合，对冲系统性风险，最终把这个阿尔法提取出来。换句话说，一般策略的阿尔法是通过各种交易手段获取的，而股票量化策略的阿尔法主要是通过选股因子选出来具有阿尔法的股票，自然获得的。



阿尔法到底是什么



股票阿尔法策略的本质是通过指标选取具有阶段性阿尔法的股票，通过在特定时点调仓从而获取整个投资区间的阿尔法收益。根据不同的选股思路，我们把股票阿尔法策略区分为A型阿尔法策略、X型阿尔法策略、B型阿尔法策略

阿尔法策略分类

- A型阿尔法策略：最直观的阿尔法策略，就是用指标对股票排序，选取其中一个组合，定期调仓，获取阶段性超越大盘的收益。策略永远满仓，但需要股指期货对冲。这是规范学术研究中用的最多的阿尔法策略。
- X型阿尔法策略：最常见的阿尔法策略，来源于技术分析和民间，也叫做战法，不区分选股和择时，往往通过择时指标来选股，也就是把择时或有上涨的股票选出来，持有一段时间，时间不确定，通过择时或者止损重构组合。
- B型阿尔法策略：基本不用指标选股，对所有股票建立可以解释波动来源的线性风险模型，然后通过表达对风险因素未来走势的看法，优化目标投资组合整体承担的各种风险暴露，这样自然确定了股票的权重，选择出来了股票，这种阿尔法策略，其实也是Smart Beta策略。



量化投资组合管理

A型阿尔法策略



单因子策略

□ 常见因子示例

- ① 盈利类因子：销售净利率、净资产收益率、总资产收益率、毛利率、营业费用比例、财务费用比例、息税前利润与营业收入比
- ② 成长类因子：每股净资产增长率、主营业务收入增长率、净利润增长率、总资产增长率、股东权益增长率、经营活动产生的现金流量净额增长率
- ③ 杠杆类因子：负债权益比、资产负债率、长期负债比率、流动负债比
- ④ 流动性因子：1个月平均换手率、3个月平均换手率
- ⑤ 动量因子：1个月股价动量、3个月股价动量、6个月股价动量、12个月股价动量
- ⑥ 规模因子：流通与总市值比、市值自然对数、流通市值自然对数、总资产自然对数、股权集中度
- ⑦ 红利因子：股息率、股息支付率
- ⑧ 现金流因子：经营活动产生的现金流量净额比总市值、经营活动产生的现金流量净额比营业收入、经营活动产生的现金流量净额比营业收入净收益
- ⑨ 相对估值因子：相对PE、相对PB、相对PS、相对PCF
- ⑩ 波动性因子：6个月振幅、12个月振幅、3个月日收益标准差、6个月日收益标准差
- ⑪ 一致预期因子：分析师预测净利润增长、分析师预测主营业务增长



因子分类

技术因子：

根据股票行情构造的技术分析类因子（主要包括趋势类、反转类、量能类、情绪类）；

① 分析师因子：

根据分析师对股票的预测结果构造的因子；

② 事件因子：

根据新闻报道、突发事件构造的因子；

③ 舆情因子：

根据舆情讨论热度监控等构造的因子；

④

单因子策略

□ 单因子策略

之所以能获取Alpha超额收益，是因为市场上各股票的收益存在分化，表现出不同的分布，从而可以通过某种恰当的因子将股票进行划分，赚取股票与股票之间的收益差、或股票与大盘之间的收益差。

其主要步骤如下：

- ① 选取合适的因子将不同的股票加以区分
- ② 研判因子值处于不同区间的股票的收益情况，持有因子值处于某一恰当区间的股票，以获得超越市场平均收益的超额收益
- ③ 在此基础上还可通过做空股指期货、或融券做空因子值处于另一恰当区间的股票，来获取相对更为稳定的alpha收益

单因子策略

□ 以价值因子PB为例

- 众多研究表明，高账面市值比公司的股票倾向于具有较高的收益，而低账面市值比公司的股票则通常具有相对较低的收益；上市公司的账面市值比反映了某种风险特征，部分投资者愿意支付一定的风险溢价来回避相应的风险，由此，价值因子从长期来看，可以获得统计显著的超额收益
- 市净率PB是账面市值比的倒数，这里通过选取低PB的股票来获取超额收益，其主要步骤如下：
 - ① 获取市场上各股票的市值和股东权益合计（即账面价值）；
 - ② 计算各股票的市净率PB；
 - ③ 按PB大小对股票进行排序；
 - ④ 选取PB较小的若干股票构建投资组合；
 - ⑤ （可选：做空股指期货或PB较大的若干股票进行对冲）

组合构建

□ 组合构建—单因子排序法

- ① 在目标股票池中使用因子暴露进行排序；
- ② 将排序后的股票池划分为合适的等份；
- ③ 根据因子测试结果选择“头部等份”或“尾部等份”（在构造多空组合时可以两部分兼而有之）；

	A	B
1	SecuCode	FactorExposure
2	601398	0.7565
3	600015	0.7456
4	601328	0.7242
5	601988	0.7147
6	600104	0.6813
7	600000	0.6649
8	601169	0.6213
9	000625	0.6208
10	600036	0.6195
11	600958	0.5894
12	000000	0.5555
13	002736	0.5021
14	000002	0.5003
15	600383	0.4918
16	600519	0.489
17	000024	0.3881
18	601985	0.3792
19	603885	0.3706
20	601766	0.3587
21	002739	0.3222

