

知乎

首发于
川流不息

解释股票截面预期收益差异的独立因子



石川

量化交易 话题的优秀回答者

已关注

51 人赞同了该文章

摘要

Green, Hand and Zhang (2017) 通过 Fama-MacBeth 回归同时检验了 94 个因子，发现其中 12 个是显著的。他们的研究方法对于 A 股的因子投资实践有很大的价值。

1 引言

在发表偏差 (publication bias) 的驱使下，多重检验 (multiple testing) 已经成为学术界在进行 asset pricing 时的普遍做法。然而，很多发表于顶级期刊的所谓能够解释股票截面预期收益差异的因子都是虚假发现 (见《出色不如走运(II)?》)。针对这个问题，2011 年，前美国金融协会主席 John Cochrane 教授在主席演讲中向学术界提出挑战——**到底哪些因子能独立解释股票预期收益截面差异？**



▲ 赞同 51 ▼

● 5 条评论

➤ 分享

★ 收藏



知乎

首发于
川流不息

we are going to have to repeat Fama and French's anomaly digestion, but with many more dimensions. We have a lot of questions to answer:

First, which characteristics really provide *independent* information about average returns? Which are subsumed by others?

Second, does each new anomaly variable also correspond to a new factor formed on those same anomalies? Momentum returns correspond to regression coefficients on a winner-loser momentum "factor." Carry-trade profits correspond to a carry-trade factor.¹⁹ Do accruals return strategies correspond to an accruals factor? We should routinely look.

Third, how many of these new factors are really important? Can we again account for N independent dimensions of expected returns with $K < N$ factor exposures? Can we account for accruals return strategies by betas on some other factor, as with sales growth?

知乎 @石川

2017 年，一篇发表于顶刊 Review of Financial Studies 上的文章通过使用 Fama-MacBeth Regression (Fama and MacBeth 1973) 同时检验 94 个学术界发表的异象回答了这个问题 (Green, Hand and Zhang 2017, 下文记为 GHZ)。

事实上，这并不是学术界就此问题的唯一尝试。比如，Harvey and Liu (2018) 提出了一个基于回归的框架来排除运气的成分以便检验真正有效的因子。不过 Harvey and Liu (2018) 这篇文章的核心内容是介绍他们提出的方法而非实证结果。而 GHZ 的研究指出了美股上真正有效的因子，具有更高的实证价值。无论是这些因子还是 GHZ 使用的方法对 A 股都有不小的借鉴意义。

本文就来介绍 GHZ 这篇论文。

2 数据和因子

在研究中，GHZ 利用了 CRSP, Compustat 以及 I/B/E/S 数据库的数据首先挑选出了 102 个因子，回测时间是 1980 年 1 月到 2014 年 12 月，频率为月频。回测期之所以从 1980 年开始是由上市公司基本面数据的可得性决定的；同时包含 NYSE、AMEX 以及 NASDAQ 的上市公司。

除此之外，**GHZ 的特别之处是考虑了微小市值公司对评价因子有效性时的影响**。为此，他们将所有股票根据市值分成三挡：

1. 市值超过中位数的股票为大市值组；
2. 市值在中位数和 20% 分位数之间的为小市值组；
3. 市值在 20% 分位数之下的为微小市值 (microcap) 组。

微小市值的这些股票的市值之和仅占有所有股票总市值的 3%。为了避免回归分析时对它们过度加权，GHZ 采取了两种回归方法：

▲ 赞同 51 ▼

● 5 条评论

➤ 分享

★ 收藏

...

作者认为上述两种方法能够更合理的找到有效因子，并通过合并这两种回归的结果确定最终的显著因子。作为比较——同时也是为了说明过度加权微小市值股票的危害——GHZ 也同样对所有的股票进行了 OLS 回归。不出意外，由于微小市值股票存在造成的偏误，这种方法能够发现更多的“有效”因子，但很多因子在非微小市值股票上并不显著。

在进行多因子同时回归时，因子间的共线性是必须考虑的问题。**共线性会造成因子收益率标准误 (standard error) 的巨大误差，从而无法有效的评价因子是否显著**（见《为什么要进行因子正交化处理?》）。GHZ 使用 variance inflation factors (VIF) 对因子进行了初步筛选，并剔除了 8 个 VIF 高于 7 的因子（下图），最终从 102 个因子中选出了 94 个待评价的因子。这 94 个因子的平均相关系数为 0.07。

