

首发于 **川流不息** 



### 从 Factor Zoo 到 Factor War, 实证资产定价走向何方?



石川 😘

量化交易 话题的优秀回答者

已关注

量化投资机器学习等83人赞同了该文章

#### 摘要

近年来,实证资产定价研究已经从 factor zoo 逐渐演变成 factor war,然而在这种愈演愈烈的"竞争"下,我们对市场的理解又增加了多少?

### 1引言

这一切都源于 Hou et al. (2019a) 怼了 Fama and French (2015) 五因子模型。

近年来,侯恪惟、薛辰以及张橹三位教授在实证资产定价领域掀起了一股华人旋风。如果你对这三个中文名字没什么印象,那么对他们 Last Names 的首字母组成的三个字母缩写 —— HXZ —— 一定不陌生。

几年前,这三位教授在一篇题为 Replicating Anomalies 的文章中惊人的复现了学术界的 447 选股异象(anomalies),并指出在排除微小市值公司、考虑了 multiple testing 以及使用了"适当"的定价模型之后,绝大部分异象都不再显著。这篇最初的 working paper 长达 146 页,

最长之一。

### **Replicating Anomalies**

#### Kewei Hou

The Ohio State University and China Academy of Financial Research

#### Chen Xue

University of Cincinnati

#### Lu Zhang

The Ohio State University and National Bureau of Economic Research

Most anomalies fail to hold up to currently acceptable standards for empirical finance. With microcaps mitigated via NYSE breakpoints and value-weighted returns, 65% of the 452 anomalies in our extensive data library, including 96% of the trading frictions category, cannot clear the *single* test hurdle of the absolute *t*-value of 1.96. Imposing the higher multiple test hurdle of 2.78 at the 5% significance level raises the failure rate to 82%. Even for replicated anomalies, their economic magnitudes are much smaller than originally reported. In all, capital markets are more efficient than previously recognized G12, G14, G17, M41)

HXZ 对于实证资产定价的贡献远不止于上面这篇 Replicating Anomalies。在资产定价领域有着不同的方法论,比如人们熟悉的 consumption-based CAPM、以 Eugene Fama 和 Ken French为代表的 empirical multi-factor asset pricing、以及最近流行的基于 behavioral finance 的 asset pricing(Stambaugh and Yuan 2017; Daniel, Hirshleifer, and Sun 2019)。而在方法论上,HXZ 三位倡导的则是 Investment-based CAPM。

在这方面,HXZ 的代表作是 Hou, Xue, and Zhang (2015) 这篇同样发表于 Review of Financial Studies 上的文章,题为 Digesting anomalies: an investment approach。他们从公司投资视角 出发提出 q-factor asset pricing model(学术界简称 q-factor model)。以 The first principle of investment 为依据,q-factor model 在市场和 SMB 的基础上加入了 profitability 和 investment 两个维度,构建了四因子模型。

而张橹教授也在其个人主页上强调了 investment CAPM 对于资产定价理论的贡献:

My unique, big-picture perspective of asset pricing, which differs drastically from both the consumption CAPM and behavioral finance, is elaborated in my article titled The investment CAPM published in 2017 at European Financial Management (Zhang 2017).



# 知乎 / imana

estimating expected stock returns 为题撰文为其造势。截至今日,Hou, Xue, and Zhang (2015) 已被引用超过 890 次,是 2015 年以来发表于 RFS 上的所有文章中引用量最高的一篇。

显然,这三位教授在 Investment CAPM 的道路上远没有止步于此。最近两年,沿着 q-factor 的 思路,他们(以及合作者 Haitao Mo)可谓相当高产,最新论文包括:

- 2018 年提出了 q-factor model 的进阶版: q5 model (Hou et al. 2018)。它在 q-factor model 的基础上增加了第五个因子 —— 预期投资增长因子;
- 2019 年,在 Review of Finance 上发表了 Which Factors 一文,使用 q 和 q5 模型和其他主流的多因子模型进行了全方位的比较(Hou et al. 2019a);
- 同样是 2019 年,在一篇刚刚出炉、还冒着热气的 working paper 中(Hou et al. 2019b),这四位使用 q5 模型解释了历史上的多位大佬(比如巴菲特和神奇公式发明者 Greenblatt)以及学术界的一些重磅因子(比如 AQR 的 QMJ 和 Piotroski 的 F-Score)的收益;这些大佬和因子的共同之处都是源于 Graham 和 Dodd 的 Security Analysis。

由于 q-factor model 和 Fama and French (2015) 五因子模型中都包括投资因子,Hou et al. (2019a) 这篇文章在比较多个因子模型之外,还用了相当多的笔墨怼了 Fama and French (2015) 中的投资因子。

终于和本文开篇第一句呼应上了。

当我第一次看到 Hou et al. (2019a) 对 Fama and French (2015) 的抨击时,着实懵逼了好长时间 —— 虽然出发点不同,但两个因子模型都有投资因子、且指标的选取一模一样(都是过去一个财年 total assets 的增长率);为什么 Hou et al. (2019a) 要怼 Fama and French (2015)? 为了搞清楚这个问题,我又系统的沿着 Hou, Xue, and Zhang (2015)、Hou et al. (2018, 2019a, 2019b) 一路读下来,加深了对 investment CAPM 的理解。不过在这个过程中,也多少感到了近年来学术界因子模型之争(factor model war,简称 factor war)的硝烟。

本文就以梳理 investment CAPM 的发展脉络为契机,映射学术界最近在 factor war 上面愈演愈烈的趋势,以此表达一些对实质资产定价研究的担忧。最后需要强调的是,在多家因子流派"干帆竞发、百舸争流"之下,factor war 是一个普遍的现象;因此本文虽然是透过 investment CAPM 的镜头一窥这一现象,并不意味着该流派是 factor war 的众矢之的。

下面先来看 q-factor model。

### 2 q-Factor Model 的经济学原理

Hou, Xue, and Zhang (2015) 提出的 q-factor model 是受到 Cochrane (1991) 的 producti based asset pricing 所启发,从公司投资的经济学原理出发。张橹教授曾在清华金融评论上撰立

