

第1章 因子投资基础

1.1 统一视角下的因子投资

1.1.1 一个公式

1.1.2 因子、多因子模型和异象

1.1.3 再论异象和因子

1.1.4 因子投资包含的内容

1. 关于 $\beta_i \lambda$ 的研究

2. 关于 α_i 的研究

3. 截面vs时序

1.2 因子投资的学术起源

1.2.1 实证资产定价

1.2.2 研究现状

1.3 因子投资的业界发展

1.3.1 因子投资和管理人

1.3.2 因子投资和投资者

获取收益、控制风险、被动资产配置、主动套利、研究预期收益率、研究收益率波动.....

1.1 统一视角下的因子投资

1.1.1 一个公式

MM定理：不考虑税收、破产成本、信息不对称，且市场有效的假设下，企业价值不会因为企业融资方式改变而改变。

CAPM：资产的预期超额收益率由市场组合的预期超额收益率和资产对市场风险的暴露大小决定。

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

APT后的多因子模型：资产的预期收益率与一系列因子的预期收益率和资产在这些因子上的暴露决定，表明多因子模型研究的是不同资产 $E(r)$ 之间的差异，而非单一资产收益率在时间序列上的变化。

$$E(R_i^e) = \beta_i' \lambda$$

beta: 因子暴露、因子载荷, lambda: 因子预期收益、因子溢价

左方代表超额收益, 不再用 $-R_f$ 的写法, 因为多空资金中性组合的超额收益是long-short

因子模型是横截面模型: 研究不同资产之间预期收益率之间的差异 (不同资产在各个因子上的暴露不一样导致的)

实际中存在错误定价:

$$E(R_i^e) = \alpha_i + \beta_i' \lambda$$

alpha: 实际预期收益率 (应该可以理解为ex-post) 和多因子模型得到的预期收益率之间的定价误差 (pricing error)

原因:

- 1) 模型错了, 因子遗漏
- 2) 模型没问题, 数据有样本误差, 需要检验alpha是否显著为0
 - 2.1) if not, 运气, 样本偏误
 - 2.2) if it is, 代表套利机会, 市场对该资产存在错误定价, 该资产属于一个**异象**。

几个重要问题:

- 选择什么样的因子?
- 如何计算资产在因子上的暴露?
- 因子收益率如何计算?
- 如何使用统计方法对alpha进行检验?

1.1.2 因子、多因子模型和异象

因子: 描述众多资产共同暴露的某种系统性风险, 该风险是资产收益率背后的驱动力; 因子收益率是系统性风险的风险溢价or风险补偿, 是这些资产的共性收益。

因子需要满足两个**必要条件**:

- 1) 因子驱动资产收益率共同运动 (co-movement), hence 因子一定和资产收益率协方差矩阵有关;
- 2) 长期因子可以获得正收益, 意味着因子必须是被定价的。

(因子特征 (暴露) 系统性影响了资产的长期预期收益, 这就意味着该特征所代表的风险或特质, 影响了市场对资产要求的回报率 (Required Rate of Return)。换句话说, 市场在给资产定价时, 已经考虑到了该因子所代表的特质或风险, 并将其融入到了该资产的要求回报率和市场价格之中。这就是“被定价”)

如何确定因子模型中的因子、因子数?

- 剔除高度相关，保持因子相互独立，每个因子对解释资产横截面收益率有增量贡献，反应不同信息；
- 简约法则：因子个数有限。（其实在机器学习框架下，感觉这条有所放松，可以由机器学习解释众多因子的相关性等等）

异象：多空组合收益率无法被多因子模型解释，且显著大于0，则该资产/投资组合是一个异象，构建该组合的指标：异象变量（anomaly variable）。

1.1.3 再论异象和因子

从学术、业界角度区分一下异象与因子

lambda：定价因子 pricing factors（多因子模型中的因子）

alpha：异象因子 anomaly factors（按照该变量构建的多空组合收益率无法被多因子模型解释）

学术界严谨区分前二者，无法加入到多因子模型中的都称为异象。

业界统称为因子，只care是否能获得定价模型无法解释的超额收益（其实业界通常把由常见定价因子带来的收益叫做beta，其余的部分称为alpha，beta是由风格带来的收益，具有不稳定性，有时需要通过择时来获取）