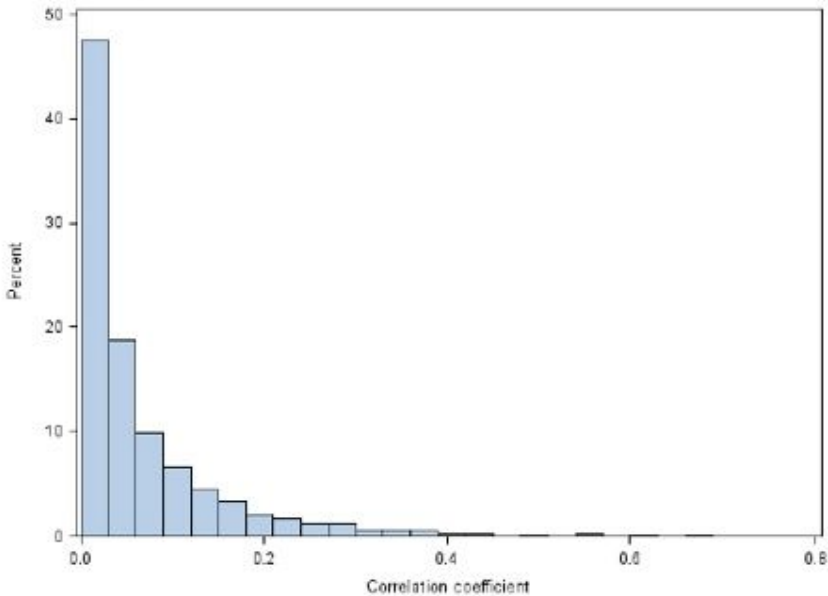
Panel A: Variance inflation factors (VIFs)

	# characteristics	Min.	10th pct.	Median	Mean	90th pct.	Max.
Before removing VIF > 7	102	1	1.1	1.8	3.8	9.8	22.1
After removing VIF > 7	94	1	1.1	1.6	2.1	3.9	6.9

Panel B: Absolute cross-correlations

	# characteristics	Min.	10th pct.	Median	Mean	90th pct.	Max.
Before removing VIF > 7	102	0	0.004	0.03	0.07	0.18	0.95
After removing VIF > 7	94	0	0.004	0.03	0.07	0.18	0.77

Panel C: Distribution of the absolute cross-correlations among characteristics after removal of characteristics with VIFs>7



出处：Green, Hand and Zhang (2017)

知乎 @石

知乎

首发于
川流不息

(Benjamini and Yekutieli 2001, 见《出色不如走运(II)?》)。GHZ 以调整后的 p-value 小于 0.05 作为因子显著的准则。

除了调整后的 p-value 之外, GHZ 也同时计算了每个因子的 t-statistic。Harvey et al. (2016) 指出由于 multiple testing 的问题, 单个因子的 t-statistic (的绝对值) 至少要大于 3 才可能显著。GHZ 也以此为显著性准则汇报了显著性因子的个数。该准则比调整后 p-value 小于 0.05 要弱一些, 因此找到了更多的因子。

下面就来看看实证结果。

3 单个因子检验

在同时对 94 个因子进行检验之前, GHZ 首先对每个因子进行了单独检验。作者这么做的目的是希望通过单独检验和共同检验分别找出显著性的因子, 以此比较这两种方法找到的因子是否一致, 这可能会对未来的 asset pricing 研究带来一些启发和灵感。

在单个因子检验中, 截面回归方程的左侧是个股的收益率, 右侧是目标因子和其他控制变量。根据有无控制变量, 单个因子的检验一共有以下四个版本:

1. 仅有单个目标因子, 无其他控制变量;
2. 除了单个目标因子外, 还包括四因子模型 (Fama and French 1993 的三因子 + Carhart 1997 的动量因子);
3. 除了单个目标因子外, 还包括 Fama-French 五因子模型 (Fama and French 2015);
4. 除了单个目标因子外, 还包括 Hou et al. (2015) 的因子模型。

对每一个版本, 通过在每期用因子暴露和下一期的个股收益率截面回归得到因子收益率, 然后使用因子收益率的时序对因子显著性进行分析, 计算得到调整后的 p-value 以及 t-statistic, 使用 p-value 是否小于 0.05 或 t-statistic 绝对值是否大于 3.0 为基准决定真正显著的独立因子。在回归时, 考虑了本文前一节提到的三种方法——对所有股票进行 VWLS, 对非微小市值股票进行 OLS, 以及 (作为“反面教材”的) 对所有股票进行 OLS。为了方便在下文中介绍结果, 将前两种回归称为 GHZ 回归。

