

主成分分析法下的股票横截面定价因子模型

——“学海拾珠”系列之七十六

报告日期：2022-01-17

分析师：严佳炜

执业证书号：S0010520070001

邮箱：yanjw@hazq.com

联系人：钱静闲

执业证书号：S0010120080059

邮箱：wuzj@hazq.com

主要观点：

本篇是“学海拾珠”系列第七十六篇，本期推荐的海外文献开发了一种运用主成分分析法（PCA）从股票收益中提取公共定价因子的方法，使用该方法，作者得到了三个定价因子，它们相比当前使用的主流因子模型在横截面收益解释程度方面表现更好。回到 A 股市场，基于主成分分析法的因子模型有助于更深入了解股票特征因子在横截面中定价中的联动效应，将具有超额收益的因子分离出来。

● 构建一种在股票收益中分离公共定价因子的方法

模型分为两步构建：第一步，对多个可以预测未来股票收益的特征变量（如公司规模，账面市值比，动量等）进行 Fama MacBeth(1973)回归，使用回归中的估计系数预测每只股票未来一个月的收益并根据预测的收益对股票排序并构造投资组合；第二步，使用主成分分析从这些组合收益中提取公共因子，作者提取了前三大主成分，并命名为“水平、斜度、曲线”因子模型（level, slope, and curve factor model）。

此命名的来源是，按预期收益排序的投资组合在第一个因子上的因子载荷是一条接近水平的线（“level”），第一个因子类似于一般市场投资组合，与主流模型中市场指数的相关系数为 0.95；第二个因子的因子载荷是一条单调的倾斜的线（“slope”），金融文献中已经确定的大多数因子都属于此；第三个因子的因子载荷是一条中间拱起的曲线（“curve”），这表明极端股票价格往往会同向变动，如果“curve”因子有正的收益收益，那么预期收益非常高和非常低的股票都会有相对较低的未来收益，而预期收益中等的股票则会有相对较高的未来收益。

● 水平、斜度、曲线模型与主流模型相比有更好的定价效果

作者发现水平、斜度、曲线模型相比 CAPM 模型、Fama 和 French 三因子模型、五因子模型以及 Hou 等人的 q 因子模型，在横截面定价时，未捕捉到的平均 alpha 绝对值更低，并能解释更多的平均收益变化。同时，在 Barillals 和 Shanken 的夏普比率测试中，该模型相比这些模型也取得了至少一样好或更好的效果。而且在样本外的测试中，水平、斜度、曲线模型显示出了很好的稳健性。

● 风险提示

本文结论基于历史数据与海外文献进行总结；不构成任何投资建议。

相关报告

1. 《基金组合如何配置权重：能力平价模型——“学海拾珠”系列之六十八》
2. 《持仓技术相似性与共同基金业绩——“学海拾珠”系列之六十九》
3. 《双重调整法下的基金业绩评价——“学海拾珠”系列之七十》
4. 《企业员工流动对股票收益的影响——“学海拾珠”系列之七十一》
5. 《贝叶斯动态面板模型下的基金业绩持续性——“学海拾珠”系列之七十二》
6. 《高点锚定效应和跨公司收益预测——“学海拾珠”系列之七十三》
7. 《债券基金交易风格与市场流动性风险——“学海拾珠”系列之七十四》
8. 《盈余公告前后的收益特征是否与投机性股票需求有关？——“学海拾珠”系列之七十五》

正文目录

1 简介	4
2 框架	5
3 排序法构造投资组合	6
4 “水平、斜度和曲线”因子模型（LSC）	8
4.1 因子数量的确定	10
4.2 与主流因子的相关性情况	10
5 资产定价检验	11
6 与主流模型的比较	12
6.1 LSC 模型与 FAMA 和 FRENCH 三因子模型的比较	12
6.2 LSC 模型与主流模型的比较	16
7 结论	17
风险提示:	17

图表目录

图表 1 25 个投资组合和前三大主成分	7
图表 2 使用预期收益的排序形成的 25 个投资组合的平均超额收益和特征	8
图表 3 对 25 个投资组合的主成分分析结果	9
图表 4 不同分组数量下前 5 种主成分的相关性展示	10
图表 5 前五个主成分与其他因子的相关性。	11
图表 6 对提取的主成分进行 25 个投资组合的时间序列回归。	11
图表 7 119 个测试投资组合两种模型的拟合比较	13
图表 8 10 个异象投资组合下模型定价能力对比	13
图表 9 模型关于 25 个市值和账面市值比投资组合的定价比较	14
图表 10 模型关于 10 个动量投资组合的定价比较	14
图表 11 模型关于 49 个行业分类投资组合的定价比较	15
图表 12 模型关于使用 5 种债券收益的投资组合的定价比较	15
图表 13 模型关于 10 个运营盈利能力投资组合的定价比较	16
图表 14 将水平、斜度和曲线与具有 119 个测试组合的领先因子模型进行比较	16

1 简介

关于资产定价因子方面研究的文献数量近年来爆炸式增长，其中一些因子是定价因子，资产对它们的暴露通常以更高的预期收益来补偿，而有些因子则不是定价因子。非定价因子代表股票收益之间的共同变动，但不会带来收益溢价。作者构造了一种寻找股票收益中共有定价因子的方法。第一步，使用几个股票收益预测因子按股票预期收益对股票进行排序构造投资组合；第二步，利用主成分法分离共有因子，即通过在已按预期收益排序的投资组合中提取共有因子的方法来剥离股票横截面定价中的联动效应。

为了区分定价和非定价的公共因子，需要使用多个不同的预测变量对预期收益进行排序。例如，在第一步中如果仅按 BP（市净率倒数）对股票进行排序，虽然可以产生对预期收益的排序，却不太容易分离出定价因子，因为按 BP 排序的投资组合中既有定价因子，也有非定价的共同变动因子（Daniel 和 Titman, 1997 年）。为了克服这个问题，作者使用多个预测组合收益的因子进行多变量回归估计后，按照估计的预期收益进行排序。该过程可以将具有非零收益溢价的因子分离出来，并把定价因子与其他非定价因子区分开来。

使用这种方法，作者发现从这些投资组合中可以提取三个共有因子，并称之为“水平、斜度和曲线”因子模型（level, slope, and curve factor model）。此命名的来源是，按预期收益排序的投资组合在第一个因子上的因子载荷是一条接近水平的线（“level”），第一个因子类似于一般市场投资组合，与主流模型中市场指数的相关系数为 0.95；第二个因子的因子载荷是一条单调的倾斜的线（“slope”），金融文献中已经确定的大多数因子都属于此；第三个因子的因子载荷是一条中间拱起的曲线（“curve”），这表明极端股票价格往往会同向变动，如果“curve”因子有正的收益收益，那么预期收益非常高和非常低的股票都会有相对较低的未来收益，而预期收益中等的股票则会有相对较高的未来收益。