头部组合

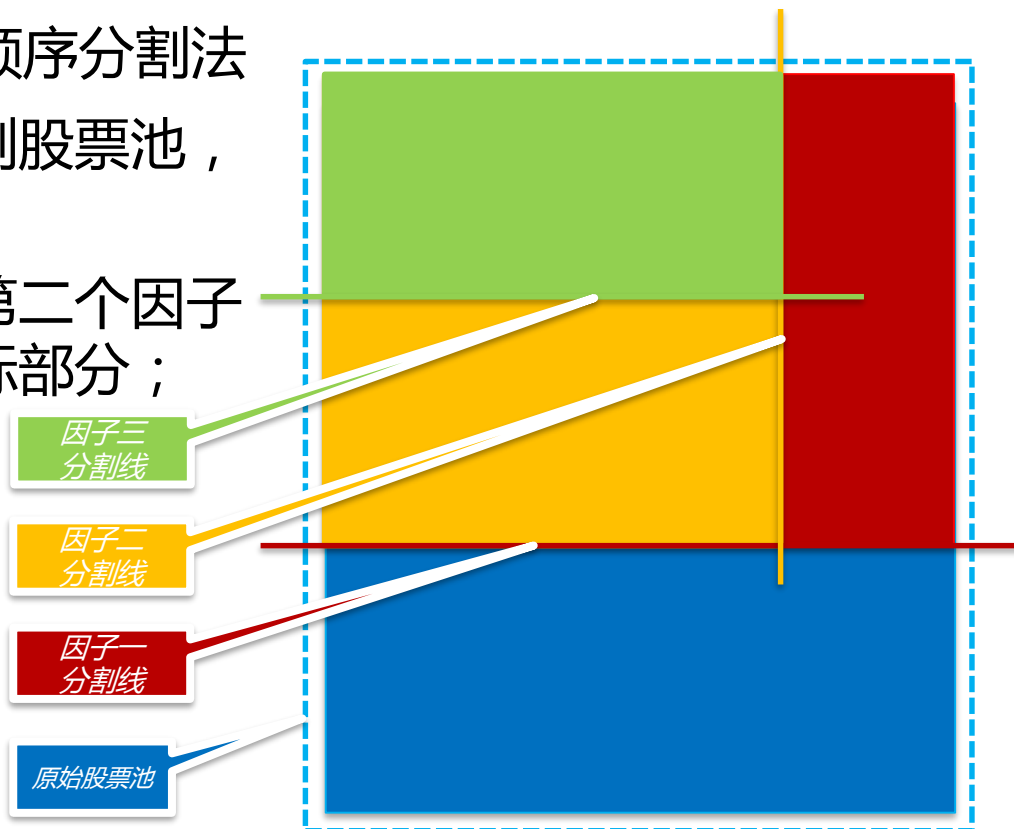
尾部组合

组合构建

□ 组合构建—多因子顺序分割法

- ① 使用第一个因子分割股票池，选取目标部分；
- ② 在选取部分再使用第二个因子进行分割，选取目标部分；
- ③

重复上面的步骤直至使用所有因子。



组合构建

□ 组合构建—多因子累积Z-Score法

- ① 在股票池中分别计算每个因子的平均值和标准差；
- ② 分别对每只股票的每个因子值进行正态化，得到它们的Z-Score：

$$Z_{i,k} = \frac{\beta_{i,k} - \mu_k}{\sigma_k}$$

其中 w_k 为因子比重，因子比重的确定方式：

- ① 等比重分配；
- ② 根据经验分配；
- ③ 根据历史数据优化；

- ③ 对每只股票计算累积Z-Score：

$$Z_i = \sum_{k=1}^K w_k Z_{i,k}$$

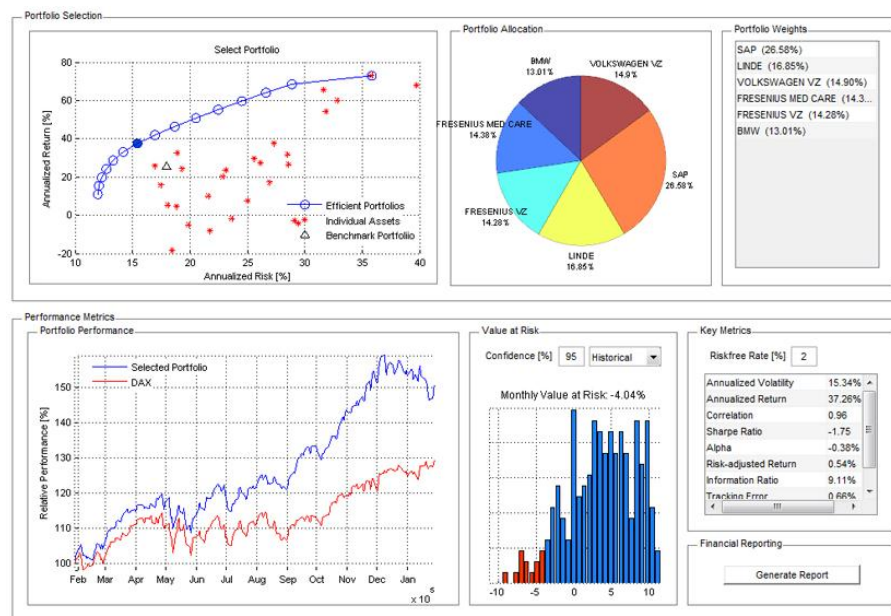
- ④ 根据每只股票的累积Z-Score划分股票池，选取目标部分构建组合。

组合构建

□ 投资权重分配

组合构建完成后，对成份股投资权重的分配方法主要包括：

- ① 等权重：最简单直接的方法；
- ② 市值权重：借鉴于市场指数的权重分配方式；
- ③ 行业限制权重：限制行业整体权重，在行业内部进行权重再分配；（同理可以使用规模、估值等其他特征的限制权重）
- ④ 均值方差优化权重：根据股票收益和波动信息计算出“最优组合”权重；（详见背景知识章节）
- ⑤ 高级权重优化技术：BETA控制、风险暴露控制等。