论文故事汇 Essays and Papers

101

# 资产定价中的因子大战

本文介绍了笔者与俄亥俄州立大学侯恪惟教授、辛辛那提大学薛辰教授共同研究开发的q-因子模型。在该模型中,一只股票的预期风险溢价由四个因子决定:市场因子、市值因子、投资因子和盈利因子。论文正式发表后,q-因子模型已迅速成为金融界实证资产定价中的主导模型之一。展望未来,q-因子模型将拓展到全球金融市场中,比如中国A股市场。

知乎 @石川

根据 Zhang (2016), q-factor model 是基于实体投资经济学理论,又称 q-theory (这也是该模型被称为 q-factor model 的原因)。它体现了公司金融学中的净现值原则 (NPV rule):如果项目现值大于投资成本,则应当投资该项目;如果项目现值小于投资成本,则不应当投资。起初,公司会有很多投资项目,此时应优先投资折现率低、盈利率高、因此现值最高的项目。随着被投资的项目越来越多,投资成本会慢慢变高,盈利率会越来越低。投资的最后一个项目应该是净现值为零(净现值原则):投资成本 = 项目现值 = 盈利率 / 折现率。

上面这个净现值原则说的是: 一家公司应该持续投资,直到投资的边际效益(被折现到今天)等于投资的边际成本。这就是对 Hou, Xue, and Zhang (2015) 所依托的经济学原理的核心概括。 Zhang (2016)继续介绍说: **q-factor model 最富想象力和创造力的是把公司金融原则当作资产定价模型。**传统资产定价理论从投资者最优证券组合角度出发,和公司变量没有直接关系。但是那一老套做了近半个世纪,结果只是大量异象现象。**q-factor model** 开辟了一个新的途径。

我对以上这段如此之高的自我评价持保留态度。

将净现值原则变化一下得出: **折现率 = 盈利率 / 投资成本**。从这个式子出发,我们可以得到两个关于折现率的条件预期结论: 当盈利率给定时,投资越多的公司(因此投资成本越高),折现率越低,股票预期收益率也越低; 当投资给定时,盈利率越高的公司,折现率越高,股票预期收益率也越高。由此可知,**股票收益率和投资成反比; 和盈利率成正比**。这就是 q-factor model 中加入 profitability 和 investment 两个维度的原因。

不过我想"杠精"一下。尽管提出者对 q-factor model 有着很高的自我评价;但我们却无法从净现值原则中推导出市场和 SMB 两个因子。为什么最后 q-factor model 是四个因子而非两个听退一步说,加入市场因子无可厚非,但为什么 q-factor model 中不是三因子,而是偏偏加入了SMB 凑出四个因子呢?

中,仅包含市场、盈利以及投资三个因子。这篇文章后被投到 Journal of Finance 却在几经修改之后还是被 rej 了。之后,Zhang 教授联手 Hou 和 Xue,最终提出了我们看到的四因子版本的 q-factor model,它经过两年的审稿和修改,终于在 2015 年被发表在 RFS 上。

反观 Fama and French (2015),他们从 discount cash flow (DCF) 出发 promote 五因子模型 (下文称 FF5) 中新加入的 profitability 和 investment 两个因子; 而如果从 DCF 的表达式来看,也是可以推出收益率和 SMB 及 HML 之间的关联(后文会说明)。从这个意义上说,FF5 更自洽一些。

那么,Hou, Xue, and Zhang (2015) 用了哪些指标作为盈利率和投资的 proxy 来构建因子呢?为此下面一小节将简单介绍 q-factor model 的数学模型。从之前的中文描述中我们已经搞清楚了该模型背后的核心经济学原理。因此在介绍数学表达时会力求言简意赅。了解 q-factor model 的数学模型也会帮助我们在本文第四节更好的理解  $q^5$  model ——它在数学上是 q-factor model 的一步拓展。

### 3 q-Factor Model 的数学模型

q-factor model 背后的数学模型十分简单;它假设一个两期(date 0 和 date 1)的公司投资决策模型。假设在 date 0,某公司 i 的资产为  $A_{i0}$  、利润率是  $\Pi_{i0}$  (假设已知)。在 date 1,该公司的利润率是  $\Pi_{i1}$  ,它是一个随机变量。对于 date 0 和 1,公司的 cash flow 为每一时刻的资产乘以利润率:  $\Pi_{it}A_{it}, t=0,1$  。

在这个两期模型中,公司的决策变量是 date 0 的投资额  $I_{i0}$  。该模型假设公司 date 0 的资产 $A_{i0}$  在 t=1 时全部折旧完,因此该公司在 date 1 的资产正是其在 date 0 的投资: $A_{i1}=I_{i0}$  。除此之外,伴随  $I_{i0}$  而来的还包括一个调整费用(adjustment cost): $(a/2)(I_{i0}/A_{i0})^2A_{i0}$  。

有了上述设定,我们就来看看最优化的目标函数是什么。

对于该公司,date 0 对于股东的回报是 cash flow 减去投资额以及投资的调整费用,即  $\Pi_{i0}A_{i0}$   $-I_{i0}-(a/2)(I_{i0}/A_{i0})^2A_{i0}$  ; 而对于 date 1,由于不再有投资,因此其在 0 时刻的预期回报可以表达为  $\mathbf{E}_0[M_1\Pi_{i1}A_{i1}]$  ,其中  $M_1$  是 stochastic discount factor、  $\Pi_{i1}A_{i1}$  是 date 1 的 cash flow、而求期望符号  $\mathbf{E}_0$  的下标 0 代表 date 0 时的预期。这两期的回报相加就是最优化的目标函数:

$$\max_{I_{i0}}\Pi_{i0}A_{i0} - I_{i0} - rac{a}{2}igg(rac{I_{i0}}{A_{i0}}igg)^2A_{i0} + \mathrm{E}_0[M_1\Pi_{i1}A_{i1}]$$

该目标函数的最优解  $I_{i0}$  应满足:



上式的左侧为投资的边际成本(第二项为 marginal adjustment cost);右边为边际效益预期被 折现到 date 0。这个关系是也说明了上一小节提到的净现值原则。此外,由 the first principle of investment 可知,投资收益率  $\mathbf{r}^{I}$  (上标 I 表示投资收益率;为了区分股票收益率)应满足:

$$\mathrm{E}_0[M_1r_{i1}^I]=1$$

比较上面两个式子就可以得到投资收益率和边际成本以及边际效益的关系:

$$\mathrm{E}_{0}[r_{i1}^{I}] = rac{\mathrm{E}_{0}[\Pi_{i1}]}{1 + a(I_{i0}/A_{i0})}$$