作者发现，水平、斜度和曲线因子模型的表现与其他几个主流的资产定价模型相比有相似甚至更好的效果。当针对不同（超额收益）排序的投资组合进行回归解释时，水平、斜度和曲线模型回归得到的平均绝对 alpha 更低，证明捕捉了大部分的组合超额收益，且该模型在横截面上解释收益的力度比 CAPM 模型、Fama 和 French (1993) 的三因子模型（添加动量，四因子模型）、Fama 和 French (2015) 五因子模型，以及 Hou 等人的 q 因子模型 (2015) 更加充分。当参考 Barillas 和 Shanken (2018) 构造投资组合时，本文模型产生了更高的夏普比，尤其是在样本外。

这种提取定价因子的方法对有关股票横截面收益方面的文献做出了重要贡献。

首先，Cochrane (2011) 呼吁重组股票收益的因子框架，需要确定哪些因子是最重要的，哪些因子应该由更深入的宏观经济模型来解释。本文使用的方法可以找出在横截面中定价的最重要且显著的风险因子，即由水平、斜度和曲线因子模型概括。

第二，本文在概念上将预测未来股票收益和创建因子模型这两件事分开。一般情况下，因子的创建是找到一个收益预测指标来形成单变量交易策略，或几个收益预测指标共同构成多变量交易策略。本文给出了另一种将收益预测指标转换为定价因子的方法，每一个指标的权重类似于它预测能力的边际贡献。而主流因子模型的改进往往受理论驱动，在实操上可能备受争议。本文弥补了由经验产生的因子模型和理论产生的因子模型之间的鸿沟，提出了一个明确的构造规则。

第三，本文构造了一个不以公司特征为核心的因子框架。虽然基于股票特征的

排序方法已有许多研究，但是，股票价格的同向变动可能不是主要由股票特征的趋同引起的。在本文中，公司特征只是预测收益的有用信号，在这种框架下，并不关心估值究竟用市净率（PB）、市息率（PD）、市销率（PS）描述更好，也不关心盈利究竟用 ROE、毛利率还是现金盈利能力描述更好，这些公司特征仅在它们改变预期收益模型的整体预测能力时才有影响。

2 框架

作者从线性因子模型产生的收益开始讨论。证券 i 的收益 R_i 是其预期收益 E_i 与其因子载荷乘以不相关的定价和非定价因子及一个误差项 ϵ_i 的和。作者将定价因子 F 和非定价因子 G 分开，非定价因子只有 0 的收益溢价。

$$R_i = E_i + \beta_{1,i}F_1 + \beta_{2,i}F_2 + \cdots + \beta_{N,i}F_N + \phi_{1,i}G_1 + \phi_{2,i}G_2 + \cdots + \phi_{N,i}G_M + \epsilon_i \quad (1)$$

如果套利定价定理成立，那么证券的预期收益是由各因子的载荷及其相关溢价产生的。

$$E[R_i] = E_i = \lambda_1\beta_{1,i} + \lambda_2\beta_{2,i} + \cdots + \lambda_N\beta_{N,i} \quad (2)$$

由于非定价因子的相关溢价为 0，只有定价因子的收益溢价 λ 和这些因子上的载荷 β 才能决定其预期收益。如果一个特征能预测收益，那么这个特征一定能预测某个定价因子的载荷。

以此为动机，为了加强股票间定价因子的共同联动并分散非定价因子的联动，作者按预期收益对股票进行排序组合。这个排序加强了与预期收益相关的共同因子投资组合的联动所产生的影响，并削弱了零风险溢价的共同因子所产生的影响。如果单独使用一个预测预期收益的因子，例如账面市值比，就很难区分定价因子联动和非定价因子联动。任何一个单一的公司特征都可以用来预测定价因子和非定价因子的因子载荷，但是，结合许多不同的收益预测特征后，加强定价因子的效果，分散非定价因子的效果，就可以创造出按预期收益风险排序的投资组合。通过从这组基础资产中提取因子，就可以创建一个新的线性因子模型。

这种方法回应了因子模型文献中出现的一种矛盾。早期的线性因子模型的概念认为，非定价的因子可能非常多，任何两只由共同的意外现金流新闻连接起来的股票，即使对预期收益没有影响，也可能是由一个共同的因子连接起来的，而这些因子可能与定价无关。

当代的因子模型方法在很大程度上依赖于这个想法。当 Fama 和 French (1993, 1996) 在公司规模和账面市值这两个被认为可以预测不同股票预期收益的特征上分出双变量特征组合时，他们也是遵循这样的逻辑：定价因子可以从整个因子结构中分离出来，而非定价因子可以被分散掉。但是，正如 Daniel 等人 (2020) 所提出的，非定价因子的可忽略性取决于这样的假设：定价因子的载荷与非定价因子的载荷不相关。如果载荷是相关的，那么非定价因子就应该出现在资产定价模型中，而不能被视为一个可忽略的特异性因素，因为对非定价因子的暴露必须从均值-方差有效组合中对冲，越来越多的文献表明，单个非零溢价因子可能会暴露于非定价的因子。

本文提出了一种方法，可以旋转因子，使得只有一个因子有非零的风险溢价，而其他所有因子都是非定价的。

$$R_i = E_i + \beta_i F_{mve} + \phi_{1,i} \tilde{G}_1 + \cdots + \phi_{N,i} \tilde{G}_{N+M-1} + \epsilon_i \quad (3)$$

这样得到的唯一的因子是一个均值-方差有效投资组合，包含了各种可投资资产。若这个表达式只有单因子，那么只有股票的第一个因子的 beta 和其对应的因子风险溢价会出现在预期收益的方程中。

$$E[X_i] = E_i = \lambda_{mve}\beta_i \quad (4)$$

如公式(4)所示, 套利定价模型真正感兴趣的特征是预期收益, 也就是定价因子的载荷(β)。通过对预期收益进行排序形成分散投资组合, 与预期收益不相关的非定价因子上的股票载荷(ϕ)被清洗掉了。剩下的因子结构是由暴露于均值-方差有效因子以及任何载荷与均值方差有效因子相关的因子决定的。这个最小化的因子模型中只包含资产定价的必要因子。

为了与传统的方法进行比较, 可以如下考虑, 如果 Fama 和 French(1993)已经能够观察到真正的股票预期收益, 而不是粗略的近似值、市值和账面价值。如果现在可观察到的预期收益与市值、账面价值和动量等公司特征一样频繁变化, 那么仍然只有很短的时间跨度来估计复杂的股票间收益协方差。预期收益的存在并没有使得估计均值-方差模型所需的股票间收益协方差矩阵变得容易。在基础资产中选择构造投资组合的资产时, 寻找一个因子模型, 似乎仍然是最好的方向。而此时构造投资组合选股的最佳特征当然是股票的预期收益。

当然, 预期收益是不可观察的。本文认为, 潜在的预期收益应该是被建模为公司特征的函数。公司规模和账面价值以及任何其他相关的特征都是重要的, 因为它们可以预测预期收益。具有类似特征的公司可能同样会受到非定价因子的影响。在传统的方法中, 基础资产是由一个或少数几个特征构成的, 即使这些特征是预期收益的可靠预测因素, 这些非定价因子载荷的存在也可能混淆所产生的因子。这种观点的改变并没有消除识别不可观察的预期收益的可靠预测因子的艰巨任务。但按照 Fama 和 French(1993)的逻辑, 虽然目的不同, 本文仍然可以提供使用各种预期收益预测变量来考虑因子模型的方法。