单个因子检验的结果如下 (由于 94 个因子太多, 故节选部分):



▲ 赞同 51 ▼

● 5 条评论

➤ 分享

★ 收藏



	Column A: Single characteristic, No benchmark model controls:						Column B: Single characteristic, Carhart benchmark model controls:						Column C: Single characteristic, 5-factor benchmark model controls:						Column D: Single characteristic, q-factor benchmark model controls:											
	All stocks, VWLS			All-but- microcap, OLS			All stocks, VWLS			All-but- microcap, OLS			All stocks, VWLS			All-but- microcap, OLS			All stocks, VWLS			All-but- microcap, OLS								
	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.	FM coef.	t- stat.						
# DFDR $p \leq 0.05$	1			12			5			15			6			4			20			2			3			19		
# p -stats ≥ 3.0	3			16			14			20			8			7			25			3			5			26		
agr	-0.21	-3.1	-0.26	-4.6	-0.42	-8.1	-0.19	-3.5	-0.24	-5.1	-0.35	-8.3																		
bm	0.17	1.1	0.14	1.8	0.32	4.8													0.13	0.9	0.10	1.4	0.19	2.6						
mom12m	0.30	2.4	0.20	1.9	0.21	2.1							0.26	2.2	0.20	2.0	0.22	2.5	0.27	2.3	0.20	1.9	0.21	2.4						
mva	-0.10	-1.0	-0.05	-1.0	-0.25	-2.3																								
operprof	0.08	1.5	0.06	2.1	0.05	1.4	0.10	2.3	0.08	2.7	0.07	1.9							0.09	1.9	0.07	2.6	0.07	2.4						
roeq	0.18	2.5	0.14	2.5	0.20	2.3	0.21	3.2	0.14	2.9	0.23	3.6	0.20	3.4	0.15	3.0	0.26	4.0												
abnacc	-0.03	-0.3	-0.08	-1.3	-0.06	-0.8	-0.07	-0.9	-0.08	-1.9	-0.08	-1.3	-0.01	-0.1	-0.04	-1.0	-0.06	-1.1	0.02	0.2	-0.03	-0.6	-0.05	-1.0						
acc	-0.20	-2.5	-0.11	-2.5	-0.16	-2.6	-0.12	-2.0	-0.08	-2.2	-0.12	-2.5	-0.17	-2.5	-0.08	-2.2	-0.07	-1.6	-0.18	-2.4	-0.09	-2.3	-0.09	-2.1						
anvol	0.02	0.3	0.00	0.0	0.08	4.7	-0.02	-0.4	-0.01	-0.6	0.04	2.4	0.01	0.2	0.00	0.1	0.06	3.6	0.01	0.2	0.00	0.1	0.05	3.1						
age	0.02	0.2	0.08	1.2	0.13	1.8	0.06	0.9	0.10	1.8	0.21	4.5	-0.03	-0.3	0.02	0.3	0.14	2.8	0.00	0.0	0.04	0.5	0.17	3.0						
buypread	-0.42	-1.2	-0.20	-1.5	0.10	0.6	-0.61	-2.2	-0.26	-2.2	-0.01	-0.1	-0.40	-1.3	-0.17	-1.3	0.00	0.0	-0.41	-1.3	-0.17	-1.2	0.01	0.1						
beta	-0.07	-0.4	-0.10	-0.7	-0.11	-0.8	-0.15	-1.0	-0.14	-1.1	-0.08	-0.6	-0.03	-0.2	-0.06	-0.4	-0.04	-0.3	-0.03	-0.2	-0.06	-0.4	-0.04	-0.3						
bm_ia	0.04	0.8	0.00	-0.1	0.01	0.2	0.03	0.7	0.01	0.2	0.02	0.4	0.04	0.9	0.01	0.3	0.02	0.4	0.05	1.1	0.01	0.1	0.00	0.1						
cash	0.11	1.1	0.04	0.3	0.10	1.0	0.11	1.4	0.05	0.5	0.14	1.5	0.17	1.9	0.10	1.0	0.17	1.8	0.15	1.6	0.08	0.8	0.14	1.5						
cashdebt	0.24	1.9	0.08	2.0	0.05	0.7	0.29	3.4	0.11	3.2	0.06	1.1	0.29	3.3	0.10	2.8	0.08	1.5	0.25	2.3	0.07	1.9	0.06	1.2						
cashpr	-0.08	-1.7	-0.10	-2.3	-0.14	-3.2	-0.03	-1.0	-0.05	-1.8	-0.04	-1.2	-0.06	-1.8	-0.06	-2.2	-0.03	-1.1	-0.07	-1.7	-0.08	-2.2	-0.09	-1.9						
cfp	0.14	1.5	0.11	1.8	0.13	1.7	0.08	1.3	0.09	2.2	0.11	2.0	0.07	1.0	0.07	1.7	0.12	2.1	0.05	0.6	0.06	1.2	0.10	1.6						
cfp_ia	-0.01	-0.2	0.00	0.1	0.04	1.0	-0.01	-0.3	0.02	0.5	0.04	1.2	0.00	0.1	0.02	0.6	0.04	1.2	0.00	0.0	0.01	0.3	0.03	0.9						
chatoia	0.07	1.6	0.08	3.9	0.08	5.0	0.06	1.5	0.07	3.6	0.07	4.7	0.02	0.5	0.04	1.9	0.05	2.9	0.02	0.5	0.04	1.8	0.04	2.6						

