

A 股多因子模型算法文档

BetaPlus 小组

This version: September 4, 2021

免责声明： BetaPlus 小组旨在针对 A 股分享独立、严谨、可复制的实证研究成果。本文中所含信息由 BetaPlus 小组基于公开信息而提供。入市有风险，投资需谨慎。在任何情况下，本文中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，BetaPlus 小组不对任何人因使用本文的任何内容所引致的任何损失负任何责任。本文中所表述的意见不代表 BetaPlus 小组成员所属机构的观点。

未经授权，严禁转载。

1 数据

- 如无特殊说明，多因子模型中因子收益率的起始日期为：1995 年 1 月 1 日。
- 股票范围：全部在市股票，包括深圳主板、中小板和创业板，上海主板和科创板。
- 剔除股票：剔除黑名单股票和不可交易股票，其中黑名单包括新股（上市不满 12 个月）、风险警示股、待退市股和净资产为负股，不可交易股票包括停牌股票和一字涨跌停股票。
- 无风险收益率：不同时间段采用不同数据，具体如表 1 所示。

表 1: 无风险收益率数据

时间区间	数据来源
2002/08/06 之前	三个月期定期银行存款利率
2002/08/07 至 2006/10/07	三个月期中央银行票据的票面利率
2006/10/08 至今	上海银行间三个月同业拆放利率

- 股票价格数据：后复权收盘价。
- 用来构建因子的变量：如表 2 所示。
- 数据更新频率：月度。

表 2: 用来构建因子的变量

因子	变量
规模	总市值，即收盘价乘以总股本
价值	账面市值比 (Book-to-Market ratio, 简称 BM)，即净资产除以总市值
动量	过去 12 个月 (剔除最近 1 个月) 累计收益率，其中一个月包含 21 个交易日
盈利	营业利润除以净资产
投资	总资产增长率或净资产增长率 ¹

¹不同版本的模型中，投资因子将使用总资产增长率或净资产增长率为变量来构建。

2 经典版本与极简版本

在针对 A 股的实证研究中，本文档提供经典版本和极简版本两个多因子模型版本。经典版本又称为学术界版本，它严格按照关于多因子模型的学术论文中描述的方法选择因子变量，构建因子投资组合，计算因子收益率。极简版本是在经典版本的基础上，针对 A 股市场特点而开发的简化版多因子模型，一边更好的适应 A 股市场。

经典版本和极简版本的主要差异包括：(1) 经典版本严格遵循学术界惯例，因此因子投资组合每年再平衡；极简版本则按照业界惯例，因此因子投资组合每月再平衡，从而更快地利用最新的因子变量取值。(2) 经典版本在构建因子投资组合时，往往采用市值和目标变量的双重排序；但在极简版本中，将使用因子变量单变量排序构建因子投资组合。

3 CAPM 模型

3.1 经典版本

- 构建市场组合：
 - * 市场组合成分为样本空间内所有股票；
 - * 加权方式为总市值加权；
 - * 年度再平衡，每年 4 月¹最后一个交易日为再平衡日。
- 市场因子，即市场组合收益率减去无风险收益率。

3.2 极简版本

- 构建市场组合：
 - * 市场组合成分为样本空间内所有股票；
 - * 加权方式为总市值加权；
 - * 月度再平衡，每月最后一个交易日为再平衡日。

¹美股版本为每年 6 月底再平衡，但国内上年年报 4 月底就已经披露完毕，因此选择在 4 月底进行再平衡。

- 市场因子，即市场组合收益率减去无风险收益率。

4 Fama–French 三因子模型

Fama–French 三因子模型出自 Fama and French (1993)，包括市场因子、规模因子 (small-minus-big，即 SMB) 以及价值因子 (high-minus-low，即 HML)。

4.1 经典版本

- 因子投资组合的换仓频率为年度，每年 4 月最后一个交易日为调仓日。
- 市场因子同 3.1 节描述的一致。
- SMB 和 HML 构造细节如下：

★ 市值采用 4 月底的总市值，BM 采用上年年末总市值和上年年报净资产计算；

★ 在组合调仓日，以主板股票总市值 (size) 中位数为断点²，将所有股票分为大市值组 (Big) 和小市值组 (Small)；相似地，以主板股票 BM 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30% (Low)、中间 40% (Middle)、后 30% (High)；