3 排序法构造投资组合

本文的样本使用美股从 1964 年 7 月到 2015 年 12 月的数据。变量的定义参考 Fama 和 French (2008), 但有两个例外。针对 Novy-Marx (2013) 和 Ball 等人 (2015) 的观点, Fama 和 French 认为营业利润率是一个比账面资产收益率更稳健的截面平均收益预测指标, 而 Aharoni 等人表明, 公司层面的资产增长是一个比每股资产增长更好、更具理论意义的预测指标。因此, 作者参考这些文献略微调整了 Fama 和 French (2008) 的回归。

为了按预期收益率对股票进行分类, 作者用多个公司的特征作为预测变量形成投资组合。Fama 和 French(2006)提供了一个三步法来实现, 首先, 运行 Fama 和 MacBeth(1973)横截面回归。第二, 使用回归中的系数估计值来预测每只股票未来一个月的收益。第三, 根据预测的收益率, 将股票排序并构造投资组合。

该方法的目标是产生一个投资组合排序, 该投资组合仅使用投资者在形成投资组合时横截面上可得的信息, 使得排序中各投资组合平均收益差异最大。确定投资者的信息集时, Fama 和 French 使用全样本的参数估计值, 以便将股票分为不同投资组合。使用这种方法的理由是, 整个时间序列的全样本信息可以最好得反映出每个异象特征对预期收益的贡献。然而, Haugen 和 Baker(1996)和 Lewellen(2015)只使用过去数据的回归来形成分类。作者在后续的主要测试中同时使用全样本和样本外的数据集来展示结果。

每个横截面回归的形式如下:

$$XRet_{i,t+1} = \beta_0 + \beta_1 LogSize_{i,t} + \beta_2 Log\left(\frac{B}{M_{i,t}}\right) + \beta_3 Mom_{i,t} + \beta_4 zeroNS_{i,t} +$$

$$\beta_5 NS_{i,t} + \beta_6 negACC_{i,t} + \beta_7 posACC_{i,t} + \frac{\beta_8 dA}{A_{i,t}} + \beta_9 posOP_{i,t} + \beta_{10} negOP_{i,t} + e_{i,t+1}$$

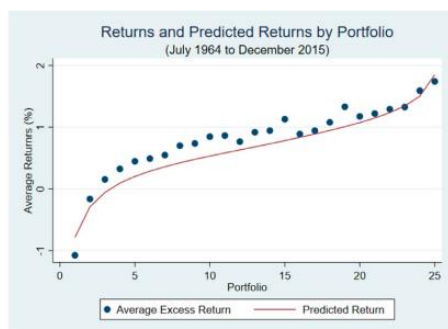
(5)

横截面回归的被解释变量是每只股票在下个月超过无风险利率 (XRet) 的超额收益率, 解释变量包括对数公司规模 (LogSize)、对数账面价值 (LogB/M)、动量 (Mom)、如果没有发行股票则指标变量等于 1 的虚拟变量 (zeroNS)、股票发行数 (NS)、负应计费用 (negACC)、正应计费用 (posACC)、资产增长 (dA/A)、正营业利润率 (posOP), 以及如果营业利润率为负值则等于 1 的虚拟变量 (negOP)。Fama 和 French (2008) 发现, 不同公司规模 (微型、小型和大型) 对特征预测变量的暴露程度不同, 因此作者对每个公司规模组的回归进行了单独估计, 允许参数估计值在这些组中有所不同。由于动量特征每个月都会更新, 因此作者对每月的数据都进行回归, 形成一组投资组合。

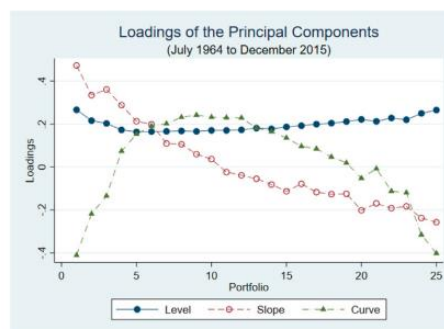
由此产生的 25 个排序的投资组合, 将其称为异象投资组合 (dissecting anomaly portfolios)。这些投资组合在排序后, 预期收益差异非常明显。图 1 的 Panel A 显示了每个组合的排序结果。由线表示的预测收益是根据每只股票的回归拟合值生成的, 由点表示的平均超额收益是每个投资组合的平均市值加权收益。

图 2 显示了 25 个排序的投资组合的汇总统计。每个投资组合的特征是投资组合中每只股票的市值加权平均数 (使用月初时的股票数据) 形成的。这种排序方法下, 各个投资组合在平均超额收益 (XRet) 方面产生了很大的差异, 在预测收益 (\widehat{XRet}) 上的各个组合也呈现出类似的差异。所有的多元回归特征除了公司规模 (JME) 之外, 其他指标在 25 个投资组合中都显示出与预期收益类似的单调增长或下降的形式, 其方向与以往研究预测的一致。

图 1 25 个投资组合和前三大主成分



(a) 25 Dissecting Anomaly Portfolios



(b) First Three Principal Components

资料来源: 华安证券研究所整理

图表 2 使用预期收益的排序形成的 25 个投资组合的平均超额收益和特征

Port	XRet	$\bar{X}Ret$	JME	B/M	Mom	dA/A	A/BE	NS	OP	XRet-EW
1	-1.08	-0.79	11,040	0.46	-0.17	0.67	0.09	0.33	-0.05	-0.80
2	-0.16	-0.29	16,861	0.46	-0.09	0.43	0.04	0.19	0.14	-0.01
3	0.15	-0.06	20,801	0.47	-0.05	0.30	0.04	0.10	0.22	0.46
4	0.32	0.09	25,562	0.48	-0.01	0.21	0.02	0.05	0.28	0.52
5	0.45	0.20	26,359	0.52	0.04	0.17	0.02	0.03	0.29	0.46
6	0.49	0.29	25,677	0.55	0.08	0.14	0.02	0.02	0.31	0.70
7	0.55	0.36	23,413	0.58	0.13	0.12	0.01	0.01	0.32	0.69
8	0.70	0.42	21,831	0.60	0.17	0.11	0.01	0.01	0.33	0.79
9	0.74	0.48	19,314	0.63	0.21	0.10	0.01	0.01	0.33	0.85
10	0.85	0.53	17,850	0.65	0.25	0.09	0.01	0.01	0.33	0.86
11	0.86	0.58	15,574	0.67	0.29	0.09	0.01	0.00	0.34	1.03
12	0.77	0.63	13,465	0.69	0.32	0.08	0.01	0.00	0.35	1.06
13	0.92	0.68	10,926	0.71	0.35	0.08	0.01	0.00	0.35	0.84
14	0.94	0.73	9,432	0.73	0.38	0.08	0.00	0.00	0.36	0.94
15	1.13	0.78	8,149	0.74	0.42	0.08	0.01	0.00	0.41	1.14
16	0.89	0.84	7,351	0.75	0.45	0.08	0.04	0.00	0.50	0.99
17	0.94	0.89	7,212	0.76	0.49	0.08	0.04	0.00	0.53	1.04
18	1.08	0.95	5,681	0.78	0.52	0.08	0.02	0.00	0.47	1.14
19	1.33	1.01	4,005	0.81	0.56	0.07	0.00	0.00	0.40	1.09
20	1.18	1.07	4,266	0.82	0.61	0.07	0.07	0.00	0.68	1.11
21	1.22	1.15	2,458	0.86	0.66	0.06	0.02	-0.01	0.54	1.21
22	1.29	1.24	2,697	0.89	0.73	0.05	-0.01	-0.01	0.54	1.34
23	1.33	1.35	1,957	0.92	0.88	0.04	-0.06	-0.01	0.46	1.30
24	1.59	1.50	2,259	1.01	1.09	0.03	-0.06	-0.01	0.47	1.40
25	1.74	1.85	4,014	1.11	1.55	-0.02	-0.33	-0.02	0.83	1.68