出处：Green, Hand and Zhang (2017)

知乎 @石川

上图中最重要的部分是红色标记的地方 —— 即使用调整后 p-value 小于 0.05 为基准确定的显著因子个数。在没有控制变量的模型中（column A），GHZ 回归一共找到了 12 个显著因子；作为对比，由于微小市值股票的影响，对所有股票的 OLS 回归发现了 30 个显著因子。这说明 18 个因子在非微小市值股票上是无效的。其他带控制变量的回归（columns B to D）也有类似的结果。

在没有控制变量的模型中，通过单个因子回归分析发现的显著因子包括：

1. 总资产增长率（asset growth）；

2. 经行业调整的销售收入增长率（growth in industry-adjusted sales）；

3. 流通股本变动百分比（growth in shares outstanding）；

4. 存货变动（growth in inventory）；

5. 盈余公告宣告收益（earnings announcement return）；

6. 账面价值增长率（growth in book value of equity）；

7. 资本支出增长率（growth in CAPEX）；

8. 长期净营运资产增长率（growth in long-term net operating assets）；

9. 机器设备厂房存货增长率（growth in PP&E plus inventory）；

10. 盈利同比增长的季度数（number of consecutive quarters with earnings higher than the same quarter a year ago）；

11. 销售收入与存货差的增长率（growth in sales less growth in inventory）；

12. 标准化的未预期盈余（standardized unexpected quarterly earnings）。

4 多个因子共同检验

在共同检验时，GHZ 将所有 94 个因子同时放入回归方程的右侧，进行 Fama-MacBeth regression。和上一节一样，仍然使用三种回归方法。结果（节选）如下：

Results of Fama-MacBeth (FM) regressions of monthly stock returns on all 94 firm characteristics simultaneously. Coefficients with a two-tailed DFDR p-value ≤ 0.05 are shown with a bolded t-statistic. The data window is Jan. 1980 – Dec. 2014.

	Set of stocks, regression method					
	Column A: All stocks, VWLS		Column B: All-but-microcap stocks, OLS		Column C: All stocks, OLS	
	FM coef.	t-stat.	FM coef.	t-stat.	FM coef.	t-stat.
# DFDR ≤ 0.05		6		9		23
# t-stats ≥ 3.0		8		12		27
<i>agr</i>	-0.02	-0.5	-0.10	-3.0	-0.14	-4.7
<i>bm</i>	0.21	3.6	0.07	2.1	0.10	3.2
<i>mom12m</i>	0.16	1.7	0.13	1.6	0.10	1.4
<i>mve</i>	-0.25	-2.5	-0.17	-3.0	-0.67	-5.9
<i>operprof</i>	0.01	0.4	0.01	0.5	0.01	0.0
<i>roeq</i>	0.08	2.1	0.05	2.1	0.09	2.8
<i>absacc</i>	-0.01	-0.2	-0.04	-1.5	-0.03	-0.8
<i>acc</i>	-0.15	-2.9	-0.06	-2.2	-0.04	-1.5
<i>aeavol</i>	-0.00	-0.1	-0.01	-0.8	0.01	1.1
<i>age</i>	-0.05	-2.0	-0.04	-2.1	0.03	1.1
<i>baspread</i>	-0.07	-0.5	0.04	0.5	0.25	2.8
<i>beta</i>	0.03	0.2	0.00	0.0	0.07	1.1
<i>bm_ia</i>	-0.24	-1.1	-0.04	-0.5	-0.10	-0.9
<i>cash</i>	0.21	3.8	0.17	3.4	0.20	4.9
<i>cashdebt</i>	0.02	0.4	0.03	1.1	-0.01	-0.5