★ 对两个 size 组和三个 BM 组取交集³，形成 $2 \times 3 = 6$ 个组：Small Low (S/L)，Small Middle (S/M)，Small High (S/H)，Big Low (B/L)，Big Middle (B/M) 以及 Big High (B/H)；以上 6 个投资组合中的股票均按照总市值加权，每年再平衡。

★ 基于上述 6 个组合，定义 SMB 和 HML 因子如下：

$$\text{SMB} = \frac{1}{3}(\text{S/H} + \text{S/M} + \text{S/L}) - \frac{1}{3}(\text{B/H} + \text{B/M} + \text{B/L}) \quad (1)$$

$$\text{HML} = \frac{1}{2}(\text{S/H} + \text{B/H}) - \frac{1}{2}(\text{S/L} + \text{B/L}) \quad (2)$$

4.2 极简版本

- 因子投资组合的换仓频率为月度，每月最后一个交易日为调仓日。
- 市场因子同 3.2 节描述的一致。
- SMB 和 HML 构造细节如下：

★ 市值采用最新交易日的总市值，BM 为最近报告期净资产除以最新交易日总市值；

²Fama and French (1993) 在决定市值和 BM 的断点时，皆使用 NYSE 股票做判断。

³将不同因子分组进行结合时，有两种方案：独立双重排序（例如 Fama and French 1993）和条件双重排序（例如 Asness and Frazzini 2013）。前者为标准算法，不同指标分组之间取交集即可，绝大多数因子收益率在构造时均采用此方法；后者主要考虑股票数量总体不多时，防止个别组分到的股票太少，先按照一个指标分组，再在此基础上做分层，例如先按照 size 分为两组，再在每个 size 组里按 BM 分为 3 组。

- ★ 在调仓日，按照总市值从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组且做空第 10 组得到 SMB；
- ★ 在调仓日，按照 BM 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组且做空第 1 组得到 HML。

5 Carhart 四因子模型

Carhart (1997) 在 Fama and French (1993) 三因子模型的基础上加入了（截面）动量因子。学术界将该模型称为 Carhart 四因子模型。

5.1 经典版本

- 市场因子、SMB 以及 HML 因子的计算同 Fama–French 三因子模型经典版本算法一致，见 4.1 节。
- 动量因子计算步骤如下：
 - ★ 换仓频率为月度，每个月最后一个交易日为调仓日；
 - ★ 在调仓日，将所有股票动量因子变量（见表 2）从小到大排列，以 30% 和 70% 分位数为断点；
 - ★ 做多动量变量最强（即累计收益率最高）的 30%，做空动量变量最弱（即累计收益率最低）的 30%，等权重加权，多空收益率即为动量因子。

5.2 极简版本

- 市场因子、SMB 以及 HML 因子的计算同 Fama–French 三因子模型极简版本算法一致，见 4.2 节。
- 动量因子计算步骤如下：
 - ★ 换仓频率为月度，每个月最后一个交易日为调仓日；
 - ★ 在调仓日，按照动量变量（见表 2）从小到大将所有股票分为 10 组，每组股票数量相同且按总市值加权，做多第 10 组且做空第 1 组得到动量因子。

6 Fama–French 五因子模型

Fama–French 五因子模型出自 Fama and French (2015)，在 Fama–French 三因子模型的基础上加入了盈利因子（robust-minus-weak，即 RMW）和投资因子（conservative-minus-aggressive，即 CMA）。