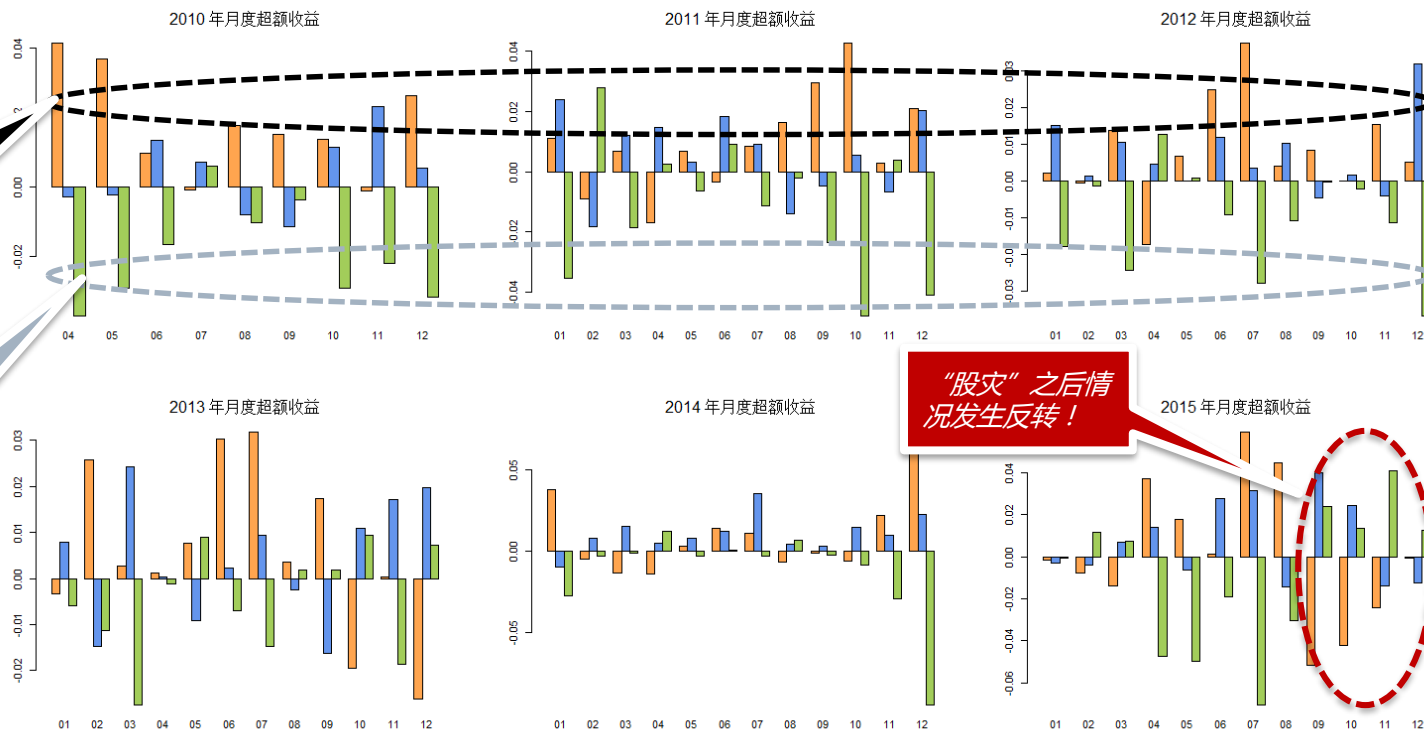


组合构建

□ 动量因子实例后验结果—三等份组合月度超额收益

头部组合绝大多数月份为正超额收益，说明它有较大概率跑赢大盘，适合进行做多套利交易

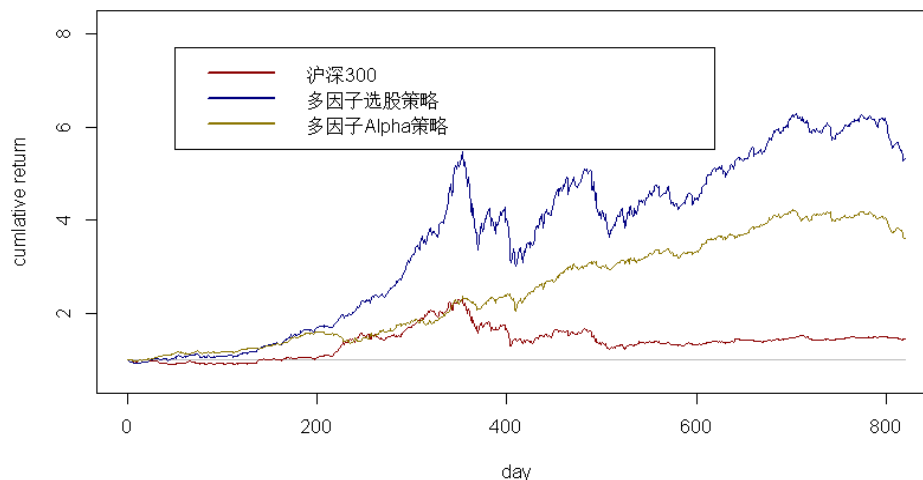
尾部组合绝大多数月份为负超额收益，说明它有较大概率跑输大盘，适合进行做空套利交易



多因子策略

□ 多因子策略

- ① 单因子模型所生产的因子往往不够纯粹，常常包含了多方面的风险因素
- ② 通过多因子模型不仅能更清晰地解构资产组合的风险来源，而且可通过恰当的因子组合，提高alpha的收益，并降低不必要的波动，从而提升整个组合的夏普比率
- ③ 如下为结合流动性因子、非线性规模因子和动量因子进行选股的示例策略的表现：



因子选择

□ 因子模型有效性检测

- 对于经济因子模型：使用单纬和多维零投资组合测试

在单纬零投资组合测试中，

- ① 将股票池根据因子暴露划分为3等份（或5等份、10等份，视情况而定）；
- ② 计算第一等份和最后等份中等权重股票组合的收益率，两者之差即为零投资组合收益率；
- ③ 重复上面的过程进行多期计算，得到零投资组合收益率序列；
- ④ 对零投资组合收益率序列进行t检验，若拒绝原假设（假设均值为零）则表明该因子有效。

在多纬零投资组合测试中，

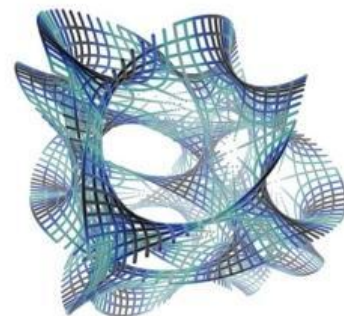
- ① 使用多个因子进行股票池的划分；
- ② 根据测试目标选择合适的不同等份构造零投资组合；
- ③ 进行与上面相同的测试步骤。

因子选择

□ 降低因子数量

主要原因：

- ① 因子越多，引入的不确定性就越大；
- ② 过多的因子带来的噪声会降低模型的信噪比；
- ③ “高维诅咒”（Curse of Dimensions）：
纬度越高，空间被切割得越细，应用模型所需的观测呈幂数增加，导致实际观测难以匹配，模型失效；
- ④ 同质化的因子无法提供增量信息，反而影响模型精度。



因子选择

□ 降低因子数量

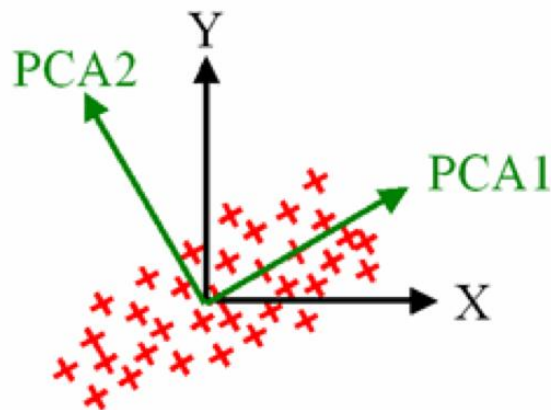
主要方法：

① 因子排除

- a) 相关性：根据相关性计算结果排除相关性高的因子（这里可以使用一般相关性和排序相关性）
- b) “距离”：构造因子距离函数，排除距离过近的因子

② 因子聚合

- a) PCA：构造一系列正交的线性因子组合作为新的因子



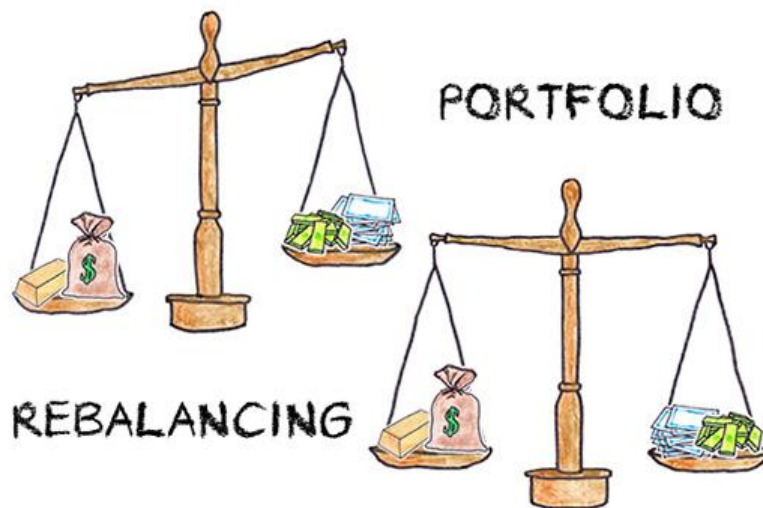
调仓与交易

□ 何时调仓？

- ① 固定时间调仓：最简单直接的方法；
- ② 根据因子特性调仓：经济因子适用于较长的周期，技术因子适用于较短周期；
- ③ 根据选股模型结果调仓：设定阈值，譬如当组合中有10%股票变动时调仓；
- ④ 根据权重优化结果调仓：设定阈值，譬如当整体权重变化10%时调仓；