资料来源：华安证券研究所整理

4 “水平、斜度和曲线”因子模型（LSC）

下一步是从这些投资组合中提取公共因子。作者使用回归的预测对按预期收益从低到高排序的 25 个投资组合执行主成分分析 (PCA)。PCA 使用特征值分解来识别投资组合中的公共因子。该方法提取测试资产收益的线性组合，以解释协方差矩阵的结构。这种方法将测试资产之间的联动从协方差矩阵转换为不相关的因子。每个因子都形成测试投资组合的一组权重。第一个因子解释了投资组合之间最大的协方差。第二个因子解释了第一个因子未捕获的下一个最大因子，依此类推。

在许多应用中，通常要么对相关矩阵进行特征值分解，要么通过标准化所有变量来重新调整协方差矩阵，因为 PCA 是量纲相关的。如果不首先转换数据，以不同单位表示的变量的分解将无法解释。由于资产定价中我们感兴趣的对象是基础资产的协方差矩阵，并且都表示为投资组合收益，因此无需先对数据进行转换即可进行 PCA。

图表 3 显示了主成分分析的结果。该表列出了前十个分量、各自的特征值和各自解释的方差。前三个分量解释了投资组合 86% 的方差，分别解释了投资组合 74%、9% 和 3% 的方差。图表 1，Panel B 显示了前三个分量的权重。

第一个因子类似于一般市场投资组合，因为它对所有 25 个投资组合的权重大致相等。该因子与所有主流因子模型中使用的 CRSP 值加权市场指数的相关系数为 0.95。因为该因子代表与股票市场整体水平的联动，可以被看作是一个“水平”因子，代表不同投资组合与股市点位同向运动。

图表 3 对 25 个投资组合的主成分分析结果

Component	Eigenvalue	Variance Explained	Cumulative
Component 1	659.89	73.80%	73.80%
Component 2	81.88	9.16%	82.96%
Component 3	27.24	3.05%	86.01%
Component 4	13.66	1.53%	87.54%
Component 5	12.34	1.38%	88.92%
Component 6	10.07	1.13%	90.04%
Component 7	8.45	0.94%	90.99%
Component 8	8.02	0.90%	91.88%
Component 9	7.33	0.82%	92.70%
Component 10	6.53	0.73%	93.43%

资料来源：华安证券研究所整理

第二个因子代表做多低预期收益股票和做空高预期收益股票。这个因子在各个投资组合上的暴露从低预期收益到高预期收益的投资组合几乎单调减少。这个“斜度”因子捕捉到了这样一个特征，即高预期收益股票收益都倾向于一起移动，而低预期收益股票则会向相反的方向，一起移动。由于斜度因子是做多低预期收益股票和做空高预期收益股票，平均而言它有一个负的收益实现。

金融文献中已经确定的大多数因子都是斜度因子。HML 因子捕捉成长型股票与价值型股票相反移动的趋势，而 SMB 因子捕捉小市值股票与大市值股票相反移动的趋势。其他例子包括动量、盈利能力、投资、波动性和流动性都是斜度因子。

该斜度因子的不同之处在于它同时代表着上述所有特征捕获的常见运动。本文感兴趣的潜在特征是投资组合的预期收益，而不是公司层面的预期收益。虽然每个公司层面的特征都提供了一些有关预期收益的信息，但仅基于公司层面特征得到的投资组合，可能会受到与预期收益无关的大量其他共同变动影响。

最后一个“曲线”因子是做空极低和极高收益的资产，同时做多中间的投资组合。这个因子表明极端股票往往会一起移动。如果曲线因子有一个正的收益，那么预期收益非常高和非常低的股票都会有相对较低的收益，而预期收益中等的股票则会有相对较高的收益。

将图表 1 的 Panel B 与图表 2 进行比较，可以得出斜度和曲线因子的特征解释。斜度因子按照公司特征对预期收益的影响程度合并了多个公司特征，并且同时体现了多个投资组合中这些预测因子相互间的协方差。斜度因子包含了账面市值比、动量、资产增长、应计费用、净股票发行量和经营盈利能力这些公司特征，同时斜度因子在实现时将包含暴露在这些股票特征上的共同变动所带来的收益。

图表 3 中曲线因子的载荷与图表 2 中的市值因子变动模式接近。市值规模较小（且波动性较大）的股票更有可能出现在极端高或低预期收益投资组合中。曲线因子的特征解释是，在考虑了水平和斜度后，曲线因子代表了一个共同运动的倾向，即预期收益处于中间部分的投资组合中的市值大的、低波动的股票预期收益往往会向相同的方向共同移动，而处于极端预期收益组合中的市值小的、高波动性的股票收益则往往会向相反的方向共同移动。虽然该因子与公司规模有关，但作者将在 4.2 节中说明它与文献中常用的市值因子不太相似。

4.1 因子数量的确定

主成分分析法不规定因子的数量，涉及到究竟选择使用多少个主成分来表示数据的问题。之前的研究将其概念化为一个模型选择问题，在拟合优度和简约性之间进行权衡，并开发了三个检验统计量（BIC1、BIC2 和 BIC3）。作者分别对 25 个和 100 个异象投资组合计算了这些检验统计量。三个检验统计量都发现存在三个因子。

图表 4 不同分组数量下前 5 种主成分的相关性展示

First Component			
	10 Portfolios	25 Portfolios	100 Portfolios
10 Portfolios	1.000		
25 Portfolios	0.999	1.000	
100 Portfolios	0.994	0.997	1.000
Second Component			
	10 Portfolios	25 Portfolios	100 Portfolios
10 Portfolios	1.000		
25 Portfolios	0.980	1.000	
100 Portfolios	-0.955	-0.976	1.000
Third Component			
	10 Portfolios	25 Portfolios	100 Portfolios
10 Portfolios	1.000		
25 Portfolios	0.928	1.000	
100 Portfolios	0.855	0.931	1.000
Fourth Component			
	10 Portfolios	25 Portfolios	100 Portfolios
10 Portfolios	1.000		
25 Portfolios	0.609	1.000	
100 Portfolios	0.461	0.611	1.000
Fifth Component			
	10 Portfolios	25 Portfolios	100 Portfolios
10 Portfolios	1.000		
25 Portfolios	0.092	1.000	
100 Portfolios	0.207	0.361	1.000

