

基于持仓数据的股票型产品风格与收益归因体系构建

一、引言

股票型产品的风格归因体系相对成熟，主要分为两大类：基于净值的归因体系（Returns-based approach-RBA）与基于持仓的归因体系（Holdings-based approach-HBA）。

RBA 被称为“外部人”归因法，核心是基于时间序列的多元线性回归，已知产品净值序列，考察其收益率相对于一系列风格指数收益率的表现，代表体系为 Fama-French 5 因子模型；而 HBA 被称为“内部人”归因法，核心是基于横截面的多元线性回归，已知产品在不同时点的具体持仓，考察其相对于特定风格因子组合的暴露程度，代表体系为 Barra 风险模型。

表 1：归因体系对比

	基于净值的归因体系(RBA)	基于持仓的归因体系(HBA)
基本原理	产品收益率序列相对于风格指数收益率序列的表现	产品具体持仓在不同风格因子组合的暴露程度
方法论	时间序列回归	横截面回归
算法流程	已知因子收益,求解因子暴露	已知因子暴露,求解因子收益
前提假设	因子暴露在一段时间内保持不变	具体持仓在一段时间内保持不变
优点	净值数据易于获取、模型解释度较高	持仓信息透明度高
缺点	模型缺陷(多重共线性问题等)	持仓信息较难获取、公布频率较低(切片化)、模型解释度较低
适用范围	外部人(内部人亦可)	内部人(外部人存在较高门槛)

本篇报告参考 Barra 风险模型框架，同时结合国内股票市场的特点，主要聚焦于基于持仓的股票型产品归因体系构建。

二、纯因子组合简介

Barra 框架通过多因子模型将个股收益和风险分解为共同因子部分和特异部分。

$$r = Xf + u$$

$$\Sigma = XFX^T + \Delta$$

其中 r 为股票组合的收益率向量， X 为因子暴露矩阵， f 为模型待估计的因子收益率向量， u 为误差向量（表示多因子模型无法解释的个股特异收益率）， Σ 为股票组合的协方差矩阵， F 为因子收益率的协方差矩阵， Δ 为个股的特异方差矩阵。

上述模型是通过横截面回归的方法对因子收益率 f 进行估计。但实证结果表明，随着股票市值的增加，其特异风险会逐步降低，也即股票的市值平方根与其特异收益的方差成反比，因此采用加权最小二乘法对因子收益率进行估计（权重为市值平方根）。由此可得到因子收益率（纯因子组合的收益率） f 的解析表达式为：

$$f = (X^T \Omega X)^{-1} X^T \Omega r$$

其中 Ω 为回归权重， $(X^T \Omega X)^{-1} X^T \Omega$ 的各行即为股票权重向量，其中第 i 行的权重向量使得组合仅对第 i 个因子有 1 单位暴露，而对其他因子 0 暴露，此即为纯因子组合的意义所在。最后再将该组合收益率序列指数化，便得到了特定的风格因子表现。

三、多因子模型的构建流程

1、描述性因子选择与预处理

参考 Barra 风险模型框架的最新版本（USE4 与 CNE5），我们选取了 10 类风格因子（factor），同时结合国内股票市场的特点，在每类因子内部确定了各自的描述性因子（descriptor）。具体分类详见附录。

随后对每一个描述性因子进行去极值（中位数去极值法）、中性化（以原始因子为因变量，以规模因子及行业哑变量为自变量进行线性回归，取回归残差替代原始因子值）与标准化（Z-score，因子值减去其均值，再除以其标准差；其中均值为简单算术平均、标准差为基于简单算术平均计算得到）。

2、风格因子合成

相同风格下的描述性因子之间具有相似的经济学逻辑，在统计上可能存在较强的相关性。因此首先将相同类别的描述性因子合成为大类的风格因子，再进行多元线性回归，从而尽可能地降低多重共线性问题，提高模型的稳定性。