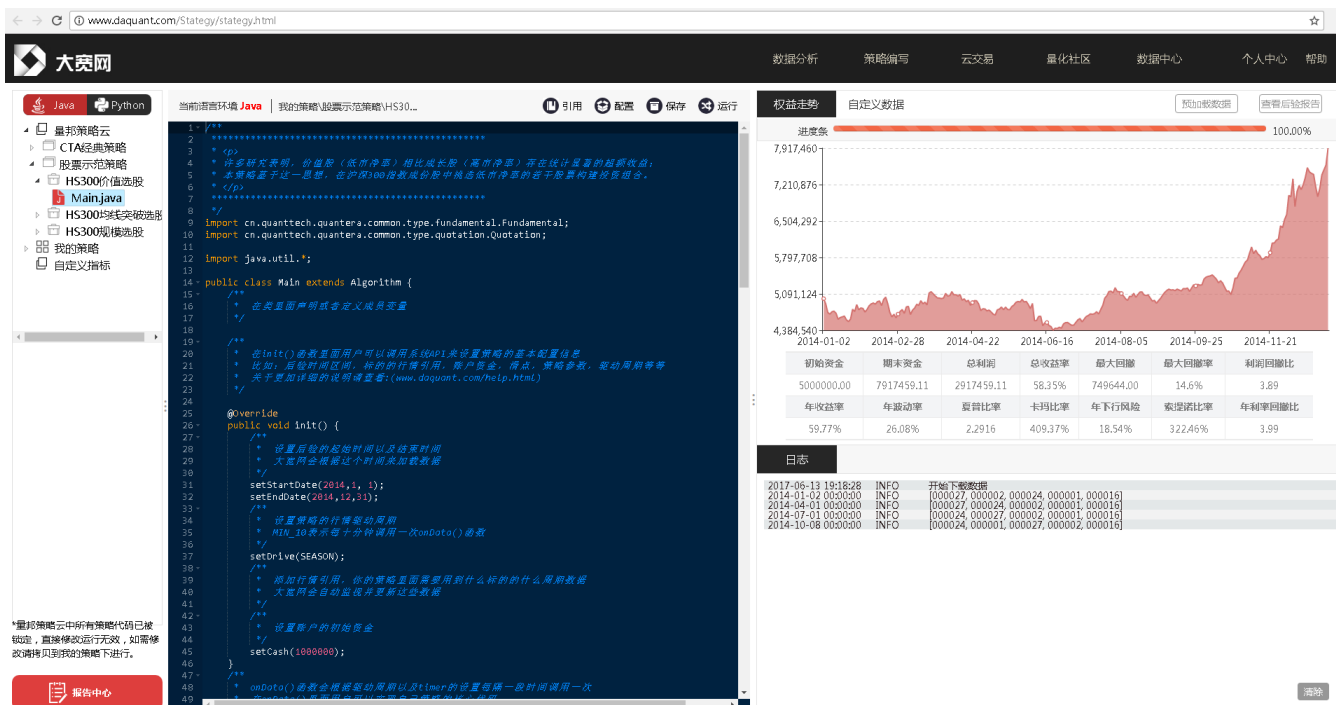
调仓时必需关注的问题：

交易费用 v.s. 预期收益



策略实现

在大宽网实现因子选股策略 支持 Java 与 Python 两种语言

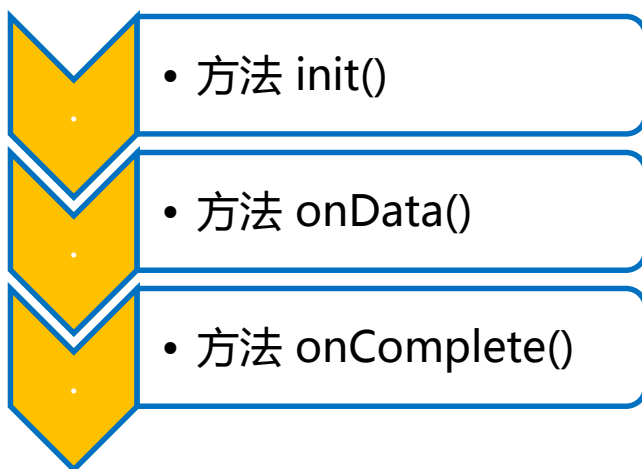


策略实现

□ 在大宽网实现因子选股策略

对大宽网而言，策略就是一个代码结构完整的程序。程序需实现三个方法（函数）：init(), onData() 及 onComplete()。具体的策略思想、交易条件在 onData() 中实现

策略主程序 Main Class



- init() 方法用于处理策略的前置条件：
定义策略后验的开始结束时间;
设定策略标的、驱动周期、初始资金、滑点;
- onData() 方法是整个策略中的重点，
onData 方法在行情数据到来时调用，并处理内部逻辑。在 onData 方法里需要实现交易思想，开平仓规则。
- onComplete() 方法在程序运行完成时调用，不作数据处理。

策略实现

□ 在大宽网实现PB价值因子选股策略

```
11 class Main(Algorithm):  
12     # 在类里面声明或者定义成员变量  
13  
14     def init(self):  
15         # 设置后验的起始时间以及结束时间  
16         # 大宽网会根据这个时间来加载数据  
17  
18         self.setStartDate(2014, 1, 1)  
19         self.setEndDate(2014, 12, 31)  
20  
21         # 设置策略的行情驱动周期  
22         # MIN_10表示每十分钟调用一次onData()函数  
23  
24         self.setDrive(TimeFrame.SEASON)  
25  
26         # 设置账户的初始资金  
27         # self.setCash(10000)  
28
```

init() 方法

设置策略后验起始与结束时间

设置策略的驱动周期，即隔多长的周期驱动一次 onData() 方法，例如这里是每季度驱动一次

设置策略的初始资金，股票组合策略默认初始资金500万，无需设置



策略实现

```
29 # onData() 函数会根据驱动周期以及timer的设置每隔一段时间调用一次
30 # 在onData()里面用户可以实现自己策略的核心代码
31 def onData(self):
32
33     # access HS300 components stock list
34     HS300 = self.getIndexStocks("000300")[0:9]
35
36     # access quotation data and financial report data of HS300 components
37     # calculate corresponding book-to-market ratios and store them into a dictionary
38     bookToMarket = {}
39     for iterCode in HS300:
40         # self.info(iterCode)
41         iterClosePrice = self.PickQuoteData(iterCode, TimeFrame.DAY, self.time().toLocalDate()).price()
42         # self.info(iterClosePrice)
43         iterFundamentals = self.getFundamentals( iterCode, Category.BALANCE, self.time().toLocalDate().plusDays(-180), self.time().toLocalDate() )
44         if iterFundamentals:
45             iterTotalStock = (iterFundamentals[0]).totalStock
46             # self.info(iterTotalStock)
47             iterTotalOwnerEquity = (iterFundamentals[0]).totalOwnerEquity
48             # self.info(iterTotalOwnerEquity)
49             iterBM = iterTotalOwnerEquity / iterClosePrice / iterTotalStock
50             # self.info(iterBM)
51             bookToMarket.update({iterCode:iterBM})
52         else:
53             self.info("Fundamental data of " + str(iterCode) + " is empty.")
54     # self.info(bookToMarket)
55
56     # sort book-to-market ratio in descending order
57     sortedBM = sorted(bookToMarket.items(), key = lambda d:d[1], reverse = True)
58     self.info(sortedBM)
59
60     # select the 5 highest book-to-market ratio stocks and stock them into a list
61     selectedStockPool = []
62     for iterSortedResult in sortedBM:
63         selectedStockPool.append(iterSortedResult[0])
64         if len(selectedStockPool) >= 5:
65             break
66     # self.info(selectedStockPool)
67
68     # set equal portfolio weight for selectedStockPool
69     self.lazyRecordWeightBook(self.createEqualWeight(selectedStockPool),self.time().toLocalDate())
```

