# 主题： which beta, which risk factors, which pricing model

**参考文献：**

1. Fama and French (2019)
2. [机器学习驱动的基本面量化投资研究 李斌 2019](https://zhuanlan.zhihu.com/p/88233451)
3. <https://www.zhihu.com/question/53768802/answer/460969682>
4. [which beta I 石川 2019](https://zhuanlan.zhihu.com/p/71072376)
5. [which beta II 石川 2019](https://zhuanlan.zhihu.com/p/84462587)
6. Fama and French(1993)
7. Fama and French(1992)
8. Carhart (1997)
9. Fama and French(2015)
10. <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html>
11. [FF和Barra体系，谁是你的菜 因子动物园 2019](https://mp.weixin.qq.com/s?__biz=MzUxNzY0NjU3Mw==&mid=2247484154&idx=1&sn=e8bbc417df2e0dcfd01999b545eed3a5&chksm=f995b378cee23a6ee62b176a6f14d81a6c3ce608717a46dcc5abe723404e00341ac06d329428&scene=21#wechat_redirect)
12. [Fama-Macbeth中的两步回归的原理](https://www.zhihu.com/question/53768802/answer/460969682)
13. common risk factor in the returns on stocks and bonds Fama and French(1993)

**概述：** 本研究旨在探讨多因子策略中定价模型（asset pricing model）的选择，风险因子（risk factor/ risk premium）的选择及其对应因子载荷（factor loadings）对策略有效性、成果的影响。FF2019以回归α作为考量模型优劣，对于策略，考虑用回测效果考量

1. **定价模型**

根据[参考文献1]FF2019，有以下三种定价模型可供选择（不参考因子类型，仅参考模型公式）：

1. 

通过对个股特征排序，选取头部底部10%股票构建多空组合作为对应特征的风险因子mimicking portfolio，其收益作factor return。即线性回归公式RHS自变量SMB, HML, RMW, CMA。 通过Time-Series回归可以得到每个股票i在选定窗口下的βs（因子载荷）。

例：假设当前在T时刻，考察monthly return（相较daily return更稳定），对前12个月窗口做时间序列回归t=T-1..T-12得到的betas，alpha估计量，作T时刻beta（逻辑链：假设过去窗口回归的betas估计量和T时刻的betas相同），sum（beta\*risk\_factor at T）+alpha = R\_iT-R\_fT

1. 

滚动TS回归得到各时期t的各股票i的betas（同1），对各期各股betas（RHS）与各期各股return-rf作横截面回归(不同于3,RHS 此处用factor loading （betas)作为input，而3用到个股characteristics)求得risk premium估计值（式中R\_MC, R\_BM, R\_OP, R\_INV）,设定窗口向t-n方向求平均值作为t时刻的risk premium。全过程成为FAMA MACBETH regression[参考文献3]。

例：该定价模型对T期Ri-Rf=sum(betas\_T \* risk\_premium\_T) +alpha（逻辑链risk\_premium\_T refers to risk\_premium\_T-1,assume unchanged）

1. 

对T-1期各股特征值（RHS）与T期各股return-Rz（市场因子[参考文献5]搜索关键词‘国家因子有异曲同工之妙’）作横截面回归求得T期risk premium估计值（式中R\_MC, R\_BM, R\_OP, R\_INV）,设定窗口向T-n方向求平均值作为T时刻的risk premium。注：此处CS回归考虑lag，即上一期char与本期return回归得到本期risk premium估计值

例：该定价模型对T期Ri-Rz=sum(chars\_T-1 \* risk\_premium\_T) +alpha（逻辑链risk\_premium\_T refers to risk\_premium\_T-1,assume unchanged）

1. **风险因子**

风险因子多样，FF1993三因子模型[参考文献6]采用market factor、size factor、value factor；[参考文献8] carhart1995四因子模型增加momentum factor; FF2015五因子[参考文献9]在三因子的基础上增加profitability factor、investment factor。

此外可考虑因子如（且不局限于）杠杆、成长、残差波动率、流动性

因子详细构建过程可[参考文献13]Pg8

美国股票市场FF因子数据提供可[参考文献10]

1. **因子载荷**

[参考文献11]若采用模型1、2（FF体系），因子载荷通过TS回归得到；若采用模型3（Barra体系），因子载荷由个股独有特征（characteristics替代），除市场因子外的其他由根据个股特征构建头尾多空组合所代表的风险因子的载荷都可以由个股char替代。问：哪种载荷作为beta所建立的定价模型更好？[参考文献1] [参考文献4][参考文献5]（文献5为文献1的解读）

从research的角度，通过比较回归由相同种类因子构建的定价模型得到的alpha大小评估定价模型优劣[参考文献1]