根据 Cochrane (1991) 和 Liu, Whited, and Zhang (2009) 的研究,在上述模型(以及更general 的模型)设定下,公司的投资收益率等于其股票收益率(记为  $r^S$ ),因此最终可以把股票收益率和上述从公司投资角度出发的经济学理论联系起来:

$$\mathrm{E}_{0}[r_{i1}^{S}] = rac{\mathrm{E}_{0}[\Pi_{i1}]}{1 + a(I_{i0}/A_{i0})}$$

上式就和本文第二节的文字解释一致了: **当预期盈利**  $\mathbf{E}_0[\Pi_{i1}]$  **给定时,股票收益率和投资**  $I_{i0}/A_{i0}$  成反比; 当投资给定时,股票收益率和预期盈利成正比。

注意,在上面这个关系式中,投资  $I_{i0}/A_{i0}$  是 date 0 的,因此是历史投资,而非未来的预期投资 (expected investment) 。 **q-factor model 背后的理论用的是 past investment,而非expected investment,这是它和 FF5 在投资因子上最大的差异 (FF5 从 DCF 模型出发promote 出来的 expected investment)** 。之所以强调这个,因为它是搞清楚 Hou et al. (2019a) 怼 FF5 的关键。

在实证研究中,Hou, Xue, and Zhang (2015) 使用 ROE 和 Total Assets 的增长率作为代表盈利和投资的指标。在构建因子时,为了体现上述**条件**预期收益率的关系,他们特意选择了通过 size、I/A 以及 ROE 将股票池独立进行 2 × 3 × 3 的 triple sort(使用 ROE 和 I/A 将股票分成三组时使用学术界常用的 30% 和 70% 分位数,中间的 40% 为 Middle 组),一共得到 18 个投资组合(每个投资组合中的股票都是按市值加权)。最终三个风格因子的定义为:

- SMB: 9 个 small size 组合的简单平均收益率与 9 个 large size 组合的简单平均收益率之差;
- I/A: 6 个 low investment 组合的简单平均收益率与 6 个 high investment 组合的简单平均收益率之差;
- ROE: 6 个 high ROE 组的简单平均收益率与 6 个 low ROE 组的简单平均收益率之差。



# 知乎 / imake

以上三个因子,再结合市场因子,就构成了 q-factor model。至于它和其他主流因子模型相比效果如何?我们留在本文的第七节 factor war 见分晓。

有了对 q-factor model 数学模型的理解,将它进行一步扩展,就得到  $q^5$  。

### 4 $q^5$

Hou et al. (2018) 这篇文章虽然目前仍然只是 working paper,但  $m{q}^{m{5}}$  这个题目绝对标新立异。

 $m{q^5}$  在 q-factor model 的四因子基础上加入了第五个代表 expected investment growth 的因子,得到一个五因子模型。在数学上,它是 q-factor model 的直接延伸。而我更关注的是这第五个因子被加入的动机。

Hou et al. (2018) 一开篇就以 Cochrane (1991) 来 motivate 出这第五个因子:

Cochrane (1991) shows that in a multiperiod investment framework, firms with **high expected investment growth should earn higher expected returns** than firms with low expected investment growth, holding current investment and expected profitability constant.

显然,这句话引入了 **expected** investment growth,同时它也指出了这是在 multiperiod investment framework 之下。由于 Hou, Xue, and Zhang (2015) 仅使用了简单的二期模型(因为 date 1 不再投资,因此它实际上是一个 one period investment framework),所以 q-factor model 中是没有投资增长的。因此,在  $q^5$  这篇文章中,Hou et al. (2018) 把投资模型扩展到了多期。

上面这个理论上的延展看来是把 q 升级到  $q^5$  背后的主要动机。不过有意思的是,Zhang 教授和另外两位 co-authors 早在 Liu, Whited, and Zhang (2009) 一文(发表于经济学顶刊 Journal of Political Economy)中就研究过多期投资模型。虽然该文的重点不是 empirical asset pricing,但 expected investment growth 这个维度早就出现过(Liu, Whited, and Zhang 2009, pp 1126, eq. 10)。所以……

我无从得知为什么 Hou, Xue, and Zhang (2015) 在 q-factor model 里没有加入 expected investment growth; 而是等到 q5 这篇才加入。也许在 q-factor model 被提出时,它已经能解释很多当时已有因子不能解释的异象了。而这几年一些新的因子模型有在效果上超过了 q-factor model,所以需要一个升级版。这当然纯属我的"阴谋论",不过 Hou et al. (2018) 中的下面 句话也十分耐人寻味:

#### 首发于 川流不息

OK,八卦结束,让我们从数学上看看这个  $q^5$  模型。

Hou et al. (2018) 在 Hou, Xue, and Zhang (2015) 的基础上考虑无限期的模型。由于是多期模 型,因此他们假设资产从 t 到 t+1 期的折旧率为  $\delta$  (在 q-factor model 中,资产从 date 0 到 date 1 折旧完,因此在该模型中  $\delta = 1$  )。投资决策是每一期的投资  $\{I_{it+s}\}, s=0,1,\ldots$ 。在任何时点 t ,目标函数是最大化股东权益的现值。根据 the first principle of investment 可知从 t 到 t+1 期的边际投资成本应该等于边际效益折现到 t , 因 此 t 到 t+1 的投资收益满足:

$$r_{it+1}^{I} \equiv rac{\Pi_{it+1} + (a/2)(I_{it+1}/A_{it+1})^2 + (1-\delta)[1 + a(I_{it+1}/A_{it+1})]}{1 + a(I_{it}/A_{it})}$$

对于 Hou, Xue, and Zhang (2015) 考虑的两期模型,由于折旧率  $\delta = 1$  且 date 1 的投资  $I_{i1}=0$  ,因此上式分子中仅有第一项,而上式也就变成了本文第三节介绍的 q-factor model 的情况。

对于多期模型, Hou et al. (2018) 如法炮制, 指出股票收益率等于投资收益率。由此, 从上式我 们可以找到股票收益率和投资以及盈利的关系。上式中第一项近似对应 "dividends" ,第二项平 方项与边际调整费用有关(由于是高阶小量故忽略),第三项近似对应 "capital gain" ,即 expected investment-to-asset growth (expected I/A growth) .