6.1 经典版本

- 因子投资组合的换仓频率为年度，每年 4 月最后一个交易日为调仓日。
- 市场因子同 3.1 节一致。
- HML 因子的构造方法与 4.1 节一致。
- SMB、RMW 和 CMA 构造细节如下：
 - ★ 市值采用 4 月底的总市值；BM 采用上年年末总市值和上年年报净资产计算；盈利指标为营业利润/净资产（记为 OP），分子分母均采用上年年报数据计算；投资指标为总资产增长率（记为 INV），采用上年年报数据计算；
 - ★ 在调仓日，以主板股票市值（size）中位数为断点，将所有股票分为大市值组（Big）和小市值组（Small）；
 - ★ 在调仓日，以主板股票 BM 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30%（Low）、中间 40%（Middle）、后 30%（High）；
 - ★ 在调仓日，以主板股票 OP 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30%（Weak）、中间 40%（Neutral）、后 30%（Robust）；
 - ★ 在调仓日，以主板股票 INV 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30%（Conservative）、中间 40%（Neutral）、后 30%（Aggressive）；
 - ★ 将两个市值分组分别和三个 BM 分组、三个盈利分组、三个投资分组做交集，共得到 18 个组合；以上 18 个组合中的股票按总市值加权，每年再平衡；
 - ★ 利用上述投资组合，构建 RMW、CMA 以及 SMB 因子（在下列公式中，使用单词首字母代表单词，即 S 代表 Small、B 代表 Big、H 代表 High、L 代表 Low、R 代表 Robust、W 代表 Weak、C 代表 Conservative 以及 A 代表 Aggressive）：

$$RMW = \frac{1}{2}(S/R + B/R) - \frac{1}{2}(S/W + B/W) \quad (3)$$

$$CMA = \frac{1}{2}(S/C + B/C) - \frac{1}{2}(S/A + B/A) \quad (4)$$

$$SMB = \frac{1}{3}(SMB_{BM} + SMB_{ROE} + SMB_{INV}) \quad (5)$$

$$\text{其中 } SMB_{BM} = \frac{1}{3}(S/H + S/M + S/L) - \frac{1}{3}(B/H + B/M + B/L) \quad (6)$$

$$SMB_{ROE} = \frac{1}{3}(S/R + S/N + S/W) - \frac{1}{3}(B/R + B/N + B/W) \quad (7)$$

$$SMB_{INV} = \frac{1}{3}(S/C + S/N + S/A) - \frac{1}{3}(B/C + B/N + B/A) \quad (8)$$

6.2 极简版本

- 因子投资组合的换仓频率为月度，每月最后一个交易日为调仓日。
- 市场因子同 3.2 节描述的一致。
- SMB、HML、RMW 和 CMA 的构建细节如下：
 - ★ 市值采用最新交易日的总市值；BM 为最近报告期净资产除以最新交易日总市值；盈利指标 OP 采用 ttm 数据，即最近 12 个月营业利润净资产比/最近 12 个月平均净资产；投资指标 INV 为最近报告期净资产同比增长率⁴；
 - ★ 在调仓日，按照总市值从小到大将所有股票分为 10 组，每组股票数量相同且按总市值加权，做多第 1 组且做空第 10 组得到 SMB；
 - ★ 在调仓日，按照 BM 从小到大将所有股票分为 10 组，每组股票数量相同且按总市值加权，做多第 10 组且做空第 1 组得到 HML；
 - ★ 在调仓日，按照 OP 从小到大将所有股票分为 10 组，每组股票数量相同且按总市值加权，做多第 10 组且做空第 1 组得到 RMW；
 - ★ 在调仓日，按照 INV 从小到大将所有股票分为 10 组，每组股票数量相同且按总市值加权，做多第 1 组且做空第 10 组得到 CMA。

7 Novy-Marx 四因子模型

Novy-Marx 四因子模型出自 Novy-Marx (2013)，包含市场、价值 (HML)、盈利 (PMU) 以及动量 (UMD) 四个因子。

7.1 经典版本

- 起始日期：1999 年 12 月 31 日。
- 换仓频率：HML 和 PMU 每年 4 月底换仓，UMD 每月换仓。
- 市场因子与 Fama–French 三因子经典版本中市场因子一致。
- HML、PMU 和 UMD 构造细节如下：
 - ★ 市值采用 4 月底的总市值；年度毛利率 (GP) 为年报毛利润/总资产；价值因子变量为 $\log(\text{BM})$ 其中 BM ratio 为年报净资产/年末总市值；动量变量每月月末更新，定义为过去 12 个月收益率（剔除最近 1 个月）；
 - ★ 在调仓日，以主板股票市值中位数为断点，将所有股票分为大市值组 (Big) 和小市值组 (Small)；