描述性因子的合成过程其实就是在各因子之间分配权重的过程，不同的加权方法可能产生表现各异的大类风格因子。为降低过拟合问题，我们仅采用等权加权的方法。

最后对合成后的风格因子进行基于流通市值加权的标准化处理：

$$x_{ns} = \frac{x_{ns}^{raw} - \mu_s^{cap}}{\sigma_s}$$

其中， x_{ns}^{raw} 为原始风格因子暴露， μ_s^{cap} 为按流通市值加权的平均值， σ_s 为简单平均标准差， x_{ns} 为标准化处理后的因子暴露。

3、多因子模型构建

基于合成后的风格因子以及行业哑变量（申万一级行业分类），便可构建多元线性回归模型以得到纯因子组合收益率序列。

Barra 的 USE4 版本下的多元线性回归模型的具体形式如下：

$$r_n = f_c + \sum_i X_{ni} f_i + \sum_s X_{ns} f_s + u_n$$

其中， f_c 为市场因子收益率， f_i 为行业因子收益率， f_s 为风格因子收益率，均为待求解的变量； X_{ni} 为取值为 0 或 1 的行业哑变量， X_{ns} 为上述标准化后的风格因子暴露， u_n 为残差项，表示股票的特异收益率。

该版本中新增加了回归截距项（称为市场因子），从而与行业因子哑变量之间产生了完全共线性问题（ $\sum_i X_{ni} = 1$ ）。为解决该问题，需加入对行业因子的限制条件（ $\sum_i w_i f_i = 0$ ，其中 w_i 为行业 i 的股票流通市值加权）。此时模型的求解用到了加权最小二乘法。

据此可以得到各待求解因子的含义如下：

(1) 市场因子（截距项）近似等于全样本股票流通市值加权组合的收益率，对其他因子 0 暴露，为市场基准组合；

(2)行业因子为资金中性的多空组合（超额收益率：100%自身行业多头 + 100%市场基准组合空头），对市场因子及各风格因子 0 暴露；

(3)风格因子为资金中性的多空组合，对市场因子、各行业因子（在各行业中也是资金中性的多空组合）及其他风格因子 0 暴露，对自身风格因子偏离基准组合 1 个标准差。