1.1.4 因子投资包含的内容

1. 关于 $\beta_i' \lambda$ 的研究

学术：

- 1) 找到最好的多因子模型，模型无法解释的异象越少越好
- 2) 解释因子背后的原
- 3) 计量方法：计算因子暴露、因子收益率、对多因子模型进行假设检验
- 4) 对主动管理人进行业绩归因：绝大多数人不能获得alpha

业界

- 1) 获得超过基准的收益
- 2) 资产配置：找到高风险溢价的因子，尽可能高的暴露配置，获得稳健的超额收益

Smart Beta ETF金融产品问世

2. 关于 α_i 的研究

学术：

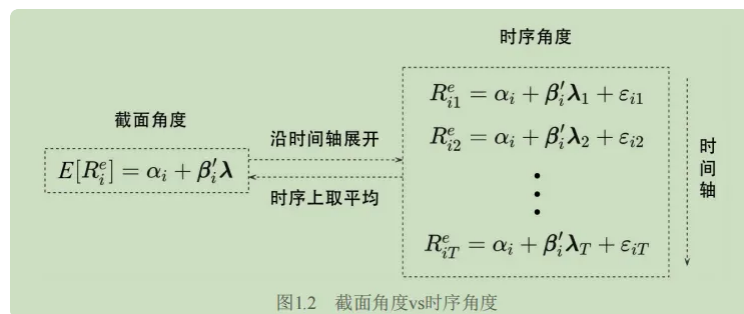
- 1) EMH：发现异象，评判市场是否有效
- 2) 评价多因子模型优劣，哪个异象少哪个好

（事实上这两点我感觉是相矛盾的，异象多可能是错误定价的原因，而不是因子模型不好的原因，市场也有可能是非有效的）

业界：不严格区分beta、alpha，关心考虑交易成本后，使用该因子是否能获得超额收益

3. 截面vs时序

多因子模型是关于均值的模型（mode of the mean），仅关心不同资产的收益率均值为什么会有差异，而非每个资产收益率如何随时间变化



（我觉得时序角度的研究，beta其实是在变化的）

时序角度：方差模型（model of the variance）和因子择时

方差模型：

$$R_t^e = \alpha + \beta \lambda_t + \varepsilon_t$$

N个资产的时序多元回归模型放在一起如上，每个字母都是一个资产向量

满足误差项的期望为0，与因子暴露的协方差为0

（我们通常讲MKT、SMB这种定价因子，讲的都是因子收益率的多空组合，不同股票在这些因子上的暴露程度是不同的，我们在谈异象因子时，通常谈的是因子暴露）

两侧求协方差矩阵后：

$$\Sigma = \beta \Sigma_f \beta' + \Sigma_\varepsilon \quad (1.6)$$

左边的N阶矩阵是N个资产的协方差矩阵，中间的那个是K个因子的协方差矩阵，最后面那个是N个随机扰动的协方差矩阵（对角阵，因为误差相互独立），学术界意味：因子必须和资产的协方差有关，验证第

一个必要条件。

目前热门的因子模型，或来自金融、经济、行为金融，都没有从因子协方差矩阵和资产协方差矩阵入手。（因子选择应使 $\beta' \Sigma \beta$ 的非对角线元素尽可能接近资产协方差的非对角元素）

Pukthuanthong et al. (2019) 指出了这种关联的重要性，并以此为依据提出了一个甄别真正因子的框架。

业界中，需要考虑投资组合波动率：weight，协方差矩阵，使用历史数据计算时，需要保证历史数据期数 $T >$ 资产数目 N ，否则协方差矩阵不可逆，当 N 很大时，可以将资产通过因子模型转化成因子 K ，简化上述问题。因子个数往往小于资产个数（否则复杂化问题， N 个资产自成因子）

还可以使用因子模型计算资产协方差矩阵进行风险管理：Barra CNE6（1国家因子、9风格因子、一系列行业因子）

因子择时：因子收益率如何随时间波动

学术：care不同因子和宏观经济、投资者情绪之间的关系

业界：不同时期配置好用的因子，不好的时候规避

1.2 因子投资的学术起源

Graham and Dodd (1934) 提出价值溢价，《Security Analysis》

Basu (1977) 便宜股效应、Banz (1981) 小市值效应

1.2.1 实证资产定价

Eugene Fama:

Fama (1970) 联合假说 (joint hypothesis) 问题：检验市场合理性必须先有一个合理的资产定价模型，只有知道定价模型给出的均衡状态下股票的预期收益率，才有可能正确的检验市场是否有效。