⁴Fama and French (2015) 在做理论推导时，投资指标对应为净资产增长率，但总资产增长率表现更好 (Fama and French 2006)，因此他们最终采用了后者。A 股方面，实证研究表明净资产增长率效果更好，因此极简算法采用前者。

- ★ 在调仓日，以主板股票 $\log(\text{BM})$ 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30% (Low 组)、中间 40% (Neutral 组)、后 30% (High 组)；
- ★ 在调仓日，以主板股票 GP 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30% (Unprofitability 组)、中间 40% (Neutral 组)、后 30% (Profitability 组)；
- ★ 在调仓日，以主板股票动量变量的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30% (Down 组)、中间 40% (Neutral 组)、后 30% (Up 组)；
- ★ 将两个市值分组分别和三个 $\log(\text{BM})$ 分组、三个 GP 分组、三个动量分组做交集，共得到 18 个组合；
- ★ 对以上 18 个组合，按总市值加权，并进行行业中性化。中性化的方式为，对每个组合中入选的股票，按其相等的权重减去该股票所属行业的收益率，最终得到 18 个行业中性后的投资组合；
- ★ 利用上述投资组合，构建 HML、PMU 以及 UMD 因子 (PMU 中的 U 为 Unprofitability 组、UMD 中的 U 为 Up 组，请勿混淆)：

$$\text{HML} = \frac{1}{2}(\text{S}/\text{H} + \text{B}/\text{H}) - \frac{1}{2}(\text{S}/\text{L} + \text{B}/\text{L}) \quad (9)$$

$$\text{PMU} = \frac{1}{2}(\text{S}/\text{P} + \text{B}/\text{P}) - \frac{1}{2}(\text{S}/\text{U} + \text{B}/\text{U}) \quad (10)$$

$$\text{UMD} = \frac{1}{2}(\text{S}/\text{U} + \text{B}/\text{U}) - \frac{1}{2}(\text{S}/\text{D} + \text{B}/\text{D}) \quad (11)$$

7.2 极简版本

- 因子投资组合的换仓频率为月度，每月最后一个交易日为调仓日。
- 市场因子与 Fama–French 三因子极简版本中市场因子一致。
- HML、PMU 和 UMD 构造细节如下：
 - ★ 涉及变量：最近 12 个月毛利率 (GP)，即最近 12 个月毛利润/平均总资产； $\log(\text{BM})$ ，其中 BM 用最近报告期净资产和最新总市值计算；动量为每月月末更新，定义为过去 12 个月收益率 (剔除最近 1 个月)；
 - ★ 在调仓日，按照 $\log(\text{BM})$ 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组且做空第 1 组得到 HML；
 - ★ 在调仓日，按照 GP 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组且做空第 1 组得到 PMU；
 - ★ 在调仓日，按照动量从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组且做空第 1 组得到 UMD。

8 Hou-Xue-Zhang 四因子模型

Hou-Xue-Zhang 四因子模型 (也被学术界称为 q-factor model) 出自 Hou, Xue and Zhang (2015)，包含市场、规模、盈利以及投资四个因子。

8.1 经典版本

- 起始日期：2003 年 12 月 31 日。
- 换仓频率：每个月最后一个交易日进行调仓。
- 市场因子与 Fama–French 三因子经典版本中市场因子一致⁵。
- 规模、盈利和投资因子构造细节如下：

- ★ 涉及变量：市值 (ME) 采用 4 月底的总市值；总资产年度变化率 (I/A) 为年报总资产每年变化幅度，每年 4 月底更新；单季度 ROE (ROE) 为最新报告期净利润和净资产计算。依照原文，上面三个变量，I/A 和 ME 每年变化一次，ROE 是每个月更新一次；
- ★ 在调仓日，以主板股票市值 (ME) 中位数为断点，将所有股票分为大市值组 (Big) 和小市值组 (Small)；
- ★ 在调仓日，以主板股票 ROE 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30% (Low 组)、中间 40% (Middle 组)、后 30% (High 组)；
- ★ 在调仓日，以主板股票 I/A 的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30% (Low 组)、中间 40% (Middle 组)、后 30% (High 组)；
- ★ 按照两个市值分组、三个盈利分组和三个投资分组，进行 $2 \times 3 \times 3$ 独立三重排序，得到 18 个组合，这 18 个组合采用市值进行加权。令 c_1 、 c_2 、 c_3 依次代表每个投资组合在市值、ROE 和总资产变化率三个变量上的划分，其中 c_1 取值为 S 或 B 代表小、大市值， c_2 和 c_3 取值为 H、M、L 代表高、中、低三档；令数学符号 $c_1/c_2/c_3$ 代表上述三个变量某个划分交集得到的分组。依照上述数学符号，使用这 18 个投资组合，这个三因子的定义如下：