因此，基于上述模型，在全部 A 股基础上剔除 ST 与停牌个股，以周度频率进行横截面回归，便可得到各风格因子的历史净值表现。

图 1：各风格因子表现

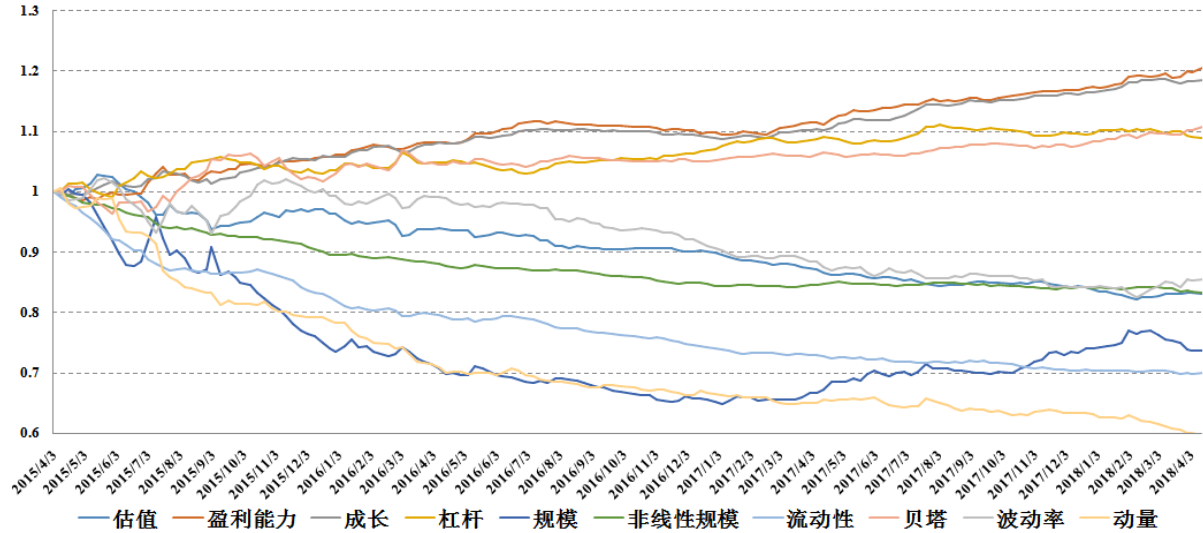


图 2：风格因子收益率相关性矩阵

	估值	盈利能力	成长	杠杆	规模	非线性规模	流动性	贝塔	波动率	动量
估值		-0.314	0.225	-0.538	-0.603	0.047	0.258	-0.637	0.802	0.057
盈利能力	-0.314		0.662	-0.443	0.340	-0.082	-0.067	-0.015	-0.166	-0.071
成长	0.225	0.662		-0.532	-0.077	0.123	0.069	-0.358	0.375	-0.037
杠杆	-0.538	-0.443	-0.532		0.086	0.078	-0.181	0.658	-0.386	-0.166
规模	-0.603	0.340	-0.077	0.086		0.154	-0.090	0.256	-0.668	0.384
非线性规模	0.047	-0.082	0.123	0.078	0.154		0.329	-0.030	0.127	0.294
流动性	0.258	-0.067	0.069	-0.181	-0.090	0.329		0.065	0.363	0.073
贝塔	-0.637	-0.015	-0.358	0.658	0.256	-0.030	0.065		-0.331	-0.233
波动率	0.802	-0.166	0.375	-0.386	-0.668	0.127	0.363	-0.331		-0.196
动量	0.057	-0.071	-0.037	-0.166	0.384	0.294	0.073	-0.233	-0.196	

四、基于持仓的归因模型实证

针对嘉实研究阿尔法（000082.OF）基金 2017 年中报与年报的全部持仓数据进行归因实证。

表 2：基金基本信息

成立日期	2013年5月28日
基金公司	嘉实基金
基金经理	张露
基金规模	4.03亿元
投资类型	股票型基金
投资风格	大盘价值风格型
比较基准	MSCI中国A股指数收益率*95%+ 银行活期存款利率(税后)*5%

图 3：基金历史表现



基于该基金 2017 年中报的全部持仓数据，其在动量因子、流动性因子、盈利因子、非线性市值因子上有显著的正向暴露，在杠杆因子有显著的负向暴露，而在其余因子上无明显暴露。同时在银行、电子、非银金融、房地产和食品饮料行业有较高的暴露。

因此，该基金的选股逻辑更偏向于选择大中市值、公司质量优良兼具成长性、且前期积累一定涨幅的股票。而这一逻辑刚好与整个 2017 年的市场风格相匹配，从而也解释了其在 2017 年相对指数表现更为优异的原因。