从多期模型出发,股票收益率除了和 q-factor model 中已经考虑的 ROE 以及 I/A (条件) 相 关,**也和 expected I/A growth 正相关**。因此,该因子作为第五个因子被补充到 q-factor model 中,最终得到了  $q^5$  。

在实际构造 expected I/A growth 因子时, Hou et al. (2018) 使用了 predictive regression 外 推的方法。即这个因子和我们了解的直接利用历史财务数据的因子均不同,它是使用历史数据构建 了一个截面回归预测模型,然后使用最新的财务数据预测下一期 expected I/A growth。从数学复 杂度上来说,我个人对这种构建因子的方法不是很感冒。**因为这里面有模型设定以及参数估计误差** 等太多的问题。

Hou et al. (2018) 自然意识到了这些问题;在文中他们也特别提到:

Our goal is a conceptually motivated yet empirically validated specification for the expected investment-to-assets changes. To this end, we turn to the investment literature in macroeconomics and corporate finance for guidance.

然而不可否认的是,在构造 expected I/A growth 因子时,由于可挖的数据太多,因此 data snooping 的风险也很高。本文不再赘述该因子的具体构造方法,感兴趣的小伙伴可以查看 Hc 🛖



# 川流不息

最后想要强调的是,我在上面提及这第五个因子时,不厌其烦的使用了 expected 这个词。这是因 为从 q-theory 出发,该因子是(未来)预期投资增长率,而非过去的投资增长率。是否带有 expected 一词,含义截然不同。

#### **5 FF5**

我们来说说 FF5 吧;它被 q-factor model 当成靶子打。

从 DCF 出发并利用 Miller and Modigliani (1961) 的结果可知  $m{t}$  时刻公司市值  $m{M}_t$  满足如下关 系:

$$M_t = \sum_{ au=1}^\infty \mathrm{E}[Y_{t+ au} - dB_{t+ au}]/(1+r)^ au$$

上式中,  $Y_{t+\tau}$  是  $t+\tau$  期的 total earnings,  $dB_{t+\tau}$  是未来  $t+\tau$  期相对前一期的 total book value 的变化,  $m{r}$  是 IRR。将上式两边同时处以  $m{t}$  时刻公司的 book value  $m{B_t}$  可得:

$$rac{M_t}{B_t} = rac{\sum_{ au=1}^\infty \mathrm{E}[Y_{t+ au} - dB_{t+ au}]/(1+r)^ au}{B_t}$$

Fama and French (2015) 通过上式 motivate 出 FF5:

- 1. 当除了  $M_t$  和 r 的其他变量给定时,更小的  $M_t$  (或更高的  $B_t/M_t$  )意味着更高的 r—— 这推出了 Small-Minus-Big (SMB) 和 High-Minus-Low (HML);
- 2. 当除了  $Y_{t+ au}$  和 r 的其他变量给定时,更高的  $Y_{t+ au}$  对应着更高的 r —— 即**预期盈利和** r**正相关**,推出了 Robust-Minus-Weak (RMW);
- 3. 当除了  $dB_{t+ au}$  和 r 的其他变量给定时,更高的  $dB_{t+ au}$  对应着更低的 r —— 即**预期投资 和 r负相关 (这就是被怼的靶子)** ,从而推出 Conservative-Minus-Aggressive (CMA) 。

在实证中如何围绕预期盈利和预期投资构建因子呢?这方面的研究可以追溯到 Fama and French  $(2006)_{\circ}$ 

2006年, Fama 和 French 在 Journal of Financial Economics 上发表了一篇题为 Profitability, investment and average returns 的文章,专门检验了预期盈利和预期投资这两个维度和 expected return 的关系:

Given B/M and expected profitability, higher expected rates of investment imply lov expected returns. But controlling for the other two variables, more profitable firms





与 FF5 一样,Fama and French (2006) 也是从 DCF 出发。它在检验预期盈利和预期投资时,既检验了使用历史数据作为预期的 naïve estimate,又考虑了使用 predictive regression 构建因子。

以预期投资为例,在使用历史数据直接外推时,该文使用 dA/A —— total asset 的变化率 —— 作为 expected investment。截面回归结果(下图)显示出它和收益率之间显著的负相关性。

Table 3

Monthly cross-section return regressions

The table shows average slopes and their Fama-MacBeth t-statistics from monthly cross-section regressions to predict stock returns.  $Y_t$ ,  $D_t$ , and  $AC_t$  are earnings, dividends, and accruals per share for the fiscal year ending in calendar year t.  $-AC_t$  is accruals for firms with negative accruals (zero otherwise) and  $+AC_t$  is accruals for firms with positive accruals.  $B_t$ ,  $A_t$ , and  $M_t$  are book equity, total assets, and stock price per share at the end of fiscal year t.  $MC_t$  is market capitalization (price times shares outstanding) at the end of June of year t+1.  $I_t$  is the  $1/B|E_t$ S consensus forecast of earnings for the coming year, sampled at the end of fiscal year t.  $I_t$  is the  $1/B|E_t$ S consensus forecast of earnings for the coming year, sampled at the end of fiscal year t.  $I_t$  is the probability of default on debt, estimated at the end of fiscal year t, and  $2-3Y_t$ , is the two-year return for the years up to the end of fiscal year t-1.  $OH_t$  is the probability of default on debt, estimated at the end of fiscal year t, from the logit regression model of Ohlson (1980).  $PT_t$  is the Piotroski (2000) composite index of firm strength.  $Neg Y_t$  is a dummy variable that is one for firms that have negative earnings for fiscal year t (zero otherwise), and  $No D_t$  is a dummy variable that is one for firms that pay no dividends during fiscal year t.  $P(Y_{t+t}/B_t)$  and  $P(I_{t+t}/A_t)$ , expected profitability and asset growth, are fitted values from the first pass regressions in Table 2 that include lagged fundamentals (in  $B_t/M_t$ , in  $MC_t$ ,  $Neg Y_t$ ,  $Y_t/B_t$ ,  $AC_t/B_t$ ,  $AC_t/B_t$ ,  $AC_t/B_t$ ,  $AC_t/B_t$ , and  $D_t/B_t$ ) and lagged fundamentals, lagged returns,  $I_t/B_t$ ,  $OH_t$ , and  $PT_t$ . Firms is the average number of firms in the regressions. The regressions are estimated monthly, beginning in July of 1963, using explanatory variables that are updated at the end of each June. The accounting explanatory variables in the regression for July of year t+1