onData() 方法

获取沪深300成份股

获取沪深300成份股的收盘价、股本、股东权益合计等数据，并计算相应的账面市值比

按账面市值比对股票进行排序，筛选账面市值比最高的若干股票

给筛选出的股票设置权重，构建股票组合，这里设置了等权重

策略实现

□ 在大宽网实现因子选股策略

```
77
78 # onComplete() 在策略运行结束的时候调用
79 def onComplete(self):
80
81     self.lazyEvaluate()
82     self.info('策略运行完毕')
83
```

onComplete() 方法

lazyEvaluate() 方法根据 onData() 方法中的 lazyRecordWeightBook() 方法所生成的股票权重进行相应的股票买入卖出操作

策略实现

□ 在大宽网实现因子选股策略

策略后验及报告



量化投资组合管理

B型阿尔法策略

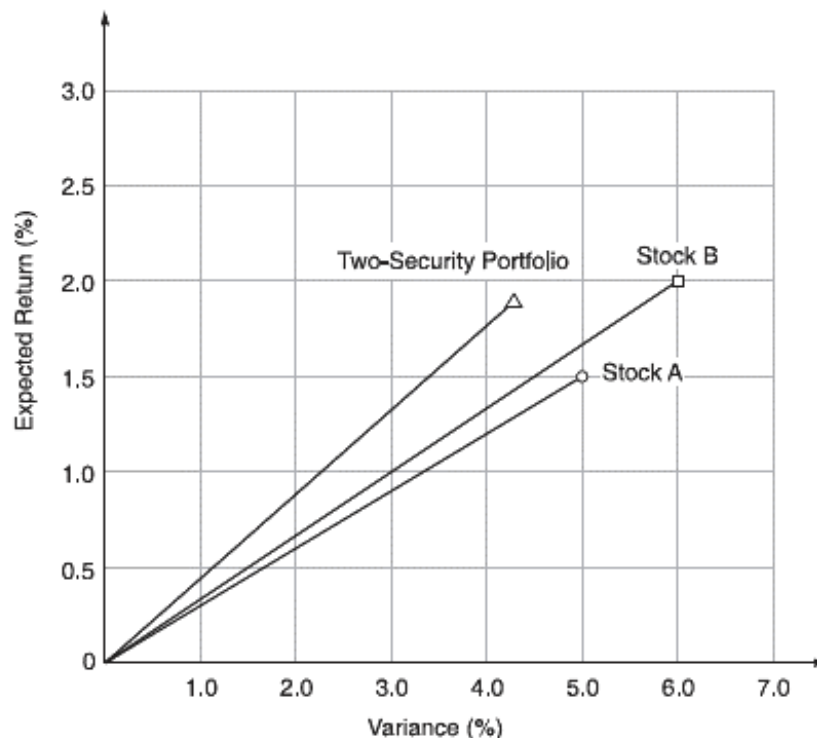
背景知识

□ 收益与风险：收益率与波动率（收益率标准差）

投资组合情形：

$$\begin{aligned} r_P &= \sum_{i=1}^n w_i r_i \\ &= w^T r \text{ (矩阵表达式)} \end{aligned}$$

$$\begin{aligned} \sigma_P^2 &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \rho_{ij} \sigma_i \sigma_j \\ &= w^T \Sigma w \text{ (矩阵表达式)} \end{aligned}$$

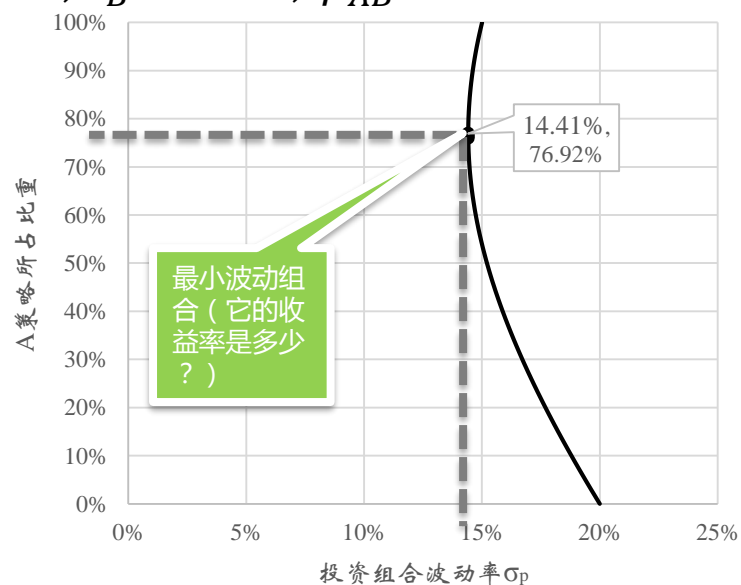
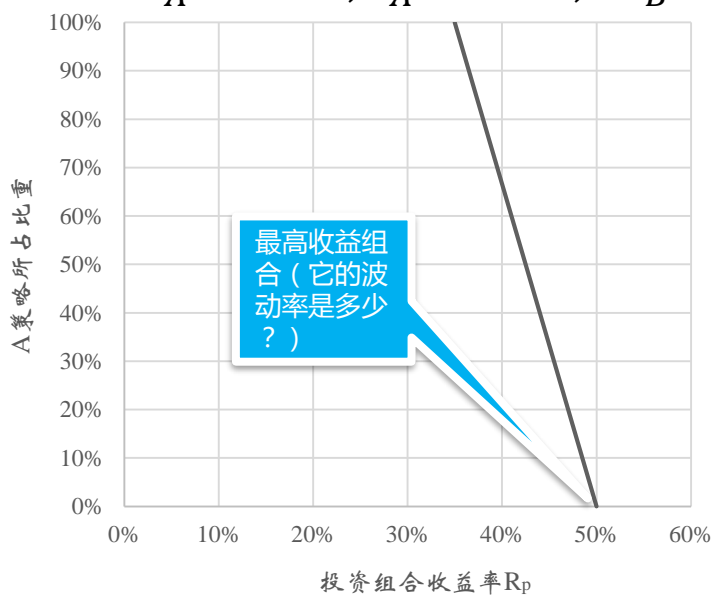