出处：Green, Hand and Zhang (2017)

知乎 @石川

根据 p-value 基准，在对所有股票进行 VWLS 时，一共发现了 6 个显著因子；在对非微小市值股票进行 OLS 回归时，一共发现了 9 个显著因子。将这两个回归的结果合并一共得到 12 个显著因子。与之相对应的是，在对所有股票进行 OLS 回归时，由于微小市值的影响，一共发现了 23 个显著因子。

把 94 个因子共同检验也发现了 12

▲ 赞同 51



● 5 条评论

➤ 分享

★ 收藏



知乎

首发于
川流不息

3. 分析师数量的变化 (change in the number of analysts) ;
4. 盈余公告宣告收益 (earnings announcement return) ;
5. 一个月的动量 (1-month momentum) ;
6. 六个月动量的变化 (change in 6-month momentum) ;
7. 盈利同比增长的季度数量 (number of consecutive quarters with earnings higher than the same quarter in a year ago) ;
8. 年度研发支出占市值的比重 (annual R&D to market cap) ;
9. 收益波动性 (return volatility) ;
10. 股票换手率 (share turnover) ;
11. 股票换手率的波动性 (volatility of share turnover) ;
12. 零交易的天数 (zero trading days) 。

虽然上一节的单一因子检验和本节的共同检验碰巧都发现了 12 个因子，但是仔细比较不难看出**两种方法找到的显著因子中大部分并不相同**。根据 McLean and Pontiff (2016) 对因子分类的定义，在单一因子检验中，10 个显著的因子都是基本面因子；而在共同检验时，仅有盈利同比增长的季度数量属于基本面因子，剩余更多的是交易行为因子。

上述两种检验结果为今后的 asset pricing 传递出两个重要信息：

1. 在排除了微小市值对回归造成的偏误以及修正了多重检验的数据挖掘之后，真正显著的因子个数很少；
2. 在 asset pricing 的研究中，在分析一个潜在异象时，应该选择共同检验发现的这些显著因子作为控制变量来进行研究。

5 样本外投资组合收益分析

在前两节的研究中，这些显著的因子是在整个回测期（1980 到 2014）期间发现的。它们也许对 asset pricing 有帮助，但是上述结果无法说明它们在实战中是否有交易的价值——**我们不能回到过去交易这些在事后被证明有效的因子。**

为了研究它们在交易中是否能够带来超额收益，GHZ 采用了滚动窗口使用因子选股并在样本外构建投资组合的方法。样本外滚动测试使用过去 10 年（120 个月）的数据来回归，因此从 1990 年开始。在选股时，由于同时考虑了多个因子，因此使用因子收益率来预测个股在下一期的收益率，并使用预期收益率的高低构建多空对冲的投资组合（做多预期收益率最高的 10%、做空预期收益率最低的 10%）。实证中一共考察了三种不同的投资组合（三个组合的区别是对小市值股票的权重依次提升）。

当使用全部 94 个因子进行滚动回归时，不同投资组合的收益率如下图所示。**随着小市值比重增加，投资组合的收益率逐渐增大。**

▲ 赞同 51 ▼

● 5 条评论

➤ 分享

★ 收藏

...

data) and coefficients from rolling 120-month Fama-MacBeth regressions of month $t-119$ returns on $t-120$ characteristics through month $t-1$ returns on $t-2$ characteristics.