	Firms	Int	$\ln B_t/M_t$	$\ln MC_t$	Neg Y,	$Y_t/B_t$	$-AC_i/B_i$	$+AC_i/B_i$	$dA_i/A_j$	OH,	$PT_t$	$\mathbb{R}^2$
Part A: Re	gressions u	se lagged pr	ofitability, asse	t growth, ac	cruals, OH,	and PT,						
Average	2,058	1.66	0.28	-0.06								0.02
-statistics		3.85	2.97	-1.20								
Average	2,058	1.69	0.28	-0.08	-0.00	1.10			-0.40			0.03
-statistics		4.36	2.74	-1.83	-0.02	2.55			-3.87			
Average	2,058	1.83	0.26	-0.10	-0.00	1.38	-0.24	-1.42	-0.19			0.04
-statistics		4.93	2.61	-2.37	-0.03	3.21	-0.80	-6.82	-1.99			
Average	2,253	1.35	0.34	-0.07						-0.04	0.06	0.03
-statistics		2.43	3.02	-1.33						-2.25	2.58	
Average	2,253	1.51	0.34	-0.09	0.11	1.51	-0.09	-1.32	-0.25	-0.03	0.04	0.03
-statistics		3.44	2.89	-1.76	0.77	3.40	-0.41	-5.42	-2.44	-1.55	7.53	@F111

出处: Fama and French (2006)

而当 Fama and French (2006) 采用 predictive regression 计算 expected investment 时,却并没有获得统计上显著的关系(下图中  $F(dA_{t+ au}/A_t)$ ):









		mated with lagged fun	aamentais				
Average slopes						578966	
1	2,058	1.61	0.37	-0.08	0.04	1.58	0.03
2	2,058	1.11	0.48	-0.08	0.39	2.05	0.03
3	2,058	0.97	0.53	-0.07	0.42	2.04	0.03
t-statistics							
1		1.86	3.42	-1.92	0.05	2.03	
2		1.92	3.87	-1.76	0.87	2.37	
3		1.89	4.14	-1.56	1.34	2.40	
Expected profit	tability and arowth estin	mated with lagacd fun	damentals, lagged ret	urns, L/B, OH, and I	PT.		
	tability and growth estin	mated with lagged fun	damentals, lagged ret	urns, $I_t/B_t$ , $OH_t$ , and $I$	PT,		
	i			CONTRACTOR OF CO		1 27	0.02
	1,530	2.06	0.20	-0.09	-0.20	1.27	0.03
	1,530 1,530	2.06 1.75	0.20 0.25	-0.09 -0.10	-0.20 0.08	1.58	0.03
	1,530	2.06	0.20	-0.09	-0.20	15007422.0	
Average slopes 1 2 3	1,530 1,530	2.06 1.75	0.20 0.25	-0.09 -0.10	-0.20 0.08	1.58	0.03
Expected profit Average slopes 1 2 3 t-statistics 1	1,530 1,530	2.06 1.75	0.20 0.25	-0.09 -0.10	-0.20 0.08	1.58	0.03
Average slopes 1 2 3	1,530 1,530	2.06 1.75 1.57	0.20 0.25 0.28	-0.09 -0.10 -0.09	-0.20 0.08 0.21	1.58 1.43	0.03

出处: Fama and French (2006)

上面的结果显示,**预测的预期投资 F(dA\_{t+ au}/A\_t) 与收益率的相关系数为正 (而这和从 DCF 推** 出的 expected investment 和收益率之间的负相关性是相左的!),但统计上非常不显著。以预 测未来一年的  $F(dA_{t+1}/A_t)$  为例,它的回归系数仅为 0.04,t-statistic 为 0.05。

对于这两种方法的差异,Fama and French (2006) 从计量经济学的角度进行了大量的探讨,并最 终选择了使用历史数据进行 naïve estimate 的方法计算预期投资 (对于预期盈利,使用历史数据 和 predictive regression 的差异很小)。上述处理方法也最终被保留到了 FF5 中。

值得一提的是,按照上述 DCF 模型,代表投资的变量应该是  $F(dB_{t+ au}/B_t)$  ,即 book value 的预期变化,而非 total asset。Fama and French (2015) 坦言他们比较了使用 book value 和 total asset 两种方法,发现使用后者排序的话,股票收益率在截面上的差异更大,因此选择了 total asset 的变化。

以上就是 FF5 背后的来龙去脉。下面就来看看 Hou et al. (2019a) 是怎么怼 FF5 的。

### 6 q vs FF5

Hou et al. (2019a) 这篇文章使用 spanning test 比较了他们提出的 q 和  $q^5$  以及其他几个主流 的多因子模型,包括:

- Fama and French (2015) 五因子模型;
- Fama and French (2018) 六因子模型 (在 FF5 的基础上加入了动量);
- Stambaugh and Yuan (2017) 四因子模型;
- Barillas and Shanken (2018) 六因子模型;
- Daniel, Hirshleifer, and Sun (2019) 三因子模型。

结论自然不出意外,q 和  $q^5$  完胜。客观的说,这篇文章最大的贡献在于复现上述模型以及 empirical test 的结果,其他创新稍显不足。这大概是它仅发表于 Review of Finance,而非传 🔺 意义上的顶刊的原因。



下面开始解惑之旅。

如果我们仔细查看 q-factor model 和 FF5 中构建投资因子的指标,会发现这两个指标完全一致 (下图 1 来自 Hou, Xue, and Zhang 2015; 图 2 来自 Fama and French 2015)。