背景知识

□ 一组股票可以构成多少种投资组合？哪个组合最好？

两只股票（A和B）的情形：

$$R_A = 35\%, \sigma_A = 15\%; R_B = 50\%, \sigma_B = 20\%; \rho_{AB} = 0.5$$



背景知识

□ 股票之间的相关性对投资组合有什么影响？

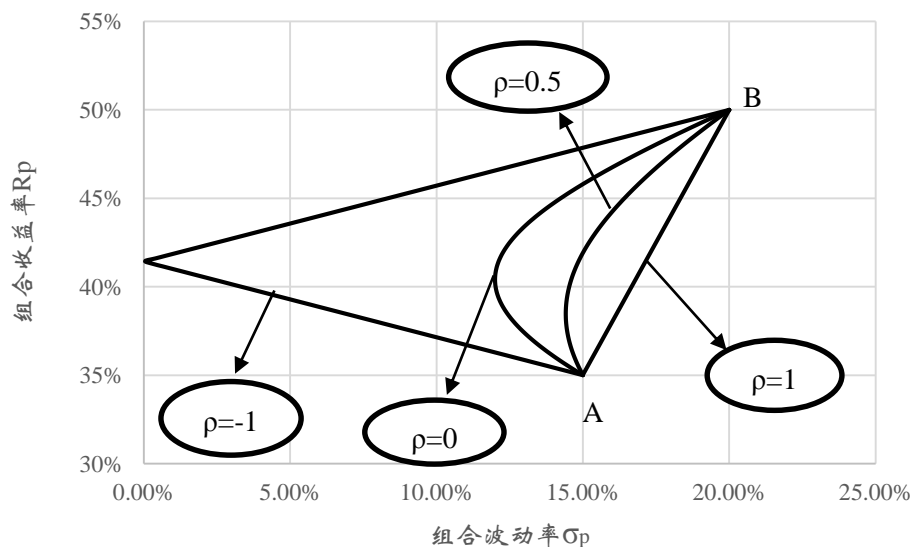
两只股票（A和B）的情形：

$$R_A = 35\%, \sigma_A = 15\%; R_B = 50\%, \sigma_B = 20\%;$$

不同相关系数下投资组合表现

图中哪个组合：

- 收益率最高？
- 波动率最低？
- 最好？



背景知识

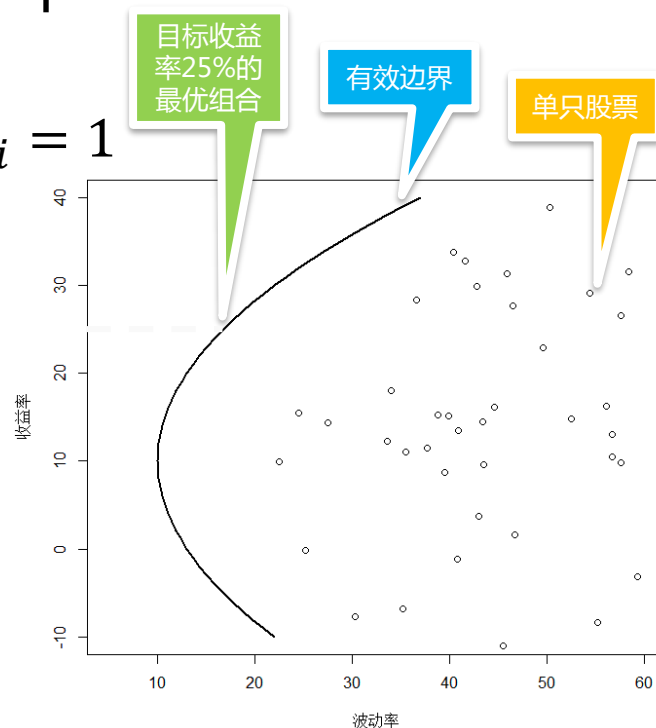
□ 均值方差优化——“最优组合”选择方法 Markowitz Mean-Variance Optimization

优化目标： $\min\{w'\Sigma w\}$

约束条件： $w'r = r_{target}$ and $\sum_{i=1}^n w_i = 1$

优化结果：

- ① 对于每一个目标收益率 (r_{target}) 都可以得到一个最优组合，这些最优组合构成一条“弧线”——称之为“有效边界”；
- ② 无法构造出“位于有效边界之外”（图中在有效边界左边）的投资组合；
- ③ 设定一个目标收益率，有效边界上满足这个收益率的组合即为风险最优组合。



背景知识

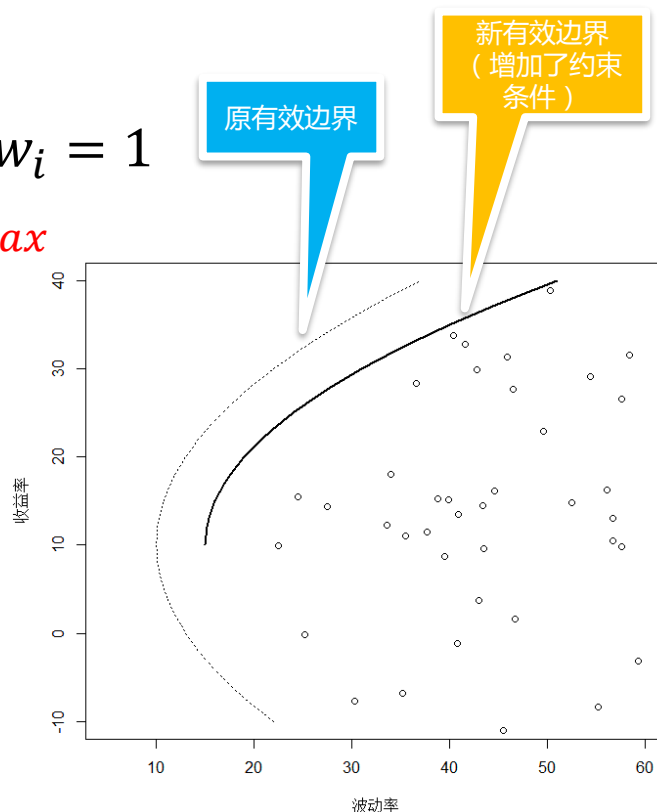
□ 均值方差优化—约束条件的影响

优化目标： $\min\{w'\Sigma w\}$

约束条件： $w'r = r_{target}$ and $\sum_{i=1}^n w_i = 1$
and $c_{min} \leq w_i \leq c_{max}$

结果：

- ① 新有效边界比原约束条件下的有效边界向右移动了，也就是说，能得到的最优组合有较小的收益率（或更大的波动率）。一般来说，约束条件越多，这种情况越明显。那么为什么还要约束？
 - a) 投资手段限制：无法进行卖空交易；
 - b) 标的仓位限制：必须在预定的范围内；
 - c) 参数（ r_i, Σ ）估计的不确定性：不增加约束代表完全信任估计参数，一旦估计不准，优化出来的组合表现有可能非常糟糕。