图 4：风格因子暴露（2017 年中报）

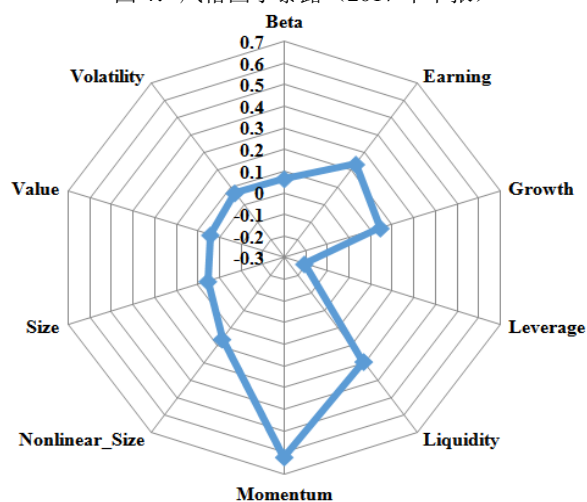


图 5：行业因子暴露（2017 年中报）

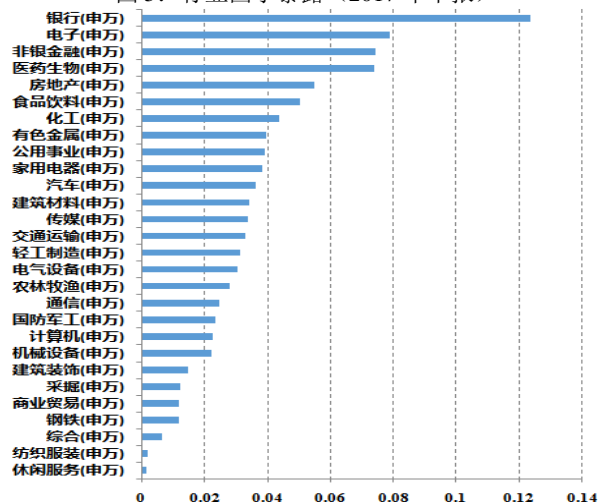


图 6：基金收益分解

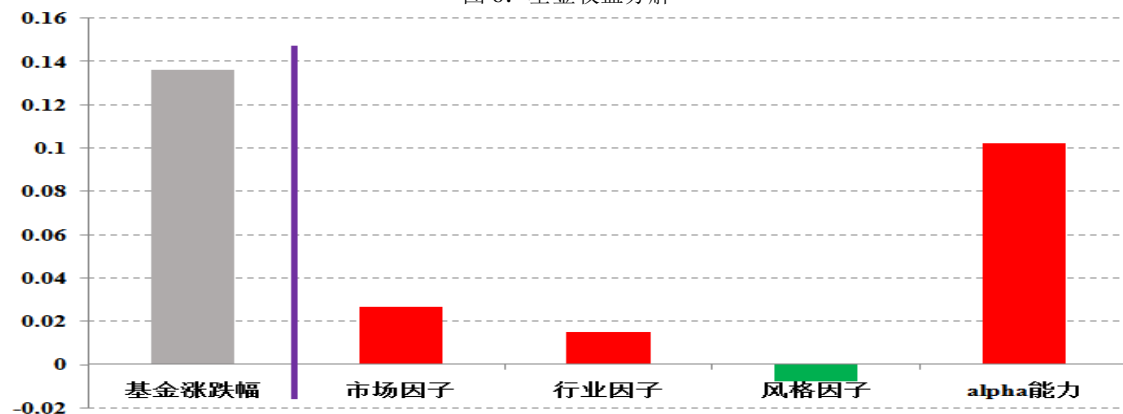
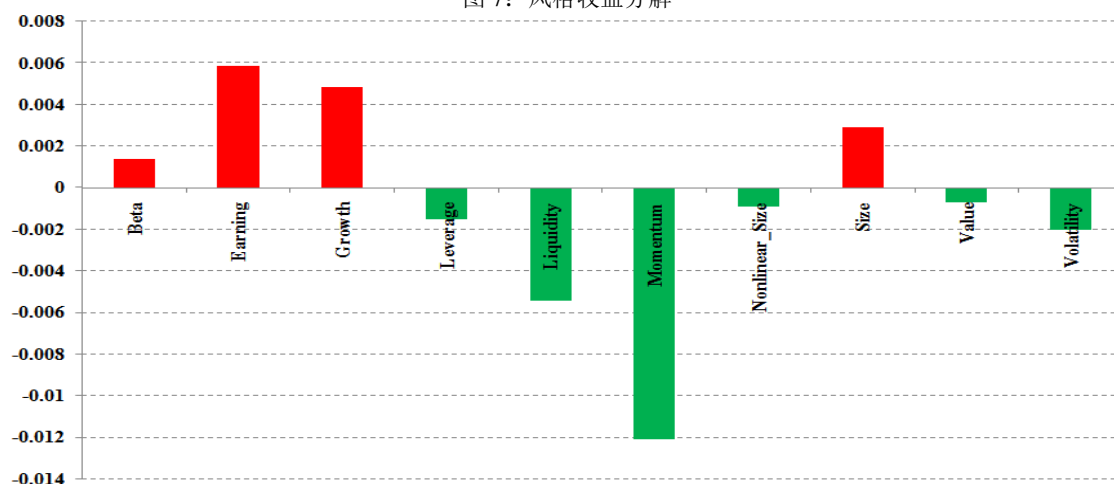