#### 2.1 Factor construction

We measure investment-to-assets, I/A, as the annual change in total assets (Compustat annual item AT) divided by 1-year-lagged total assets. We measure profitability as ROE, which is income before extraordinary items (Compustat quarterly item IBQ) divided by 1-quarter-lagged book equity. We construct the *q*-factors from a triple 2-by-3-by-3 sort on size, I/A, and ROE. Because 出处: Hou, Xue, and Zhang (2015), pp 659

Panel C of Table 1 shows average excess returns for 25 Size-Inv portfolios again formed in the same way as the 25 Size-B/M portfolios, but the second variable is now investment (Inv). For portfolios formed in June of year t, Inv is the growth of total assets for the fiscal year ending in t-1 divided by total assets at the end of t-2. In the valuation Eq. (3), the investment variable is the expected growth of book equity, not assets. We have replicated all

出处: Fama and French (2015), pp 4

知乎@石川

#### 虽然指标一致,但背后的出发点截然不同:

- 1. q-factor model 来自投资中的 q-theory,从其数学原理出发,收益率和 **past**investment 成 反比,因此 q-factor model 中使用最近一个财年 total assets 的变化率没有任何问题;
- 2. FF5 来自 DCF 模型,从其数学原理出发,收益率和 **expected** investment 成反比;而 FF5 基于 Fama and French (2006)的检验结果,采用 past investment ——即过去一个财年 total assets 的变化率——来对 expected investment 做 naïve estimate。

Hou et al. (2019a) 怼 FF5 的两个核心观点是:

1. DCF 中的 IRR 和未来单期的 expected return 不同;如果从 valuation theory 出发,从理上是可以推导出 expected investment 和 expected return 之间存在**正相关(正如同 Fam and French 2006 里用预测的**  $F[dA_{t+\tau}/A_t]$  和收益率的检验结果),而非 FF5 使用 D 所隐含的负相关 —— 这是从理论上开炮;





#### 知平 川流不息

结合上述两点,FF5 使用 past investment 构建的 CMA 因子 (且该因子有效) 其实是阴差阳错 的利用了 past investment 和收益率之间的负相关(q-theory);而非他们从 DCF 推出的 expected investment 和收益率之间的负相关性。这无疑直接撼动了 FF5 中投资因子的根基, 也质疑了 FF5 在学术界的地位。

因子大战, 硝烟弥漫。

对于上面第一点,Hou et al. (2019a) 的论述如下,感兴趣的小伙伴可以看看。

#### 4.3 The Relations among Past Investment, the Expected Investment, and the **Expected Return**

Fama and French (2015) argue that Equation (2) predicts a negative relation between the expected investment and the IRR. However, this negative relation does not apply to the one-period-ahead expected return,  $E_t[r_{it+1}]$ . From the definition of return,  $P_{it} = (E_t[D_{it+1}] + E_t[P_{it+1}])/(1 + E_t[r_{it+1}])$ , and the clean surplus relation, we can reformulate the valuation Equation (2) in terms of the one-period-ahead expected return:

$$P_{it} = \frac{E_t[Y_{it+1} - \triangle Be_{it+1}] + E_t[P_{it+1}]}{1 + E_t[r_{it+1}]}.$$
 (3)

Dividing both sides of Equation (3) by  $Be_{it}$  and rearranging, we obtain:

$$\frac{P_{it}}{\operatorname{Be}_{it}} = \frac{E_t \left[ \frac{Y_{it+1}}{\operatorname{Be}_{it}} \right] - E_t \left[ \frac{\triangle \operatorname{Be}_{it+1}}{\operatorname{Be}_{it}} \right] + E_t \left[ \frac{P_{it+1}}{\operatorname{Be}_{it+1}} \left( 1 + \frac{\triangle \operatorname{Be}_{it+1}}{\operatorname{Be}_{it}} \right) \right]}{1 + E_t [r_{it+1}]},\tag{4}$$

$$\frac{P_{it}}{\text{Be}_{it}} = \frac{E_t \left[ \frac{Y_{it+1}}{\text{Be}_{it}} \right] + E_t \left[ \frac{\triangle \text{Be}_{it+1}}{\text{Be}_{it}} \left( \frac{P_{it+1}}{\text{Be}_{it+1}} - 1 \right) \right] + E_t \left[ \frac{P_{it+1}}{\text{Be}_{it+1}} \right]}{1 + E_t [r_{it+1}]}.$$
 (5)

Fixing everything except  $E_t[\triangle Be_{it+1}/Be_{it}]$  and  $E_t[r_{it+1}]$ , high  $E_t[\triangle Be_{it+1}/Be_{it}]$  implies high  $E_t[r_{it+1}]$ , because  $P_{it+1}/Be_{it+1}-1$  is likely positive in the data. This prediction is consistent with the weakly positive  $E_t[\triangle Be_{it+1}/Be_{it}] - E_t[r_{it+1}]$  relation documented in Fama and French (2006).

出处: Hou et al. (2019a)

知平 @石川

客观的说,关于 expected investment 和 expected return 的关系,我也认为 Hou et al. (2019a) 比 FF5 更合理。但上面这种"竞赛"还是让人感到不安。它似乎传递出一个信号:我们总 能使用不同的金融学或经济学模型,推导出截然相反的关联 —— 上面的模型较 DCF 模型,通过 把 IRR 换成单期收益率,巧妙的干掉了 dB 前面的负号,因此把"负相关"逆转为"正相关"。 从实证结果出发总能讲出最适合的故事,但对着历史数据讲故事真的让我们离真相更进一步了「

OK, 抛开理论上的差异不说, 从实证结果来看, 虽然 FF5 和 q-factor model 使用了同样的历中 total asset 增长率构建投资因子,但由于构建方法不同,q-factor model 中投资因子的溢价也



从 q-theory 出发,投资和收益率的负相关是在控制了 ROE 之后的。因此,q-factor model 在构建投资因子时利用 size,ROE 和 total assets 增长率三个指标使用 2 × 3 × 3 triple sort,从而更好的反映了在控制 ROE 之后,投资和收益率的关系。反观 FF5,它仅使用 size 和 total asset 增长率做了 2 × 3 double sort,没有考虑控制 ROE 的影响,效果不如 q-factor model。这似乎佐证了 q-factor model 背后的 q-theory 能更好的解释 past investment 和 expected return 之间的关系。