$$\begin{aligned} \text{规模} = & \frac{1}{9}(S/L/L + S/M/L + S/H/L + S/L/M + S/M/M + S/H/M \\ & + S/L/H + S/M/H + S/H/H) \\ & - \frac{1}{9}(B/L/L + B/M/L + B/H/L + B/L/M + B/M/M + B/H/M \\ & + B/L/H + B/M/H + B/H/H) \end{aligned} \quad (12)$$

$$\begin{aligned} \text{盈利} = & \frac{1}{6}(S/H/L + S/H/M + S/H/H + B/H/L + B/H/M + B/H/H) \\ & - \frac{1}{6}(S/L/L + S/L/M + S/L/H + B/L/L + B/L/M + B/L/H) \end{aligned} \quad (13)$$

$$\begin{aligned} \text{投资} = & \frac{1}{6}(S/L/L + S/M/L + S/H/L + B/L/L + B/M/L + B/H/L) \\ & - \frac{1}{6}(S/L/H + S/M/H + S/H/H + B/L/H + B/M/H + B/H/H) \end{aligned} \quad (14)$$

式 (12)、(13) 和 (14) 的文字含义如下。规模因子是等权做多 9 个小市值组合 ($S/c_2/c_3$)、同时等权做空 9 个大市值组合 ($B/c_2/c_3$)；盈利因子是等权做多 6 个

⁵需要说明的是，Hou, Xue and Zhang (2015) 剔除了金融类公司。为了便于比较，我们在构造该模型时，投资范围和 Fama–French 三因子模型保持一致。

高 ROE 组合 ($c_1/H/c_3$)，同时等权做空 6 个低 ROE 组合 ($c_1/L/c_3$)；投资因子是等权做多 6 个低总资产变化率组合 ($c_1/c_2/L$)，同时等权做空 6 个高总资产变化率组合 ($c_1/c_2/H$)。

8.2 极简版本

- 因子投资组合的换仓频率为月度，每月最后一个交易日为调仓日。
- 市场因子与 Fama–French 三因子极简版本中市场因子一致。
- 规模、盈利和投资构造细节如下：
 - ★ 涉及变量：单季度 ROE (ROE)，最新总市值 (ME) 和最新报告期总资产年度变化率 (I/A)；
 - ★ 在调仓日，按照 ME 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组 (**市值最小**) 且做空第 10 组 (**市值最大**) 得到规模因子；
 - ★ 在调仓日，按照 ROE 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组 (**盈利最强**) 且做空第 1 组 (**盈利最差**) 得到盈利因子；
 - ★ 在调仓日，按照 I/A 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组 (**I/A 最小**) 且做空第 10 组 (**I/A 最大**) 得到投资因子。

9 Stambaugh-Yuan 四因子模型

Stambaugh-Yuan 四因子模型出自 Stambaugh and Yuan (2017)，其在市场和规模的基础上加入两个错误定价因子，分别为 MGMT (管理因子) 和 PERF (表现因子)。

9.1 经典版本

- 换仓频率：每个月最后一个交易日进行调仓。
- 市场因子与 Fama–French 三因子经典版本中市场因子一致。
- 规模、MGMT 和 PERF 因子构造细节如下：
 - ★ 涉及变量：MGMT 用到股票净发行量、复合股票发行量、应计利润、经营业资产、总资产增长率和投资与总资产比；PERF 用到 O-Score、12 个月动量 (剔除最近一个月)、毛利率和 ROA⁶；
 - ★ 在调仓日，用每一个变量对股票进行排序，排序方向和股票未来收益成反比，即被高估的股票排名较高，被低估的股票排名较低；
 - ★ 对 6 个管理类变量排名取平均，得到管理综合变量 (MGMT)；对 4 个表现类变量排名取平均，得到表现综合变量 (PERF)；