从策略的角度，通过比较回归由相同种类因子构建的定价模型得到的alpha大小评估定价模型生成的构建组合信号所产生的组合进行区间回测，由表现指标判断模型优劣。即下文内容：

1. **策略**

策略思路：根据任意定价模型，转译为交易策略时，是一个预测未来收益的过程，逻辑链为只要在合适的因子（这里用是否显著评价）有较好的暴露（对于数值为正的风险因子给与对应因子载荷排序头部的个股组合，负🡪尾部组合），该组合存在超额收益（超出市场部分 Jensen’s alpha）

步骤：

1. 模型确定

目标对比三个定价模型转译的交易策略优劣，三种模型逐一考虑

1. 因子确定

保持三个模型采用的因子统一的前提下，考虑多个因子组合,尝试单因子，双因子，三因子，四因子，五因子所有可能的组合;六七八因子………

**细节注意：**[参考文献7]pg439，现存因子可能会因为新加入因子而变得不显著，可能导致的结果三因子列表[因子1，因子2，因子3]，四因子列表[因子1，因子2，~~因子3~~，因子4]，由于因子3因为因子44的加入变得不显著导致四因子列表实际只剩三因子。但文献中是将因子拆解，可能是共线性导致该情况，而独立因子的加入可能不会造成这样上述可能的结果？待考证

**衍生：1.**全程用一致的因子可能不是最优的，甚至有时是不合适的。后期用机器学习（线性回归🡪非线性回归）输出最优因子列表（[参考文献2]表5），基于此列表动态调整各期策略因子**2.**可以考虑独立回归，而不是一起回归，就不会出现因为其他因子的加入造成的某因子长期不显著

1. 数据
   1. 原始数据

**PULIC：**各时点个股价格数据（用于求return）

对模型1、2请求现成的各时点风险因子收益数据（risk factor return）；对模型3请求各时点个股特征数据(characteristics)

* 1. 数据处理

对于模型1、2，若有现成risk factor数据库，可直接采用。若直接制作须[参考文献1]pg6过程，细节需谨慎

对于模型3，考虑两点：1.应采用哪个特征对应到特定risk factor（如盈利因子，用EBITDA/NET PROFIT..etc.?）2.具体lag多少期？（对本期的回归采用前面第几期的数据作RHS？偏主观可[参考文献7]pg439）

* 1. 数据计算

**PULIC：**Return calculation（log return/simple return?）

**模型1：**stock return(LHS)同risk factor(RHS）作[股票数量]次TS regression（para：TS window）确定最近一期各股betas

**模型2：**模型1步骤求得betas后，对各时间T对stock return（LHS），betas（RHS）做N（para：CS window）次横截面回归后得到risk premium estimator取mean()作为最近一期的risk factors return及对应T值

**模型3：**对各时间T对stock return（LHS），chars(RHS)做N（para：CS window）次横截面回归后得到risk premium estimator取mean()作为最近一期的risk factors return及对应T值

**细节注意：**按照模型，LHS是Stock return-Rf(model 1,2)/Rm(model 3),考虑是否简化rf

1. Betas分组

对最近一期各股betas用同一beta的分位数进行进行高低分组（para：group\_num）以量化beta相对大小，如用25%，50%，75%分位点分4组，第一组内的股票对应beta相对其他股票最高，第四组内股票对应beta相对最低

**衍生：**考虑非分组的其他量化高低排序的方法

1. 交易组合确定

若不考虑做空，对风险因子为正（若采用模型2、3还须满足risk premium显著），筛选股票在该风险因子的暴露为最大即对应beta所在组1，为负(同上条件)，筛选股票同步满足beta所在组为4。将共同筛选条件均满足的股票作为交易组合

若考虑做空，附加做空组合：对风险因子为正（同上条件)，筛选股票在该风险因子的暴露为最小即对应beta所在组4，为负(同上条件)，筛选股票同步满足beta所在组为1。将共同筛选条件均满足的股票作为临时做空组合，进一步筛选组合中可融券股票列表作为最终确定的做空组合

**细节注意：**

1. 过多因子存在同步筛选后无股票符合要求的情况，此时考虑空仓或放宽筛选要求，如从组1放宽至组2，组4放宽至组3
2. 筛选出来的组合数量小于20（para：portfolio\_stock\_min\_num）则考虑不构建组合即不交易，或放宽筛选条件。因为需足够分散以避免特异性风险

**衍生：**考虑其他筛选股票的方法，如综合权重打分，考虑因子大小。或者组合内权重偏向考虑因子大小

1. 策略执行

每个月月初执行一次（para: rebalancing freq=1M）

1. 回测：

对大涨部分单独回归，对大跌部分单独回归，对全时段回归

1. 参数优化
2. 模型确定
3. 因子确定
4. factor\_list

Default:[‘size’,’value’,’market’]