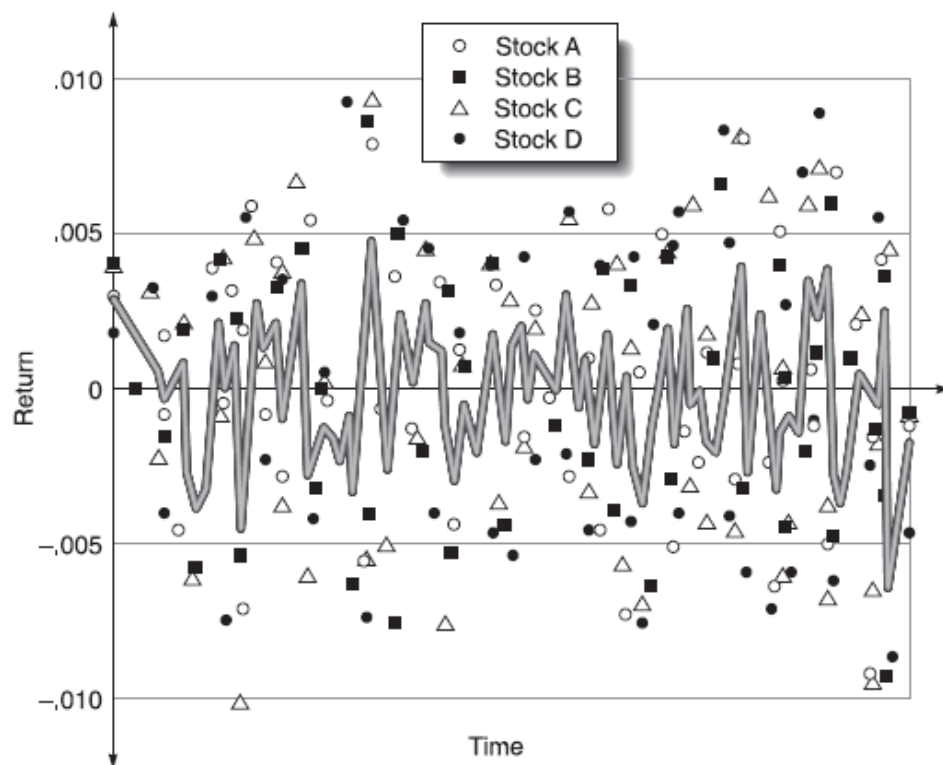
背景知识

□ 分散的效用

通过优化方法构建的有效股票组合可以在保持收益能力的情况下减小风险程度

实例：

右图中4只股票的波动率为16%，由它们构成的组合收益曲线明显更加稳定，波动率仅为5%！

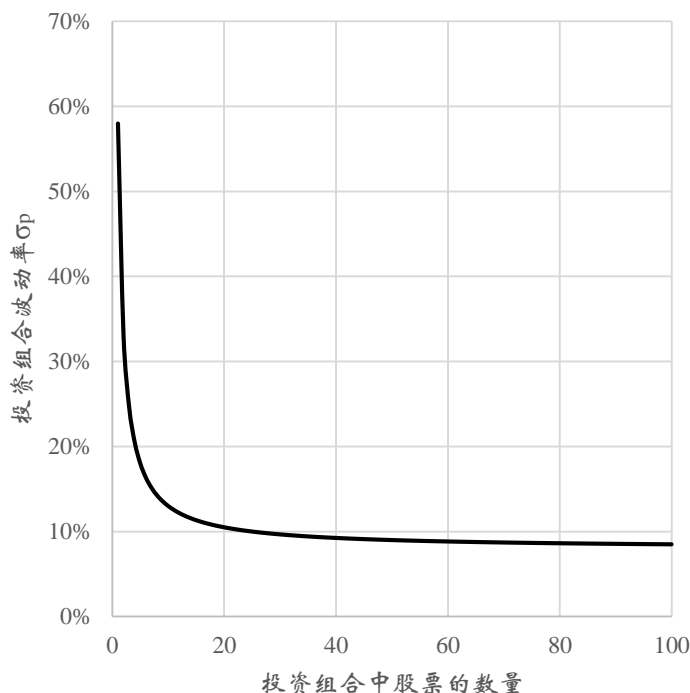


背景知识

□ 分散的局限

Evans-Archer的研究结果：

- ① 随着投资组合中股票数目的增加，组合波动率不断减小，不断的分散的确带来了好处；
- ② 分散的好处在组合中股票数目较少时较为明显，这时增加股票能够得到非常大的风险降低，但组合中股票越多的话这种风险降低效应会越小；
- ③ 当组合中股票数目达到一定数目时，进入“饱和”状态（在图中这种状态大致在20只股票左右），再增加股票所得到的风险降低效应微乎其微。



CAPM模型

□ Capital Asset Pricing Model (CAPM), Sharpe (1964)

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f)$$

相对应的股票定价模型为：

$$R_{it} = \alpha_i + \beta_i R_{mt} + e_{it}$$

从中可以看出，

Total Return = Alpha + Beta*Market Return

Total Risk = Specific Risk + Beta*Systematic Risk

结论：

- ① 股票风险有2部分组成：个股风险和市场风险；
- ② BETA反映股票跟市场的关联程度；
- ③ 构建有效组合可以分散风险，但是无法同时消除所有的个股风险和市场风险；

CAPM模型

□ Capital Asset Pricing Model (CAPM)