⁶原模型在 PERF 因子里面还考虑了财务困境。由于指标在 A 股无法计算，因此我们不考虑该变量。

- ★ 以主板股票市值中位数为断点，将所有股票分为大市值组 (Big) 和小市值组 (Small) ；
- ★ 以所有股票管理综合变量 (MGMT) 的 20% 和 80% 分位数为断点，将所有股票分为 3 组：前 20% (L_{MGMT})、中间 60% (M_{MGMT})、后 20% (H_{MGMT} 组) ；
- ★ 以所有股票表现综合变量 (PERF) 的 20% 和 80% 分位数为断点，将所有股票分为 3 组：前 20% (L_{PERF})、中间 60% (M_{PERF})、后 20% (H_{PERF}) ；
- ★ 将两个市值分组分别和三个管理分组、三个表现分组做交集，共得到 12 个组合；对以上 12 个组合，按总市值加权，持有一个月。依照上述数学符号，使用这 12 个投资组合，MGMT、PERF 以及规模 (SMB) 三个因子的定义如下：

$$\text{MGMT} = \frac{1}{2}(\text{S}/\text{L}_{\text{MGMT}} + \text{B}/\text{L}_{\text{MGMT}}) - \frac{1}{2}(\text{S}/\text{H}_{\text{MGMT}} + \text{B}/\text{H}_{\text{MGMT}}) \quad (15)$$

$$\text{PERF} = \frac{1}{2}(\text{S}/\text{L}_{\text{PERF}} + \text{B}/\text{L}_{\text{PERF}}) - \frac{1}{2}(\text{S}/\text{H}_{\text{PERF}} + \text{B}/\text{H}_{\text{PERF}}) \quad (16)$$

$$\text{SMB} = \frac{1}{2}(\text{S}/\text{M}_{\text{MGMT}} + \text{S}/\text{M}_{\text{PERF}}) - \frac{1}{2}(\text{B}/\text{M}_{\text{MGMT}} + \text{B}/\text{M}_{\text{PERF}}) \quad (17)$$

由 (17) 可以看出 Stambaugh–Yuan 四因子模型中的规模因子的构建方法与传统方法（例如 Fama–French 三、五因子）而言差异非常大（它刻意使用了通过 MGMT 和 PERF 变量划分得到的中间组 M_{MGMT} 和 M_{PERF}）。对于这种构造方法，Stambaugh and Yuan (2017) 的解释是，传统的双重排序方法会中性化错误定价对市值的影响。然而，套利不对称性致价格被高估的股票的错误定价难以被消除。此外，大量实证结果显示，错误定价在小市值股票中更为突出。这些特点使得传统的构造方法无法在规模因子的多、空两头对称地消除错误定价的影响，造成规模因子有被高估的偏误，因此不宜采用。正因如此，Stambaugh and Yuan (2017) 才采用了 (17) 的方法构造规模因子。他们的实证表明，如此构造的规模因子比传统方法得到的规模因子在美股中有更高的风险溢价。

9.2 极简版本

- 因子投资组合的换仓频率为月度，每月最后一个交易日为调仓日。
- 市场因子与 Fama–French 三因子极简版本中市场因子一致。
- MGMT、PERF、和规模 (SMB) 因子的构造细节如下：
 - ★ 涉及变量：管理综合变量 (MGMT) 和表现综合变量 (PERF) 同经典版本；
 - ★ 在调仓日，按照 MGMT 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组且做空第 10 组得到 MGMT 因子；
 - ★ 在调仓日，按照 PERF 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组且做空第 10 组得到 PERF 因子；
 - ★ 在调仓日，按照总市值从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组且做空第 10 组得到 SMB 因子。

10 Daniel-Hirshleifer-Sun 三因子模型

Daniel-Hirshleifer-Sun 三因子模型出自 Daniel, Hirshleifer and Sun (2020), 其市场因子的基础上加入两个错误定价因子, 分别为 FIN 和 PEAD。