解惑之旅结束。

#### 7 Factor War

让我们最后来看看近几年学术界的 Factor War。以下先后顺序使用论文的发表时间为基准。

在 Hou, Xue, and Zhang (2015) 提出 q-factor model 时,他们使用了 80 个异象(那篇写 Replicating Anomalies 花的功夫绝对不会被浪费)比较了 q-factor model 和 Fama and French (1993) 三因子模型和 Carhart (1997) 四因子模型。结论自然是 q-factor model 战胜了它们。由于当时 FF5 还没有发表,因此并没有被拿来比较。

颇有意思的是,关于 q-factor model 和 FF5 写作背后的时间先后也拿来被当作 factor war 的炮弹。Zhang (2016) 介绍了两篇文章写作的背景(下图),指出 Fama 和 French 在 2013 年 6 月首先加入了盈利因子,然后才又加入了投资因子、构建了 FF5,而这个 timeline 远远晚于 q-factor model 的发展历程(见本文第二节),因此巨人 Fama 是站在了他们的肩膀上。不过 Zhang (2016) 似乎忘记了 Fama and French (2006) 这篇早就对盈利和投资进行检验的文章—— 虽然 Hou et al. (2019a) 对 FF5 的时候想起了它。

Fama-French起初于2013年6月提出一个四因子模型,把一个盈利因子加入他们的三因子模型。在当年11月又将一个投资因子加到四因子模型中,而得出了现在的五因子模型,与我们的q一因子模型抗衡。在学术界有句俗话,写好论文需要站在巨人肩膀上。我常在博士生面前自嘲,说Fama-French五因子模型是巨人站在了我们的肩膀上,还差点没给我压死。

出处: Zhang (2016)

知乎 @石川

2017 年, Stambaugh 和 Yuan 在 Review of Financial Studies 上提出了两个 mispricing factors (Stambaugh and Yuan 2017) , 并结合市场和 SMB 构建了四因子模型。新的模型 1 🛖



Our model's overall ability to accommodate anomalies exceeds that of both the four-factor model of Hou, Xue, and Zhang (2015a) and the five-factor model of Fama and French (2015). This conclusion obtains not only within the set of anomalies used to construct the factors but also for the substantially larger set of 73 anomalies examined previously by Hou, Xue, and Zhang (2015a,b). For example, when applied to the 51 of those anomalies having 出处: Stambaugh and Yuan (2017)

2018 年,Barillas 和 Shanken 在 Journal of Finance 上发表了题为 Comparing asset pricing models 的文章。该文提出了一个 Bayesian asset pricing test 检验不同的定价模型(Barillas and Shanken 2018)。通过对比,q-factor model 和 FF5 都败下阵来,而 Barillas and Shanken (2018) 也根据他们的结果提出了一个六因子模型,包括:市场因子、FF5 的 SMB、q-factor model 的 ROE 和 I/A,Asness and Frazzini (2013) 提出的按月更新的 HML,以及 UMD。

同年,在美国金融协会年会上,Daniel, Hirshleifer, and Sun (2019) ——这里引用使用了即将发表于 RFS 的版本 ——提出了基于行为金融学的两个因子,并结合市场因子构建了一个复合三因子模型(见<u>《一个加入行为因子的复合模型》</u>)。在 AFA 年会的报告中,Daniel, Hirshleifer, and Sun (2019) 将该模型和主流模型进行了对比,除了考察因子模型解释异象及其他因子的能力,也特别从模型复杂度角度进行了对比(见《Anomalies, Factors, and Multi-Factor Models》)。

在 AFA 年会报告的讨论环节,Stambaugh 质疑 Daniel, Hirshleifer, and Sun (2019) 的三因子模型无法解释 SMB 因子,但他们却没有加入这个因子。我现在来看,这反而是 Daniel, Hirshleifer, and Sun (2019) 的一个优点。该模型从行为金融学的理论 promote 出长、短两个周期的行为因子;而从行为金融学的理论是无法引出 SMB 的因子的。所以该模型仅使用三个因子无可厚非;如果生硬的加入了 SMB,反而更有 data mining 之嫌。

自 2015 年 q-factor 被提出后,由于其理论上的创新和实证方面的优秀表现,获得了学术界的广泛关注。但是,随着新的模型被发表,它又被别人比了下去。在这种背(yin)景(mou)之(lun)下,  $m{q}^5$  横空出世。Hou et al. (2019a) 的实证结果显示,  $m{q}^5$  战胜了上述全部 candidates。





models in spanning regressions. The Stambaugh–Yuan (2017) factors are sensitive to their construction, and once replicated via the traditional approach, are close to the q-factors, with correlations of 0.8 and 0.84. Neither the original nor the replicated Stambaugh–Yuan model can explain the q and  $q^5$  factors in the Gibbons–Ross–Shanken (1989) test, but the  $q^5$  model can explain both their original and replicated factors. The Daniel–Hirshleifer–Sun (2018) factors are also sensitive to their construction, and once replicated via the traditional approach, are close to the q-factors, with correlations of 0.69. Their three-factor model cannot explain the size, investment, and expected growth factors, and the q and  $q^5$  models cannot explain their earnings factor. Finally, the Barillas–Shanken (2018) model, which embeds the investment and Roe factors from the q-factor model, cannot explain the expected growth factor in the  $q^5$  model. Although the q-factor model cannot explain the Asness–Frazzini (2013) monthly formed HML factor in the Barillas–Shanken specification, the monthly formed q-factor model can.

出处: Hou et al. (2019a)

从现有的实证结果来看,出自 investment CAPM 的  $q^5$  的五因子模型是当下"最强"的实证资产定价模型。当然,在不远的将来,学术界一定会搞出新的 factor model 来打败它。

近年来,随着 John Cochrane 所提的 factor zoo 越来越深入人心,学术界对挖异象的热潮似乎有所减退,转而把研究重点放到了 factor war。然而,如果仅以各种 test 的结果或者谁解释的异象更多为标准,非要在众多模型中分出高下,又有多少意义?如果仅以"打赢" factor war、而非推动人们对市场的理解为目标,而使用不同的金融学或经济学理论 promote 出一个"更好使"的因子或者互相"怼",这又有多少意义?