Long top / short bottom decile hedge portfolio	Mean return	Std. dev.	t-stat.
Portfolio A: VW all stocks hedge portfolio with NYSE decile breakpoints	1.2%	5.3%	3.8
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps decile breakpoints	1.4%	5.5%	4.5
Portfolio C: EW all stocks hedge portfolio with NYSE decile breakpoints	3.1%	4.7%	11.3

出处：Green, Hand and Zhang (2017)

知乎 @石川

如果不使用全部 94 个因子，而仅仅使用 GHZ 通过共同检验找出的 12 个独立因子会怎样呢？下图给出了答案。

*Panel B: Descriptive statistics for raw monthly hedge portfolio returns, where predicted returns are based on limiting the firm characteristics to the subset of the 12 shown in Table 6. The 12 characteristics are *bm*, *cash*, *chmom*, *chnanalyst*, *ear*, *mom1m*, *nincr*, *rd_mve*, *retvol*, *std_turn*, *turn*, *zerotrade*. From these we omit *chnanalyst* because analyst data is not available early in the sample data window.*

Long top / short bottom decile hedge portfolio	Mean return	Std. dev.	t-stat.
Portfolio A: VW all stocks hedge portfolio with NYSE decile breakpoints	0.4%	5.7%	1.2
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps decile breakpoints	1.0%	5.2%	3.4
Portfolio C: EW all stocks hedge portfolio with NYSE decile breakpoints	2.4%	4.2%	9.6

出处：Green, Hand and Zhang (2017)

知乎 @石川

可以很明显的看到，仅仅使用 12 个独立因子显著降低了投资组合的收益率。这意味着我们好不容易辛辛苦苦找出了 12 个对解释 asset pricing 有效的独立因子，但是从实战选股的角度来说，这却不如使用全部 94 个因子——不管它们是不是过拟合。

如何解释这种现象呢？GHZ 这篇文章并没有对此进行过多的讨论。就我的理解，上述结果说明了两点：

- 这 12 个因子在整个回测期内显著；当使用滚动窗口时，不同的因子可能在不同历史时期显著，因此使用 94 个因子能够更好的捕捉因子风格的轮动。
- 实盘交易的目标是为了获得超额收益，而不是理解 asset pricing 的机制。对于选股模型来说，最核心的是对未来收益率的预测！

赞同 51

5 条评论

分享

收藏

...

有必要指出的是，上面的实证结果并没有考虑实际交易中的手续费等问题，且同时使用 94 个因子一定存在过拟合的问题。因此，虽然数据显示 94 个因子比 12 个因子的效果更好，这也不意味着我们在实际交易中一定能取得显著的优势。

6 2003 年前后的差异

GHZ 这篇文章研究的最后一个部分发现因子在 2003 年前后截然不同的表现 —— 因子对截面收益率差异的解释力度骤然下降。上一节提到的三个样本外投资组合的收益率在 2003 年前后的表现如下图所示，前后出现了明显的差异。这些投资组合的净值曲线在 2003 年之后变得更加平坦。