资料来源：华安证券研究所整理

4.2 与主流因子的相关性情况

图表 5 显示了水平、斜度和曲线因子与其他几个常见因子的相关性。三个提取的因子中没有一个与 HML 的相关性超过 0.35。SMB 与水平因子的相关性为 0.46，与曲线因子的相关性为 0.49，在某种程度上呼应了图表 2 中特征的形状，但与斜度因子的相关性低于 0.30。斜度因子与动量的相关性为 0.77，是表中最强的相关性。盈利能力因子 RMW、ROE 和 PMU 与斜度因子和曲线因子的相关性较低。投资与水平、斜度和曲线因子的相关性较低。这三个因子中没有一个与流动性表现出很大的相关性。斜度和曲线因子与这些常用模型中的因子相关但又不同。

图表 5 前五个主成分与其他因子的相关性。

Variables	PC 1	PC 2	PC 3	PC 4	PC 5
Mkt-RF	0.96	0.20	0.13	0.03	-0.03
SMB	0.47	-0.29	-0.47	-0.11	0.10
HML	-0.31	-0.13	0.23	0.00	-0.11
RMW	-0.31	-0.01	0.41	0.09	-0.03
CMA	-0.39	-0.26	0.20	-0.02	-0.05
MOM	-0.04	-0.76	0.15	-0.03	-0.05
ME	0.46	-0.33	-0.46	-0.13	0.10
INV	-0.40	-0.23	0.28	-0.01	-0.12
ROE	-0.23	-0.26	0.33	0.05	-0.07
Liq-T	-0.07	0.06	0.08	-0.04	0.06

资料来源：华安证券研究所整理

5 资产定价检验

根据套利定价理论，股票投资组合所产生的超额收益将通过每个投资组合对公共因子的载荷来解释。

图表 6 给出了投资组合超额收益对前一、二、三、四个主成分的时间序列回归结果。由于这些因子不相关，因此回归的贝塔系数的模式由图表 1 的 Panel B 中所示的因子载荷所示。该表展示了 25 个投资组合对依次加入前一、二、三、四个主成分的模型进行时间序列回归时，每一个回归的 α 、t 统计量和 R 平方。

图表 6 对提取的主成分进行 25 个投资组合的时间序列回归。

Port	XRet	α_1	α_2	α_3	α_4	t_1	t_2	t_3	t_4	R^2_1	R^2_2	R^2_3	R^2_4
1	-1.08	-2.15	-0.94	-0.61	-0.15	-9.48	-6.16	-5.34	-1.92	0.61	0.85	0.91	0.97
2	-0.16	-1.03	-0.18	0.00	-0.09	-6.03	-1.30	-0.01	-0.73	0.63	0.82	0.85	0.85
3	0.15	-0.66	0.26	0.37	0.13	-3.84	1.96	2.76	0.97	0.59	0.83	0.84	0.86
4	0.32	-0.37	0.37	0.31	0.12	-2.60	2.92	2.33	0.90	0.61	0.82	0.82	0.84
5	0.45	-0.21	0.33	0.21	0.10	-1.80	3.13	2.16	0.99	0.65	0.79	0.82	0.83
6	0.49	-0.17	0.34	0.19	0.11	-1.53	3.50	2.20	1.30	0.69	0.82	0.86	0.86
7	0.55	-0.12	0.16	0.00	-0.16	-1.18	1.64	-0.01	-2.07	0.75	0.79	0.84	0.86
8	0.70	0.02	0.30	0.11	0.10	0.25	3.17	1.46	1.28	0.76	0.80	0.86	0.86
9	0.74	0.07	0.22	0.03	0.02	0.76	2.25	0.38	0.29	0.78	0.79	0.86	0.86
10	0.85	0.16	0.25	0.07	0.06	1.92	2.75	0.95	0.81	0.82	0.82	0.88	0.88
11	0.86	0.18	0.11	-0.07	-0.06	2.05	1.23	-0.95	-0.83	0.81	0.81	0.87	0.87
12	0.77	0.07	-0.03	-0.21	-0.17	0.79	-0.36	-2.85	-2.27	0.82	0.82	0.88	0.88
13	0.92	0.19	0.05	-0.10	-0.03	2.23	0.53	-1.26	-0.42	0.83	0.84	0.87	0.88
14	0.94	0.23	0.02	-0.11	-0.06	2.55	0.21	-1.29	-0.74	0.81	0.83	0.86	0.86
15	1.13	0.38	0.09	-0.02	-0.02	3.95	0.93	-0.25	-0.21	0.79	0.82	0.84	0.84
16	0.89	0.11	-0.09	-0.17	-0.16	1.30	-0.99	-1.94	-1.90	0.84	0.86	0.87	0.87
17	0.94	0.14	-0.16	-0.23	-0.14	1.43	-1.74	-2.56	-1.59	0.82	0.85	0.86	0.86
18	1.08	0.26	-0.07	-0.11	-0.09	2.55	-0.79	-1.22	-0.95	0.82	0.86	0.86	0.86
19	1.33	0.48	0.16	0.14	0.27	4.85	1.67	1.52	2.88	0.82	0.86	0.86	0.87
20	1.18	0.28	-0.24	-0.20	-0.08	2.32	-2.00	-1.89	-0.80	0.77	0.85	0.85	0.86
21	1.22	0.36	-0.07	-0.07	0.00	3.55	-0.79	-0.71	0.05	0.82	0.88	0.88	0.89
22	1.29	0.37	-0.12	-0.03	0.11	3.03	-1.11	-0.29	0.97	0.78	0.85	0.85	0.86
23	1.33	0.44	-0.03	0.07	0.31	3.35	-0.25	0.56	2.76	0.75	0.81	0.82	0.85
24	1.59	0.59	-0.03	0.23	-0.02	3.99	-0.20	2.06	-0.24	0.73	0.81	0.86	0.88
25	1.74	0.67	0.01	0.33	-0.05	4.00	0.07	2.78	-0.49	0.72	0.80	0.87	0.92
α		0.39	0.16	0.16	0.10					0.75	0.83	0.86	0.87
GRS		5.21	3.10	2.54	1.41								
p-val		[0.00]	[0.00]	[0.00]	[0.09]								