10.1 经典版本

- 换仓频率: FIN 因子每年四月底进行调仓; PEAD 每个月最后一个交易日进行调仓。
- 市场因子与 Fama-French 三因子经典版本中市场因子一致。
- FIN 和 PEAD 因子构造细节如下:
 - ★ 涉及变量: FIN 因子采用过去 5 年的复合股权发行量 (CEI) 变量⁷, PEAD 因子则使用的是 $t-2$ 、 $t-1$ 、 t 以及 $t+1$ (t 为最近财报公告日) 这 4 日累计超额收益率 (CAR) 变量⁸;
 - ★ 以主板股票市值中位数为断点, 将所有股票分为大市值组 (Big) 和小市值组 (Small);
 - ★ 以主板股 CEI 的 20% 和 80% 分位数为断点, 将所有股票分为 3 组: 前 20% (L_{CEI})、中间 60% (M_{CEI})、后 20% (H_{CEI});
 - ★ 以主板股 CAR 的 20% 和 80% 分位数为断点, 将所有股票分为 3 组: 前 20% (L_{CAR})、中间 60% (M_{CAR})、后 20% (H_{CAR});
 - ★ 将两个市值分组分别和三个 FIN 分组、三个 PEAD 分组做交集, 共得到 12 个组合; 对以上 12 个组合, 按总市值加权, 定期调仓。依照上述数学符号, 使用这 12 个投资组合, FIN 和 PEAD 两个因子的定义如下:

$$FIN = \frac{1}{2}(S/L_{CEI} + B/L_{CEI}) - \frac{1}{2}(S/H_{CEI} + B/H_{CEI}) \quad (18)$$

$$PEAD = \frac{1}{2}(S/H_{CAR} + B/H_{CAR}) - \frac{1}{2}(S/L_{CAR} + B/L_{CAR}) \quad (19)$$

10.2 极简版本

- 因子投资组合的换仓频率为月度, 每月最后一个交易日为调仓日。
- 市场因子与 Fama-French 三因子极简版本中市场因子一致。
- FIN 和 PEAD 因子的构造细节如下:
 - ★ 涉及变量: FIN 因子采用过去 5 年的复合股权发行量 (CEI) 变量, PEAD 因子则使用的是 $t-2$ 、 $t-1$ 、 t 以及 $t+1$ (t 为最近财报公告日) 这 4 日累计超额收益率 (CAR) 变量;

⁷原模型中, FIN 因子还考虑了过去一年股票净发行。由于 A 股回购数据不够充分, 在后续无法用来排序构造组合, 因此我们没有考虑这一变量。

⁸每日的超额收益为个股收益减去同期市场指数收益率。

- ★ 在调仓日，按照 FIN 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组且做空第 10 组得到 FIN 因子；
- ★ 在调仓日，按照 PEAD 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组且做空第 1 组得到 PEAD 因子。

11 BetaPlus A 股混合四因子模型

BetaPlus 小组在市场和规模的基础上加入两个行为金融学因子，分别为 OVER（反应过度）和 UNDER（反应不足），提出一个混合四因子模型。需要特别指出的是，这个模型是专门针对 A 股市场提出的。

11.1 经典版本

- 起始日期：1999 年 12 月 31 日。
- 换仓频率：每个月最后一个交易日进行调仓。
- 市场因子与 Fama–French 三因子经典版本中市场因子一致。
- 规模、OVER 和 UNDER 因子构造细节如下：
 - ★ 涉及变量：OVER 用到账面市值比、短期反转、特质波动率、MAX 以及异常换手率；UNDER 用到 SUE、ROA、应计量（基于资产负债表）、中期动量以及流动性冲击；
 - ★ 在调仓日，将股票在截面上根据每个变量和预期收益的相关性排序，正相关则从小到大排序，负相关则从大到小排序，排名越靠后得分越高。需特别指出的是，由于反应过度变量和收益率负相关，因此得分高（低）的股票意味着反应过度最微弱（严重）；由于反应不足和收益率正相关，因此得分高（低）的股票反应不足最严重（微弱）；
 - ★ 对 5 个 OVER 变量排名取平均，就得到反应过度综合得分；对 5 个 UNDER 类变量排名取平均，就得到反应不足综合得分；
 - ★ 以主板股票市值中位数为断点，将所有股票分为大市值组（B）和小市值组（S）；
 - ★ 以所有股票反应过度综合得分（OVER）的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30%（LOVER）、中间 40%（M_{OVER}）、后 30%（H_{OVER} 组）；
 - ★ 以所有股票反应不足综合得分（UNDER）的 30% 和 70% 分位数为断点，将所有股票分为 3 组：前 30%（L_{UNDER}）、中间 40%（M_{UNDER}）、后 30%（H_{UNDER}）；
 - ★ 将两个市值分组分别和三个反应过度分组、三个反应不足分组做交集，共得到 12 个组合；对以上 12 个组合，按总市值加权，持有一个月。依照上述数学符号，使用这 12 个投资组合，OVER、UNDER 以及规模（SMB）三个因子的定义如下：