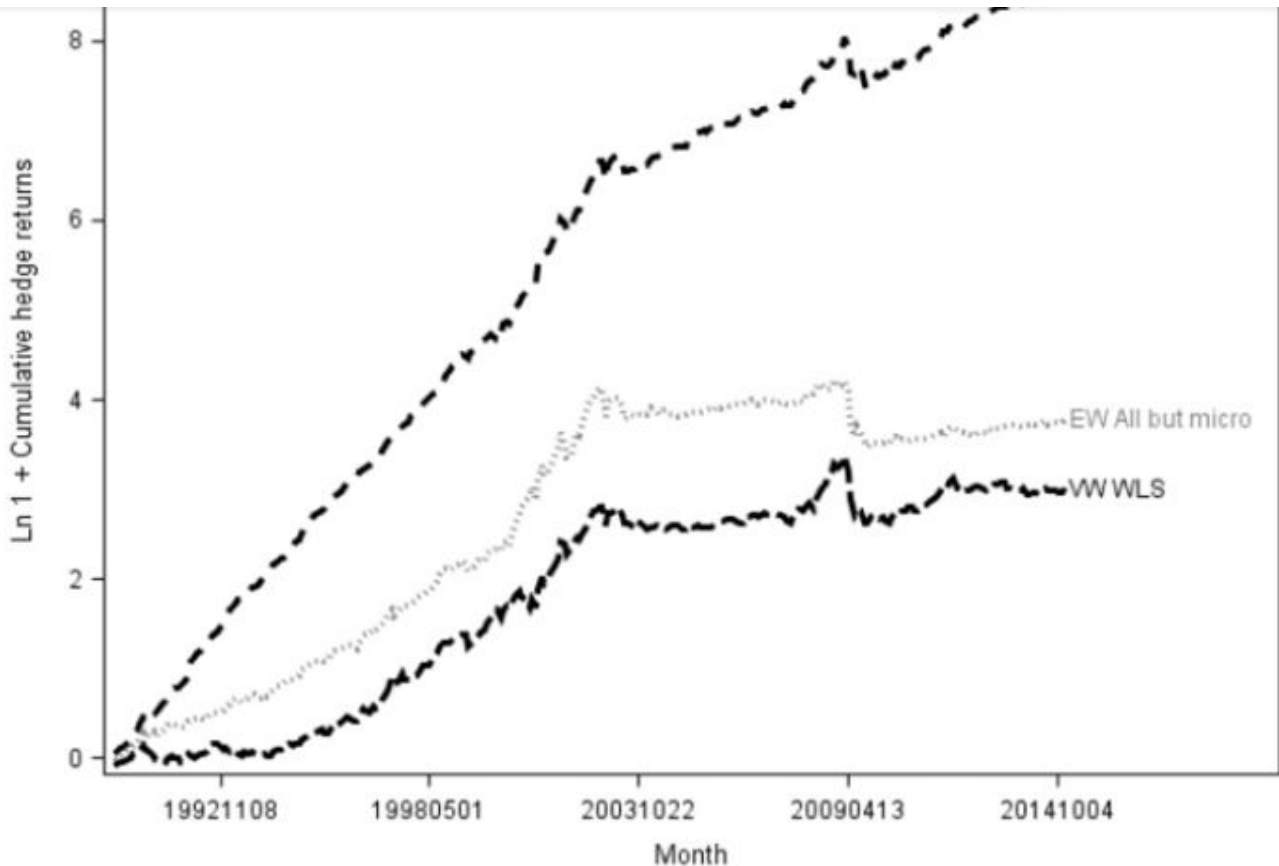
Portfolio	Pre-2003			Post-2003			Post- minus pre-2003	
	Mean return	Std. dev.	t-stat.	Mean return	Std. dev.	t-stat.	Diff.	t-stat.
Portfolio A: VW all stocks hedge portfolio with NYSE decile breakpoints	1.9%	5.5%	4.4	0.5%	4.9%	1.1	-1.4%	-2.3
Portfolio B: EW all-but-microcaps hedge portfolio with all-but-microcaps breakpoints	2.8%	6.0%	5.7	0.1%	4.4%	0.2	-2.7%	-4.4
Portfolio C: EW all stocks hedge portfolio with NYSE decile breakpoints	4.4%	4.4%	12.5	1.7%	4.4%	4.4	-2.8%	-5.2

出处：Green, Hand and Zhang (2017)

知乎 @石川



知乎

首发于
川流不息

出处: Green, Hand and Zhang (2017)

知乎 @石川

在 2003 年之后，仅仅有 2 个因子仍然显著，它们分别为盈利同比增长的季度数量以及经行业调整后的员工数。究其原因，**GHZ 认为在 2003 年之后，美股上的套利成本大幅降低造成了这种现象。**

2002 年 7 月，美国股市通过了 Sarbanes-Oxley 法案；同年 10 月，SEC 加速了对上市公司披露 10-Q 和 10-K 文件的要求。与此同时，在 2003 年初（1 月到 5 月），纽交所引入自动报价系统，减少了交易的摩擦成本。**这一系列变化极大降低了交易股票市场异象的量化对冲策略的成本和技术难度。**无疑，2003 年前后的差异对于今后研究提出了挑战。GHZ 也呼吁今后 asset pricing 的研究应该给予 2003 年之后的数据更多的权重。

7 结语

研究因子时最艰难的部分往往是数据的准备。一旦有了高质量的数据，各种用于检验的回归及其他统计手段是非常丰富的。作为一篇同时研究 94 个因子的文章，GHZ 的一个巨大贡献在于因子数据的计算，以及为了计算因子而对公司基本面和交易数据的收集、整理。此外，GHZ 在截面回归中用个股的收益率和这 94 个因子进行回归，可以想见计算量之巨大。这和 Harvey and Liu (2018) 中给出的简单例子完全不同。

简单总结一下这篇文章给多因子投资

▲ 赞同 51 ▼

● 5 条评论

➤ 分享

★ 收藏

...