Fama and MacBeth (1973)：检验CAPM，提出Fama Macbeth回归

Fama and French (1992)、(1993)：Fama-French三因子模型

Hansen (1982)：广义矩估计 (Generalized Method of Moments Estimator, GMM) 检验CCAPM

Robert Shiller (1984)：噪声交易者模型：噪声交易者的存在造成价格和内在价值出现了偏移——行为金融，错误定价因子 (mispricing factors)，多因子模型

统计学方法：主流的统计手段包括Gibbons et al. (1989) 检验（简称GRS检验）、Huberman and Kandel (1987) 的均值-方差张成（mean-variance spanning）检验以及Barillas and Shanken (2018) 的贝叶斯方法。

机器学习

1.2.2 研究现状

John Cochrane: factor zoo:

- 1) 哪些因子独立？
- 2) 哪些因子重要？
- 3) 因子驱动资产价格的原因是什么？

Harvey et al. (2016) 研究了316个因子，提出了一种能够利用不同因子之间相关性的全新分析框架，t value from 2 to 3, Chordia et al. (2020) 认为316是被发表的，被研究过的，低估多重假设检验的影响，他使用模型定量刻画学术界挖因子流程，t to 3.4; Green et al. (2017) 使用FM回归分析100个因子，发现真正独立因子少之又少；Harvey and Liu (2018) 以“幸运因子”为题提出一个基于正交化、自助法分析框架，从一揽子因子中逐一识别对解释资产预期收益率截面差异有增量贡献的因子。

新因子被发表随着使用，样本外效果下降；没有充分考虑交易费用；不合理考虑做空限制高估因子收益；

另一个关注点：多变量构建复合异象或因子

Piotroski (2000) 提出的F-Score以及Mohanram (2005) 提出的G-Score (Chapter 5)

AQR提出质量因子 (Asness et al. 2019) , Novy-Marx (2015a) 指出变量数量增多可能导致过拟合风险加大，需考虑每个变量对于提升收益率的增量贡献

行为金融：人的各种偏差也是因子成因的必要补充，投资者情绪等等；研究包括时序可预测性、截面差异性。

最后一个关注点：因子收益与宏观经济之间的关系，对因子择时也很有帮助 (Claessens and Kose (2018)) 做了综述，但是之间关系没那么强，还有大量研究要做。

1.3 因子投资的业界发展

主动管理人：围绕某个因子构建portfolio时需要考虑可投资性的约束（风险暴露、交易成本、流动性等等），理论到实践的细节。

普通投资者：每个因子背后的原因，选择适合自己风险偏好的因子，使用合适金融工具（eg：Smart Beta ETF）交易因子。

1.3.1 因子投资和管理人

1) 收益预测vs风险预测

截面获得超额；时序计算协方差矩阵进行风险管理；both，然后组合优化

2) 资金流入削弱因子收益率

资金越多——因子拥挤（factor crowding），使用相似指标排序、接近的调仓频率加剧这种影响，时常会由非有效变得更有效，降低因子的预期收益率。

还有可能造成流动性冲击，市场大量抛压，股票价格大量下降，造成大量亏损。应考虑如何计算因子拥挤度，规避因子拥挤带来的负面影响。

3) 因子择时

Bender et al. (2018) 客观地综述了使用不同预测指标在不同时间尺度下对不同风格因子择时的效果。这些预测指标包括因子估值、因子动量、投资者情绪以及各类宏观经济指标。

按因子估值（factor valuation）：USA西海岸 Research Affiliates

按因子动量（factor momentum）：USA东海岸 AQR

4) 区分alpha、beta收益

收益来源于异象还是配置了某些风格因子

主动配置风格因子（择时、风险控制）也很厉害，最顶尖的beta管理人比alpha获得更优质的收益

5) 创新价值

新数据反应增量信息：舆情、专利、新闻等；机器学习算法，预测财务指标，挖掘因子与收益率之间的非线性关系

1.3.2 因子投资和投资者

风格因子指数：起初评估主动管理人业绩，后来出现复制指数的基金，后构造一系列Smart Beta ETF

挑选同时被动追踪标普500ETFs的ETFs，只需要比较费率，而很多人被收益率迷惑，追高收益率。

针对不同ETFs，可以从因子敞口、费用、流动性等角度考虑，种类比较多很难选，需要首先明确目标，eg：分散化风险or获得相对市场的超额，其次了解风格因子背后逻辑和风险。