$$\text{OVER} = \frac{1}{2}(\text{S}/\text{H}_{\text{OVER}} + \text{B}/\text{H}_{\text{OVER}}) - \frac{1}{2}(\text{S}/\text{L}_{\text{OVER}} + \text{B}/\text{L}_{\text{OVER}}) \quad (20)$$

$$\text{UNDER} = \frac{1}{2}(\text{S}/\text{H}_{\text{UNDER}} + \text{B}/\text{H}_{\text{UNDER}}) - \frac{1}{2}(\text{S}/\text{L}_{\text{UNDER}} + \text{B}/\text{L}_{\text{UNDER}}) \quad (21)$$

$$\text{SMB} = \frac{1}{2}(\text{SMB}_{\text{OVER}} + \text{SMB}_{\text{UNDER}})$$

$$\begin{aligned} \text{其中 } \text{SMB}_{\text{OVER}} &= \frac{1}{3}(\text{S}/\text{H}_{\text{OVER}} + \text{S}/\text{M}_{\text{OVER}} + \text{S}/\text{L}_{\text{OVER}}) \\ &\quad - \frac{1}{3}(\text{B}/\text{H}_{\text{OVER}} + \text{B}/\text{M}_{\text{OVER}} + \text{B}/\text{L}_{\text{OVER}}) \end{aligned} \quad (22)$$

$$\begin{aligned} \text{SMB}_{\text{UNDER}} &= \frac{1}{3}(\text{S}/\text{H}_{\text{UNDER}} + \text{S}/\text{M}_{\text{UNDER}} + \text{S}/\text{L}_{\text{UNDER}}) \\ &\quad - \frac{1}{3}(\text{B}/\text{H}_{\text{UNDER}} + \text{B}/\text{M}_{\text{UNDER}} + \text{B}/\text{L}_{\text{UNDER}}) \end{aligned}$$

11.2 极简版本

- 因子投资组合的换仓频率为月度，每月最后一个交易日为调仓日。
- 市场因子与 Fama–French 三因子极简版本中市场因子一致。
- OVER、UNDER、和规模（SMB）因子的构造细节如下：
 - ★ 涉及变量：反应过度综合得分（OVER）和反应不足综合得分（UNDER）同经典版本；
 - ★ 在调仓日，按照 OVER 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组（**得分最高**）且做空第 1 组（**得分最低**）得到 OVER 因子；
 - ★ 在调仓日，按照 UNDER 从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 10 组（**得分最高**）且做空第 1 组（**得分最低**）得到 UNDER 因子；
 - ★ 在调仓日，按照总市值从小到大将所有股票分为 10 组，每组股票数量相同且市值加权，做多第 1 组（**市值最低**）且做空第 10 组（**市值最高**）得到 SMB 因子。

参考文献

- Asness, C. S. and A. Frazzini (2013). The devil in HML’s details. *The Journal of Portfolio Management* 39(4), 49–68.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance* 52(1), 57–82.
- Daniel, K. D., D. A. Hirshleifer, and L. Sun (2020). Short- and long-horizon behavioral factors. *Review of Financial Studies* 33(4), 1673–1736.
- Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.

- Fama, E. F. and K. R. French (2006). Profitability, investment and average returns. *Journal of Financial Economics* 82(3), 491–518.
- Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics* 116(1), 1–22.
- Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies* 28(3), 650–705.
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics* 108(1), 1–28.
- Stambaugh, R. F. and Y. Yuan (2017). Mispricing factors. *Review of Financial Studies* 30(4), 1270–1315.