Alpha和Beta的估算—线性回归

最直观和常用的方法是对股票收益率和市场收益率进行回归，回归参数即为该股票的Alpha和Beta

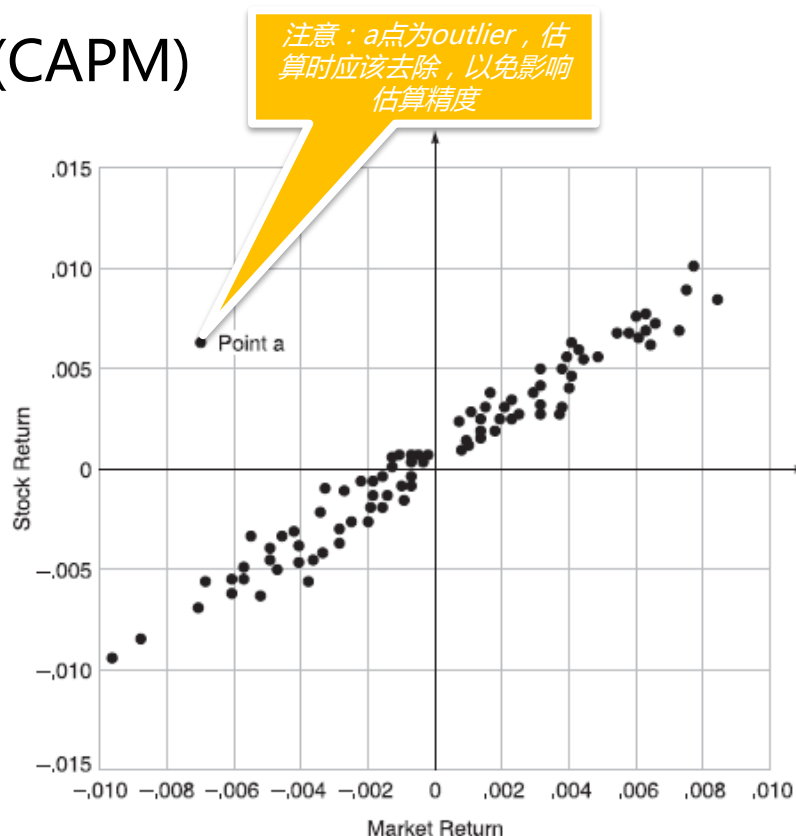
右图的估算结果为：

$$R_{it} = 0.001 + 1.060R_{mt} + e_{it}$$

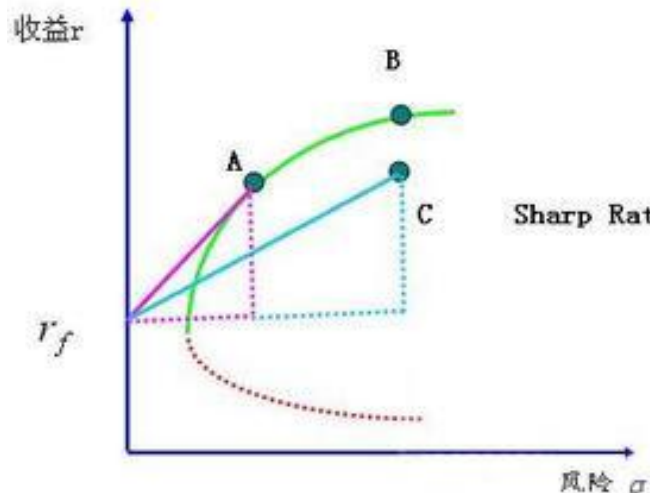
■ Alpha = 0.001

■ Beta = 1.060

市场收益变化1单位，股票收益变化1.06单位。



背景知识



- ❑ 风险预算
- ❑ 夏普比率：收益率除以标准差
- ❑ 无风险收益替换为投资基准收益，夏普比率变成信息比率
- ❑ 定性投资者觉得年化收益率最实在，量化投资者看夏普比率
- ❑ 奥妙在于，如果A策略的收益率低于B，但夏普比率高于B，A策略可以通过杠杆承担和B策略同样大的风险，但收益率会大大提高
- ❑ 夏普比率决定举债能力！

多因子模型

□ 多因子模型 (Multi-Factor Model)

以时间序列回归模型为例：

$$R_{it} = \alpha_i + \beta_{i1}f_{1t} + \cdots + \beta_{iK}f_{Kt} + e_{it}$$

β_{ik} ：因子暴露 (factor exposure)

f_{kt} ：因子溢价 (factor premium)

可以看出，多因子模型在市场因子是唯一因子时等价于CAPM

因子选择

□ 因子模型有效性检测

在使用因子构建投资组合之前，需要对因子在模型中的有效性进行检测！

测试方法：

- 对于基本面因子模型：可以使用单变量和多变量横截面回归

$$R_{it} = \alpha_t + \beta_{it}f_t + e_{it}$$

$$R_{it} = \alpha_t + \beta_{i1,t}f_{1,t} + \cdots + \beta_{iK,t}f_{K,t} + e_{it}$$

- ① 此时因子暴露 β_{it} 可直接观测，可通过横截面回归估计因子溢价
- ② 通过时间序列上的回归模型可以检测因子溢价的统计有效性
- ③ 统计学中的模型变量选择技术可被应用于此进行因子选择

因子分类

□ 一种常见的因子分类方式：

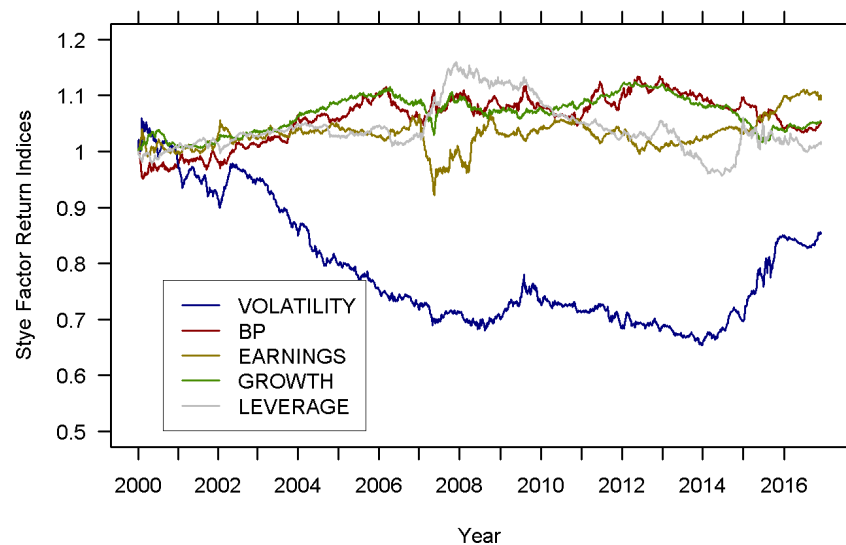
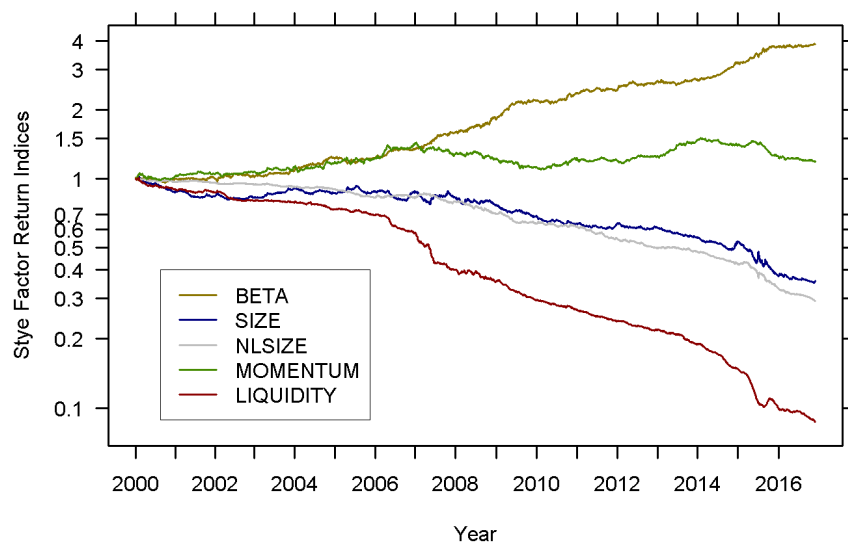
因子类型	输入	参数估计方法	输出
宏观经济因子	证券收益与宏观经济变量	时间序列回归	证券的因子beta
统计学因子	证券收益	迭代的时间序列或横截面回归	统计学因子和证券的因子beta
基本面因子	证券收益与证券特征	横截面回归	基本面因子

- ① 宏观经济因子示例：GDP增长率、失业率、通货膨胀率、消费指数、商品价格指数等；
- ② 基本面因子示例：公司规模、估值、运营效率、金融风险、流通能力、分红、行业属性等；
- ③ 统计学因子示例：运用主成份分析、因子分析等方法从证券收益的协方差矩阵中提炼因子

因子选择

□ 因子模型有效性检测

基本面因子溢价走势



量化投资组合管理

X型阿尔法策略



X型阿尔法策略

- 最常见的阿尔法策略，来源于技术分析和民间，也叫做战法，不区分选股和择时，往往通过择时指标来选股，也就是把择时或有上涨的股票选出来，持有一段时间，时间不确定，通过择时或者止损重构组合。

疑问

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他人回答问题

□ 量化投资策略研发平台：<http://www.daquant.com>

相关微信号推荐



量化投资华山之巅



量化投资大家学



联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：大数据分析挖掘
- 新浪微博：ChinaHadoop