资料来源：华安证券研究所整理

在第三列中，较大的 α_1 表明超额收益的巨大差异未被第一个因子捕获。这种回归几乎与传统的 CAPM 相同，第四列显示第二个因子解释了 α 的很大一部分，

添加斜度因子后，25 个回归的平均 R 平方从 75% 上升到 83%，而平均 alpha 从每月 0.39% 下降到每月 0.16%。第五列显示添加曲线因子将平均 R 平方增加到 86%，而平均 alpha 没有变化。第四个因子几乎没有增加额外的 R 平方，虽然它降低了平均 alpha，但它在 25 个投资组合的因子载荷图像上是一个令人困惑的锯齿形图案，GRS 检验显示前三个因子在 5% 的水平下被拒绝，而第四个因子在 10% 的水平下被拒绝。因此推断，“水平、斜度和曲线”三因子模型已捕获了平均收益的大部分变化，以及 25 个投资组合预期收益的大部分方差。

6 与主流模型的比较

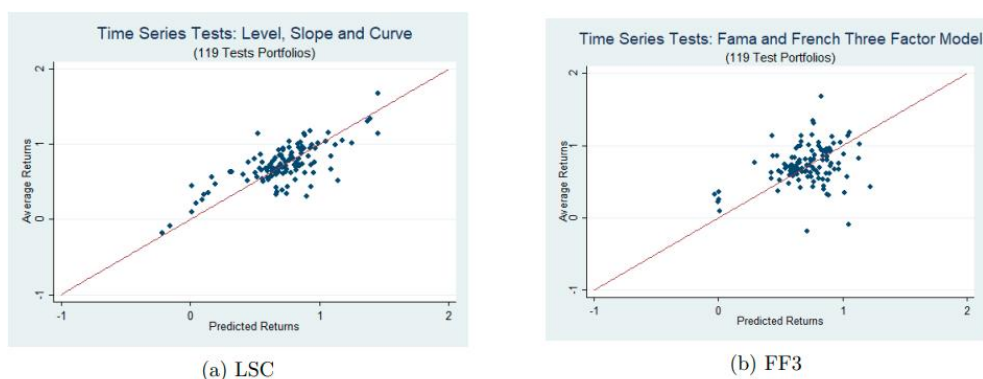
在这一节中，作者进行了一些资产定价测试，以比较水平、斜度、曲线模型和其他主流模型的效果。目前学术上主流的资产定价模型有：Fama-French 三因子模型 (Fama and French, 1993)：使用市场投资组合和两个多空投资组合，一个做多高市净率股票和做空低市净率股票 (HML)，另一个做多小市值股票和做空大市值股票 (SMB)；Fama 和 French 四因子模型，也称为 Carhart (1997) 模型，增加了一个动量因子 (MOM)，即做多过去 12 个月上涨的股票和做空过去 12 个月下跌的股票；Fama 和 French (2015a) 五因子模型：排除了动量，纳入了做多低投资股票和做空高投资股票的因子 (CMA)，以及做多高利润股票和做空低利润股票的因子 (RMW)；在此基础上，作者添加动量因子创建出 Fama 和 French 六因子模型。此外，还有 Hou 等人的四因子模型，模型使用市场收益 (MKT)、公司市值 (SIZE)、资产投资收益率 (INV) 和股本收益率 (ROE) 作为因子衡量盈利能力。作者将与这些模型比较水平、斜度、曲线因子模型的定价效果。

对于测试资产，作者使用了 119 个股票和债券构成的投资组合。其中包括第 4 节中的由“扩展窗口异象”回归形成的 10 个投资组合；25 个按规模和账面市值比排序构成的投资组合；10 个按动量排序的投资组合；5 个美国国债（1 年、5 年、10 年、20 年和 30 年）构成的投资组合；49 个行业组合；10 个基于运营盈利能力排序构成的组合；以及 10 个基于投资（资产增长）排序构成的投资组合。

6.1 LSC 模型与 Fama 和 French 三因子模型的比较

首先，作者使用时间序列检验将水平、斜度和曲线模型与 Fama 和 French 三因子模型进行了比较。图表 7 以图形方式显示了 119 个测试组合在各自的三因子模型上的时间序列测试结果。X 轴是模型的预期收益，即每个测试资产的因子载荷乘以因子平均收益。Y 轴是平均收益。定价误差 (alpha) 是每个绘制的测试资产与 45 度线的垂直距离。零 alpha 的完美定价会将每个测试资产放在 45 度线上。

图表 7 119 个测试投资组合两种模型的拟合比较



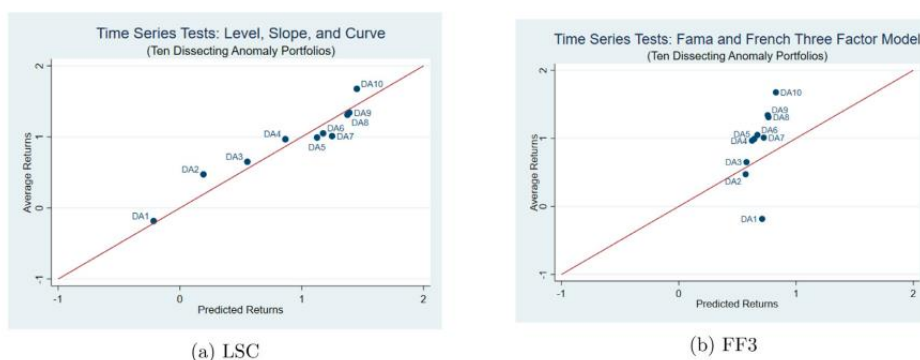
资料来源：华安证券研究所整理

右图是 Fama 和 French 三因子模型的结果，在横截面上有些混乱。左侧面板显示水平、斜度和曲线模型捕获了平均收益的分化，在横截面上非常清晰。为了更清晰的展现拟合优度，将平均收益对预期收益进行回归，考察回归的 R 平方（同时仍然施加时间序列限制）。Fama 和 French 三因子模型的 R 平方为 12%，而水平、斜度和曲线模型的 R 平方为 56%。

图表 8 到图表 13 分解了图表 7 聚集的点，分别显示了每组投资组合下两种模型的定价比较结果。

图表 8 显示了针对 10 个投资组合测试时，水平、斜度、曲线模型与 Fama 和 French 三因子模型的定价误差比较，这 10 个投资组合是使用对七种公司特征的扩展回归排序形成的。水平、斜度和曲线模型几乎解释了各个投资组合平均收益的差异。Fama 和 French 三因子模型显示出相当大的定价误差，对一些高预期收益的投资组合只有一点点定价能力。