知乎

首发于
川流不息

解释；

3. 共同检验和单一检验发现的显著因子往往不同，前者的发现更有效；
4. 因子投资和资产定价的目标不同；前者是为了提高样本外预测的准确性，后者是为了确保因子收益率的无偏估计；在投资实务中，更应该注重前者，但也要考虑实际的交易限制和成本。

GHZ 通过同时回归找到的 12 个因子虽然是对美股，但很多在 A 股上也是有效的。此外，GHZ 使用多个因子同时进行 Fama-MacBeth regression 的做法和 Barra 如出一辙。今后的研究会使用 Barra 中国股票模型中的风格因子进行类似的实证，检验不同因子在 A 股上的显著性，希望以此更好的指导我们在 A 股上实践因子投资。

参考文献

- Benjamini, Y. and D. Yekutieli (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *Annals of Statistics*, Vol. 29, 1165 – 1188.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, Vol. 52(1), 57 – 82.
- Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, Vol. 66(4), 1047 – 1108.
- Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, Vol. 33(1), 3 – 56.
- Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics*, Vol. 116(1), 1 – 22.
- Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, Vol. 81(3), 607 – 636.
- Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, Vol. 29(1), 5 – 68.
- Harvey, C. R. and Y. Liu (2018). Lucky Factors. Working paper, available at SSRN: ssrn.com/abstract=25287...
- Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: an investment approach. *Review of Financial Studies*, Vol. 28(3), 650 – 705.
- Green, J., J. R. M. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies*, Vol. 30(12), 4389 – 4436.
- McLean, R.D., J. Pontiff. (2016). Does academic research destroy stock return predictability? *Journal of Finance* 71(1), 5–32.

免责声明：文章内容不可视为投资意见。市场有风险，入市需谨慎。



▲ 赞同 51 ▼

● 5 条评论

➤ 分享

★ 收藏



编辑于 2019-07-03

资产定价 多因子模型 回归分析

文章被以下专栏收录



川流不息

北京量信投资管理有限公司是一家在中国基金业协会备案登记的专业私募基金管理人...

关注专栏

推荐阅读



有没有哪个趋势指标更好使？

石川



不赚钱，发个锤子文章

joea....

发表于量化哥



7种回归分
必须掌握！

DataH...

5 条评论

切换为时间排序

写下你的评论...



Walker Sleep

5 个月前

您好，看完感觉一言难尽。。
么搞很复杂的因子检测逻辑在

赞同 51



5 条评论

分享

收藏



知乎

首发于
川流不息

👍 赞



石川 (作者) 回复 Walker Sleep

5 个月前

回归之前，尽量从业务逻辑上选择相关性低的因子，必要的话可以做一定的正交化处理；如果线性相关的话，回归得到的参数的 standard error 会不准。另外对机器学习，冷静是必要的，但也不用特别悲观。还是要看模型要干什么，是要做 prediction 还是做 estimation。从选股来说，这些模型都是为了 prediction，目标是 loss function minimization，而不是 parameter adjudication。用人话说就是，只要预测的好，那么因子之间即便很相关也 ok，因为它的目的不是为了无偏的估计这些因子的收益率（参数）。前 Harvard 的 Sendhil Mullainathan（现在去了 UChicago）曾在 AFA 做了一个题为 Machine Learning and Prediction in Economics and Finance，值得一看。

👍 赞



Walker Sleep 回复 石川 (作者)

5 个月前

谢谢解答，要是从 prediction 出发的话，还是挺打击我的。毕竟我刚学多因子模型，都是在那测因子 T 值再筛选，这个我还得消化下。。。

另外想请教下，一般机器学习模型更多的是采用面板数据一起进行训练；而传统的线性模型很多是横截面上先做 Fama-MacBenth 回归，然后再考虑时序的变化情况，这种还能计算一些标准差什么的。不知道有没有文献讨论这种区别？

👍 赞



石川 (作者) 回复 Walker Sleep

5 个月前

石川：[股票多因子模型的回归检验](#)

👍 赞



Walker Sleep 回复 石川 (作者)

5 个月前

谢谢，这个之前有拜读过，应该还是关于线性模型的。本意是想问下机器学习模型怎么很少见讨论面板数据与横截面数据训练的，不知是不是自己读的 paper 太少。或者说机器学习模型比如说深度学习，参数众多意义也不明显，这种情况下再去讨论参数在时序上的均值或者标准差是不是没有意义？

👍 赞



▲ 赞同 51 ▼

● 5 条评论

➦ 分享

★ 收藏

...