图 7：风格收益分解



五、说明与改进

基于持仓的归因模型主要为横截面回归，其最大的缺点便是模型的解释力度偏低，平均大概在 10%-40% 之间。同时由于纯因子的多空特性，即使增加更多的风格因子，也很难让解释力度有一个质的提升（特别是考虑 adjR^2 后）。

因此可以考虑对正统的归因模型进行适当修正，以增强其解释力度：

思路一：基于净值的时间序列归因模型，此类方法的解释力度一般都在 50% 及以上。但最大问题是时间窗口期的选择以及多重共线性问题。

思路二：基于持仓的 Brinson 风格模型，也即将原始 Brinson 模型中的板块/行业分类，替换为风格分类。但最大问题是风格指数成分股之间可能存在交集，不再具备互斥性。解决办法是可以考虑将某成分股按不同风格进行拆分，如某股票的 30% 属于价值、70% 属于市值等。

思路三：仍基于持仓的归因模型，但风格因子不再是纯因子组合，而是某种风格下的纯多头组合。但最大问题也是风格指数成分股之间可能存在交集，不再具备互斥性，需要对个股进行风格层面的拆分。

关于风格因子拓展的思考：

最近集中调研了几家做市场中性策略为主的公募专户及私募投顾，根据换手率高低可以大概分为低换手率策略（偏基本面因子为主，大概年换手为 2-3 倍，在 Barra 模型的风格因子上存在明显暴露，如估值、盈利及成长等）和高换手率策略（偏价量等市场高频因子为主，大概年换手为 100-240 倍，在 Barra 模型的风格因子上无明显暴露，无法解释部分占比高达 90% 及以上）。

基于上述实践，可以将归因模型进行有针对性地分别构建：

对于偏基本面因子为主的策略，可在传统的 Barra 模型基础上，深挖财报相关的因子并进行合成（即本文尝试的方向）；

而对于偏技术面因子为主的策略，传统 Barra 模型的效果将大打折扣，需要开发更偏高频的风格因子，如 WorldQuant 开发的 101 个 alpha 因子的角度（《101 Formulaic Alphas》）等。

附录

风格因子	描述性因子
估值类	市净率 PB(LF)
	市销率 PS (TTM)
	市盈率 PE (TTM)
	股息率(TTM)
盈利类	净资产收益率（平均）
	总资产收益率
	销售净利率
成长类	归属母公司股东的净利润同比增长率
	单季度_营业收入同比增长率
	营业收入同比增长率
	净资产收益率（摊薄）同比增长率
	单季度_归属母公司股东的净利润同比增长率
杠杆类	资产负债率
	市场杠杆
规模	log(流通市值)
非线性规模	log(流通市值)^3
流动类	近一个月交易日平均换手率的自然对数
	近三个月交易日平均换手率的自然对数
	近六个月交易日平均换手率的自然对数
	近十二个月交易日平均换手率的自然对数
Beta	beta（252 日）
波动率	残差波动率
	波动率
动量	return_1m
	return_3m
	return_6m
	return_12m
	均线偏离度_1m
	均线偏离度_3m
	均线偏离度_6m
	均线偏离度_12m