图表 8 10 个异常投资组合下模型定价能力对比

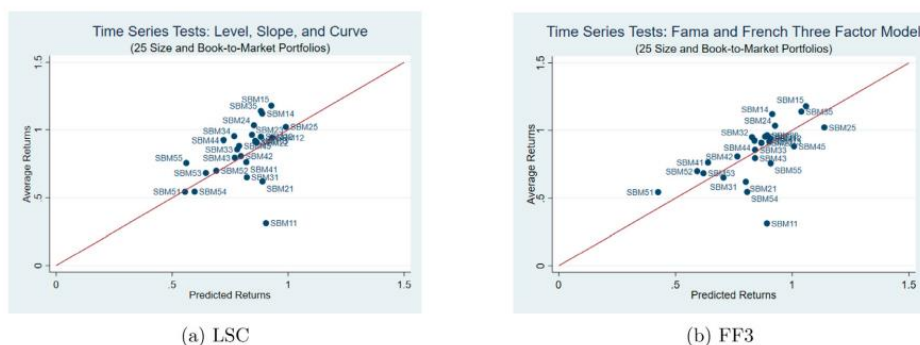


资料来源：华安证券研究所整理

图表 9 显示了根据市值和账面市值比排序得到的 25 个投资组合进行定价时，Fama 和 French 三因子模型的初步成功。图中每个测试资产都由其排序的投资组合标记，首先根据市值分成五组（最小到最大），然后再根据账面市值比分成五组（成长到价值）。显然，小市值成长投资组合(SBM11)使两个因子模型定价失效。虽然水平、斜度和曲线模型解释了大部分平均收益，但它在极值投资组合(SBM15、SBM35)

上留下了 alpha 值，并且对第二小的成长投资组合（SBM21）定价效果不如 Fama 和 French 三因子模型。

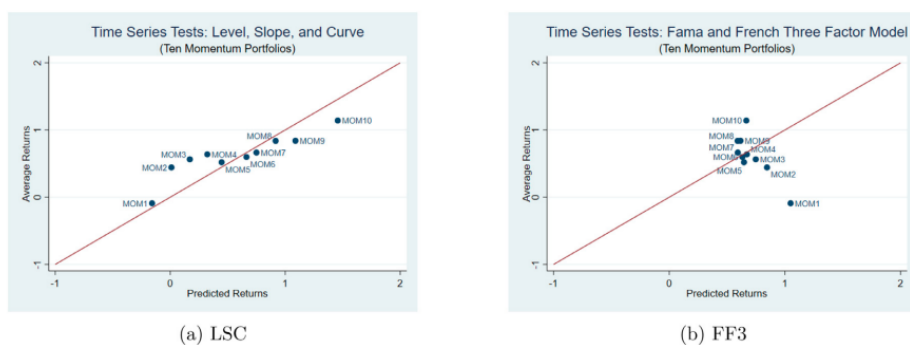
图表 9 模型关于 25 个市值和账面市值比投资组合的定价比较



资料来源：华安证券研究所整理

图表 10 显示了两种模型在根据动量排序的投资组合测试下的鲜明对比。水平、斜度和曲线模型基本可以解释动量带来的收益差异，然后捕获了一些略微逆转动量异象。Fama 和 French 三因子模型的定价误差大于平均收益的差异，无法有效地解释动量带来的收益。

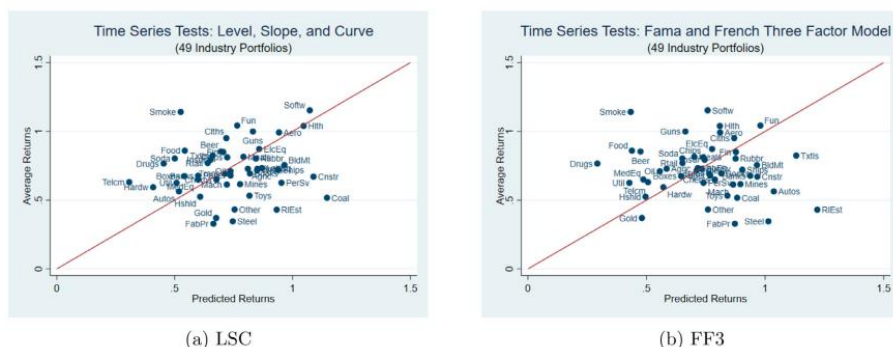
图表 10 模型关于 10 个动量投资组合的定价比较



资料来源：华安证券研究所整理

图表 11 显示了在 49 个行业上形成的投资组合下的两个模型定价效果的比较。虽然行业投资组合中各行业的平均收益差异较小（注意轴的比例），但水平、斜度和曲线模型的定价误差比 Fama 和 French 三因子模型小。两个模型下的平均绝对 alpha 值分别为 15 和 21 个基点。

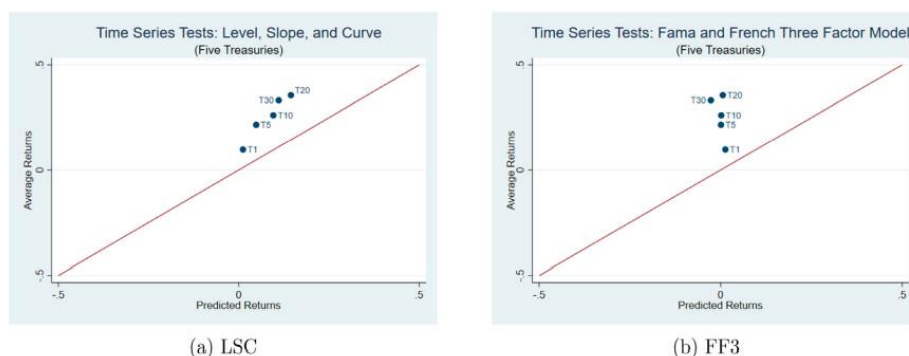
图表 11 模型关于 49 个行业分类投资组合的定价比较



资料来源：华安证券研究所整理

图表 12 显示了使用期限从 1 年到 30 年不等的美国国债投资组合对两种模型进行测试时的比较结果。水平、斜度和曲线模型显示预期收益定价的差异开始与平均收益保持一致。Fama 和 French 三因子模型几乎无法解释国债的预期收益差异。对于水平、曲线、斜率因子来说，随着到期日的增加，差异来自曲线因子的显著正载荷。这表明曲线因子反映了该因子长期持有的大市值低波动性股票，即曲线因子表现得像长期债券的特点。该证据表明，曲线因子代表与久期相关的潜在风险因子。

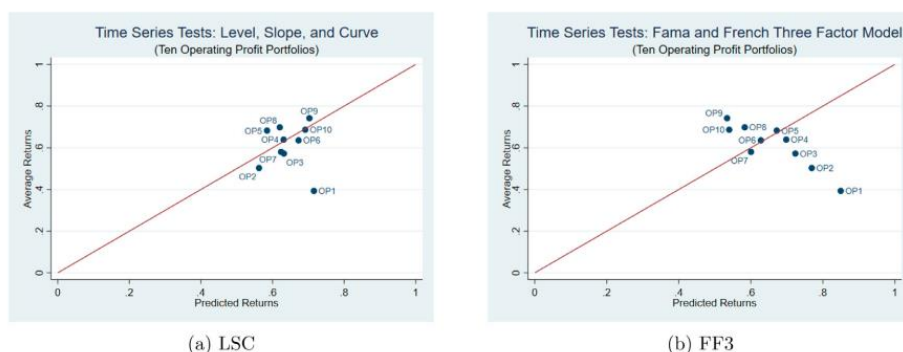
图表 12 模型关于使用 5 种债券收益的投资组合的定价比较



资料来源：华安证券研究所整理

图表 13 显示了使用基于运营盈利能力排序得到的十个投资组合进行测试时两个模型效果的比较。大多数 alpha 实际上是由模型产生的，因为从利润最高的前百分之十投资组合到利润最低的后百分之十的平均收益的差异仅为 29 个基点。最低利润的股票投资组合(OP1)具有高水平和市场贝塔值这一事实导致这两种模型对其定价效果均较差。虽然它在左图中作为异常值突出显示，但其他九个投资组合的平均收益被预测的相当好。在右图中，低利润和高利润股票都混淆了 Fama 和 French 三因子模型，因此预测收益的排名几乎与平均收益的排名相反。