因子大战, 烽鼓不息。

### 8 结语

在上一节介绍 factor war 时,我特意保留了一篇,它就是 Fama and French (2018) 这篇题为 Choose Factors 的文章,发表于 Journal of Financial Economics。这篇文章拓展了 Gibbons, Ross, and Shanken (1989) test,提出使用模型所包含因子能够构成的 max squared Sharpe ratio 来比较不同的模型。

之所以保留这篇文章是因为它没有和别人比,而是"自己和自己比"。该文比较了 CAPM, Fama and French (1993) 三因子、Fama and French (2015) 五因子以及 Fama and French (2018) 六因子模型。除了方法论之外,Eugene Fama 在这篇文章中表达出来的对 factor war 的担忧深深引起了共鸣。

Fama and French (2018) 认为,虽然 CAPM 以及 consumption-based CAPM 被 empirica data 拒绝了,但是这些理论模型在收益和风险之间建立了有效的联系。而另一方面,实证资产,



#### efficient (MVE) tangency portfolio 为目标对着历史数据进行挖掘:

There is an obvious danger that, in the absence of discipline from theory, factor models degenerate into long lists of factors that come close to spanning the **ex post** MVE tangency portfolio of a particular period.

出于这个原因, Fama and French (2018) 在 FF5 加入动量因子的时候显得异常小心谨慎:

We include momentum factors (somewhat reluctantly) now to satisfy insistent popular demand. We worry, however, that opening the game to factors that seem empirically robust but lack theoretical motivation has a destructive downside — the end of discipline that produces parsimonious models and the beginning of dark age of data dredging that produces a long list of factors with little hope of sifting through them in a statistically reliable way.

对于一个因子模型来说,它是为了解释股票收益率如何共同运动的,因此必须和个股的协方差矩阵密切相关(Pukthuanthong, Roll, and Subrahmanyam 2019)。**仅靠因子模型之间相互spanning tests,或者用一箩筐其实相关性都不低的异象(全都是投资组合;连个股都没有用)来检验因子模型,恐怕和前面的目标还有一定的距离。**如此 factor war 意义十分有限。

最后,我想用 Fama and French (2018) 里的一段话作为本文的结尾。他们二位呼吁高质量和有意义的因子模型比较。此外,对于 factor modeling,必须要有理论依据,避免建模本身成为对着历史数据进行数据挖掘。然而,实现这一切需要我们对市场更深的理解。

In general, if inference is to have content, the list of models considered in a study must be relatively short. Moreover, if factor modeling is not to degenerate into meaningless dredging for the ex post MVE portfolio, the number of factors in models must also be limited. Establishing ground rules, however, awaits more experience.

感谢阅读,全文完。

#### 参考文献

- Asness, C. and A. Frazzini (2013). The devil in HML's details. *The Journal of Portfolio Management*, Vol. 39(4), 49 68.
- Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *The Journal of Finance*, Vol. 73(2), 715 754.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, Vol. 52(1), 57 82.
- Chen, L. and L. Zhang (2007). Neoclassical factors. NBER Working Paper, No. 13282.





# 知乎 / imake

- Daniel, K., D. Hirshleifer, and L. Sun (2019). Short- and long-horizon behavioral factors. *The Review of Financial Studies*, forthcoming.
- 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 (2006). Profitability, investment and average returns. *Journal of Financial Economics*, Vol. 82(3), 491 518.
- 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 K. R. French (2018). Choosing factors. *Journal of Financial Economics*, Vol. 128(2), 234 252.
- Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica*, Vol. 57(5), 1121 1152.
- Hou, K., H. Mo, C. Xue, and L. Zhang (2018). q5. Charles A. Dice Center Working Paper No. 2018-10; Fisher College of Business Working Paper No. 2018-03-010. SSRN: ssrn.com/abstract=31911....
- Hou, K., H. Mo, C. Xue, and L. Zhang (2019a). Which factors? *Review of Finance*, Vol. 23(1), 1 35.
- Hou, K., H. Mo, C. Xue, and L. Zhang (2019b). Security analysis: an investment perspective. Working paper.
- Hou, K., C. Xue, L. Zhang (2015). Digesting anomalies: an investment approach. *The Review of Financial Studies*, Vol. 28(3), 650 705.
- Hou, K., C. Xue, and L. Zhang (2018). Replicating anomalies. *The Review of Financial Studies*, forthcoming.
- Liu, L. X., T. M. Whited, and L. Zhang (2009). Investment-based expected stock returns. *Journal of Political Economy*, Vol. 117(6), 1105 1139.
- Miller, M. H. and F. Modigliani (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business*, Vol. 34(4), 411 433.
- Pukthuanthong, K., R. Roll, and A. Subrahmanyam (2019). A protocol for factor identification. *The Review of Financial Studies*, Vol. 32(4), 1573 1607.
- Stambaugh, R. F. and Y. Yuan (2017). Mispricing factors. *The Review of Financial Studies*, Vol. 30(4), 1270 1315.
- Zhang, L. (2016). Factors war. *Tsinghua Financial Review*, Vol. 37, 101 104, in Chinese.
- Zhang, L. (2017). The investment CAPM. European Financial Management, Vol. 23(4), 545 – 603.

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。





发布于 2019-07-10

资产定价 资产定价模型 (CAPM) 金融学

▲ **赞同 83** ▼ **9** 5 条评论 **7** 分享 ★ 收藏 …

#### 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





股重供求影不影响股重裁而收益? 加里山脊公司和交易所押过土用中的亚奇委托及挂合数据公



也许没有固定模式,不是回归就能解释。大家都是在尝试新的模型去更好的匹配历史股票截面 收益率差异,是不是有点典型的过度挖掘,牵强附会?

┢ 赞



#### Hunter小能手

1个月前

这些factor model很大程度上data mining,而且是基于已经发现的anomaly,如果以后再出 现新anomaly, 这些factor model会马上继续加新factor

┢ 赞



### **M** Julian Wong

1个月前

作為一個價投,看到DCF出現在估值模型,有種2001太空漫遊裡,猩猩總於學會拾起骨頭當 武器的既視感, 笑。方向是對的。

┢ 赞



#### 🤻 零度交易员

1个月前

以对某因子未来的预测值来预测未来 和DCF好像 不是很看好 人类可笑的预测能力无非就是线 性外推和物极必反两种方法

┢ 赞