1. 数据
2. Return\_type

Default: monthly return

1. TS\_window

Default:12

1. CS\_window

Default:12

1. Betas分组
2. Group\_num

Default:4

1. 交易组合确定
2. T\_treshold\_val

Default:2

1. Portfolio\_stock\_min\_num

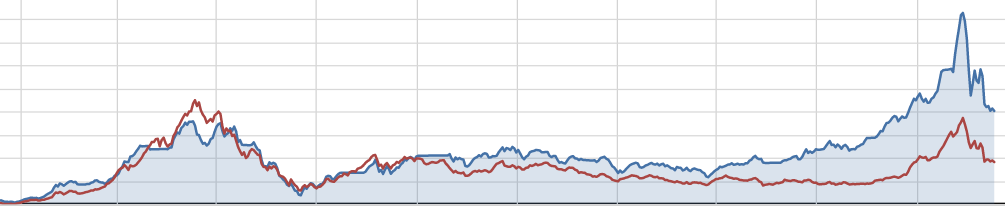
Default:20

1. 策略执行
2. Rebalance\_freq

affect return type in[3];e.g., 1M🡪montly return; 1W🡪weekly return

default: 1 month

试图理解：



factor list保持不变，在部分时段无超额收益，有时却很高？

猜测1：虽然显著但risk premium数值右侧大于左侧，是不是动态factor list能解决这个问题（找到当前risk premium有效且最高的组合）。动态factor list保证输出的factor在当前都是有效的，而静态factor list在各时刻有效的因子各不相同

在预期下跌（缺信号）时，1.采取空仓防御，2.换仓至beta为显著负的股票

建立A股risk factor组合机制 实时更新

6.4

FF1993 risk factor portfolio 构建逻辑搬运至A股，建立A股FF factor return database

前提概述：对因子选择的优化。我可以给定100因子模型，跑原策略，哪个有效用哪个因子累计筛选，但直接给定原定价模型合理吗？如果可以动态调整给定的因子模型是否会合理？即先给定有效因子列表，再在这个因子列表做CS，原100个因子可能有些都是无效因子但也被扔进来一起做回归了，先确定因子有效性，生成有效因子列表，再确定因子模型。此时涉及因子有效性检验，对所有因子逐一检验有效性，思路如下

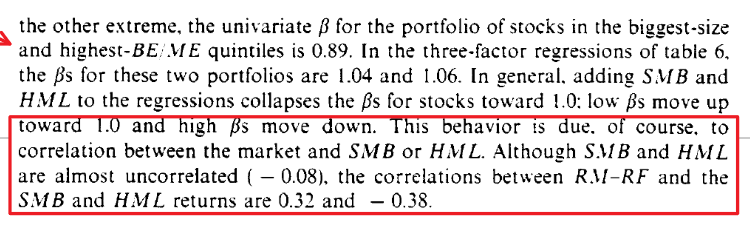
新的检验因子有效性思路，构建相关因子排序组合，检验某时间段内因子大小组合的回归beta（loading）是否类单调变化，若有该现象则因子有效。也就是说不看loading绝对大小，看相对大小及对应T值有效性（有效即可，不看T的绝对超过2的多少），检验HML 或SMB因子有效性时，按两个特征交叉分组5X5，

考虑不交叉直接按想要检验的因子直接分5组看5组的规律，因为对除这两个因子以外的其他因子检验没有交叉分组控制影响因子的其他变量的理由。

考虑用对应要检验的单因子模型检验，而非多因子，因为对于HML SMB因子 二因子模型三因子模型都有对应规律（FF1993 pg21），先检验这两个因子的单因子模型是否可以得到单调规律，若可以，考虑其他的因子检验也采用单因子，因为你并不知道该用最后确定几因子模型

若在所有有效的因子所构建的模型下回归，发现有的有效 有的无效，是不是因为covariance的原因，两个因子相关，比如同时有size和smb。

CS回归发现有因子无效，目前推断两个可能 1.因子在这段时间本身无效 2.两个相关因子造成其中一个因子无效。上述方法解决第一个原因，故若还有无效因子，考虑第二个原因。此时检验cov matrix。



另一个研究新加入因子导致原因子loading减小是因为两者相关，那么两者几乎不相关是否可以让loading几乎不减？考虑用HML SMB单因子与两者结合的双因子回归检验（先保证两因子return cov足够小）

构建定价模型可能出现的情况 双阴 变 双假阳 或单假阳，双阳变单阴单阳 或双阴 单阴单阳变假阳阳 假阳阴 阴阴 提前独立确定有效性再放入模型