图表 13 模型关于 10 个运营盈利能力投资组合的定价比较



资料来源：华安证券研究所整理

6.2 LSC 模型与主流模型的比较

接下来，作者将进行水平、斜度和曲线模型与使用其他资产定价主流模型的比较。作者发现水平、斜度和曲线模型的性能至少与其他主流模型相当，并且通常表现得更好。图表 14 显示了对 119 个测试组合的时间序列测试结果，比较了水平、斜度和曲线模型(LSC)和样本外水平、斜度和曲线扩展窗口模型(LSC-NP)，资本资产定价模型 (CAPM)、Fama 和 French 三因子模型 (FF3)、Carhart 模型 (CAR)、Fama 和 French 五、六因子模型，以及 Hou、Xue、Zhang 四因子模型 (HXZ)。第一行显示水平、斜度和曲线模型的平均绝对 alpha 为 15 个基点。CAPM 的平均绝对 Alpha 值最高，为 23 个基点，其次是 Fama 和 French (1993)，为 21 个基点，Fama 和 French (2015a) 为 20 个基点，Hou 等人(2015)的模型具有 18 个基点。Carhart 模型以及 Fama 和 French 六因子模型的表现几乎与 LSC 模型一样好，具有 16 个基点的 alpha。

图表 14 将水平、斜度和曲线与具有 119 个测试组合的领先因子模型进行比较

Time Series Results								
	LSC	LSC-EW	CAPM	FF3	CAR	FF5	FF6	HXZ
Avg $ \alpha $	0.15	0.15	0.23	0.21	0.16	0.20	0.16	0.18
$ t > 1.96$	18	21	41	38	27	37	29	26
Avg R^2	0.70	0.69	0.66	0.72	0.74	0.74	0.76	0.72
CS (TS) R^2	0.54	0.49	0.03	0.09	0.46	0.23	0.45	0.37
Sh - Full	0.38	0.33	0.14	0.21	0.30	0.35	0.38	0.43
Sh - Split	0.23	0.23	0.12	0.15	0.19	0.21	0.22	0.23
Sh - Rolling	0.27	0.25	0.12	0.14	0.19	0.19	0.22	0.24

资料来源：华安证券研究所整理

第二行显示在 5%置信水平下显著的 t 统计量的数量。LSC 和 LSC-NP 模型留下的显著 alpha 最少。

第三行显示时间序列回归的平均 R 平方。这不是对模型的测试，预期收益的正确模型不需要解释时间序列变化，但了解每个模型对时间序列变化的捕捉程度可能是有意义的。

第四行显示了时间序列检验的横截面 R 平方，它用作模型预期收益捕获的平均收益变化百分比的汇总统计量。它充当以上各图中以图形方式呈现的证据的汇总统计量。LSC 和 LSC-NP 模型都解释了平均收益的最大变化，R 平方分别为 54%和 49%。其次是 Carhart 模型 46%和 Fama 和 French 六因子模型 45%。

最后三行表示均值方差最优化(mean-variance efficiency)下的投资组合的夏普, Sh-Full 表示样本内结果, Sh-Split 表示作者将样本分成两半, 仅使用样本的前半部分计算事后最优权重, 并使用样本的后半部分显示该投资组合的最终夏普比率。水平、斜度和曲线模型都优于所有 Fama 和 French 模型。Sh-Rolling 表示通过每月向前滚动并使用所有先前数据重新计算最佳均值方差有效因子权重来构造组合, 水平、斜度和曲线模型仍然优于 Fama 和 French 模型。令人惊讶的是, 三因子、五因子和六因子模型在使用滚动窗口形成的权重时相比“Split”样本表现得更差, 即使“Rolling”包含一个更大的信息集。Hou 等人的模型在“Split”样本中的表现与水平、斜度和曲线模型一样好, 基准和扩展窗口模型的夏普比率为 0.23, 但在滚动窗口测试中却一般: Hou 等人模型仅提高到 0.24, 而水平、斜度和曲线模型分别提高到 0.27 和 0.25。

7 结论

本文开发了一种提取股票收益横截面定价因子的新方法。第一步对多个预测变量使用横截面回归, 将股票按高预期收益到低预期收益分类到投资组合中。第二步使用主成分从这些投资组合中提取因子。这种方法的目标是根据预期收益对投资组合进行排序, 然后提取与预期收益相关的因子, 分散非定价因子的风险敞口。最终形成的因子模型是“水平、斜度和曲线”因子模型。作者测试新模型的资产定价效果, 将水平、斜度和曲线模型与几个主要的主流模型进行比较, 发现该模型在使用样本内和样本外测试时都表现得非常好。

文献来源:

核心内容摘选自 Charles Clarke 在《Journal of Financial Economics》上的论文《The level,slope,and curve factor model for stocks》。

风险提示:

本文结论基于历史数据与海外文献进行总结; 不构成任何投资建议。

重要声明

分析师声明

本报告署名分析师具有 PRC 证券业协会授予的证券投资咨询执业资格，以勤勉的执业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

免责声明

华安证券股份有限公司经 PRC 证券监督管理委员会批准，已具备证券投资咨询业务资格。本报告中的信息均来源于合规渠道，华安证券研究所力求准确、可靠，但对这些信息的准确性及完整性均不做任何保证。在任何情况下，本报告中的信息或表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送，未经华安证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。如欲引用或转载本文内容，务必联络华安证券研究所并获得许可，并需注明出处为华安证券研究所，且不得对本文进行有悖原意的引用和删改。如未经本公司授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

投资评级说明

以本报告发布之日起 6 个月内，证券（或行业指数）相对于同期沪深 300 指数的涨跌幅为标准，定义如下：

行业评级体系

- 增持—未来 6 个月的投资收益率领先沪深 300 指数 5%以上；
- 中性—未来 6 个月的投资收益率与沪深 300 指数的变动幅度相差-5%至 5%；
- 减持—未来 6 个月的投资收益率落后沪深 300 指数 5%以上；

公司评级体系

- 买入—未来 6-12 个月的投资收益率领先市场基准指数 15%以上；
- 增持—未来 6-12 个月的投资收益率领先市场基准指数 5%至 15%；
- 中性—未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差-5%至 5%；
- 减持—未来 6-12 个月的投资收益率落后市场基准指数 5%至；
- 卖出—未来 6-12 个月的投资收益率落后市场基准指数 15%以上；
- 无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。市场基准指数为沪深 300 指数。