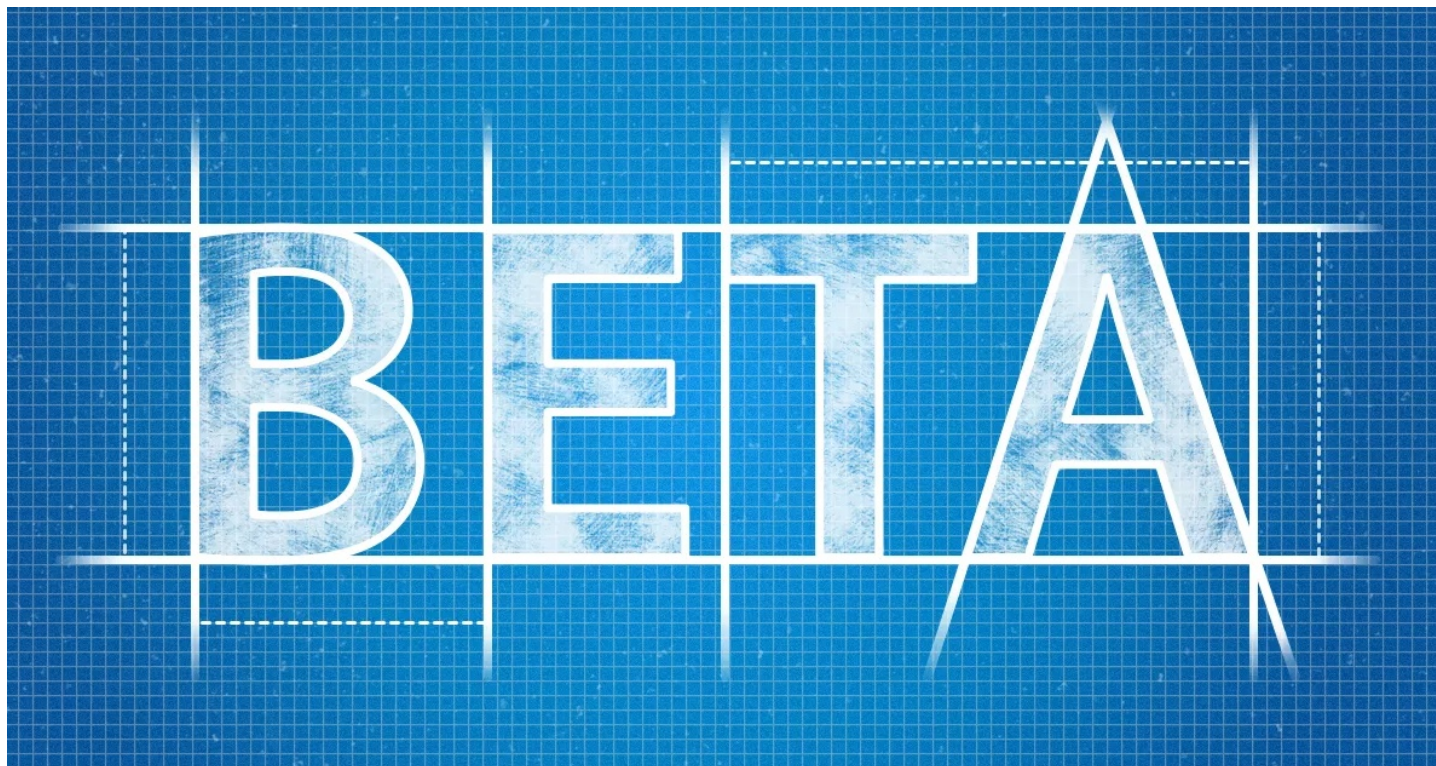


知乎

首发于  
川流不息

## Which Beta ?



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

为检验因子是否有显著的风险溢价，首先需要正确计算因子载荷。本文讨论了收益率时序回归  $\beta$  和 firm characteristics 作为  $\beta$  时的差异，并给出了一些建议。

### 1 引言

Spoiler alert!

这篇文章关注的是 **empirical asset pricing test 中的 test**。它既不介绍异象，也不讲述因子，而是着眼于检验。本文的内容虽然重要，但行文非常 technical 且结论又看似 boring；以“真香”论来看，搞清楚文中的解读不如记住最后的结论。

OK，如果你没有被上述剧透打败，那让我们开始吧。

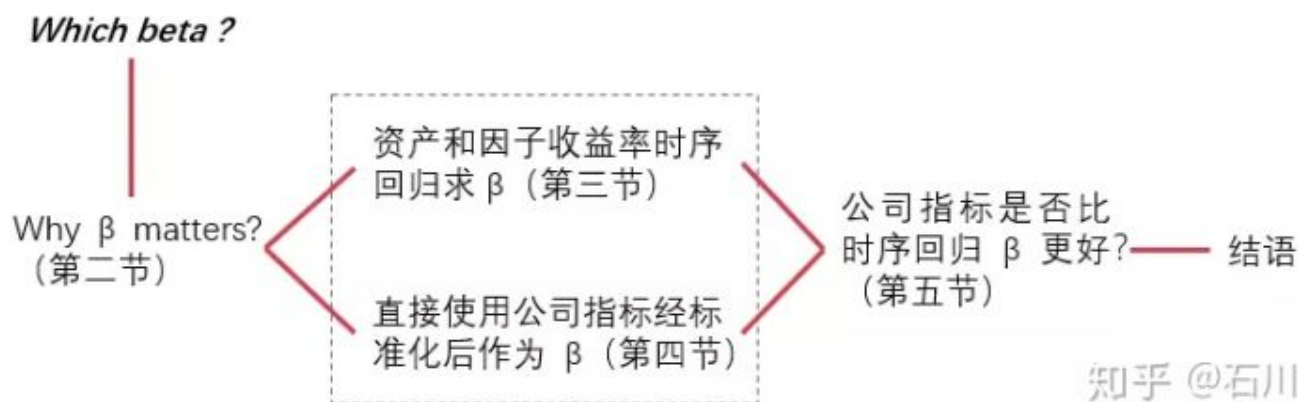
在检验因子是否有显著的 **risk premium** 时，常用的工具是 **Fama-MacBeth Regression**；**i** 方法首先需要明确股票在每个因子上的因子载荷（factor loading，即我们常说的  $\beta$ ）。计算 **尸** 载荷有两种方法：

知乎

首发于  
川流不息

2. 像 Barra 那样直接使用 firm characteristics 作为因子载荷（例如，直接用个股的 P/B 取值经过必要的标准化后作为因子载荷）。

比较这两种方法的差异就是本文的目标，这也是本文取名 which beta 的原因。由于篇幅较长，为避免各位小伙伴在阅读过程中 get lost，下面先给出文章的思维导图。



下面让我们从 Fama-MacBeth Regression 说起。

## 2 Fama-MacBeth Regression

Fama-MacBeth Regression 是学界和业界检验因子 risk premium 的主流方法之一（Fama and MacBeth 1973，见《股票多因子模型的回归检验》）。它的检验过程分为两步：

1. **第一步是时序回归**：把待检验的因子收益率放在回归方程的 RHS，把用来检验这些因子的资产收益率逐一放在回归方程的 LHS，使用 multivariate regression 计算每个资产在这些因子上的因子载荷  $\beta$ ；
2. **第二步是截面回归**：使用第一步得到的  $\beta$  作为解释变量放在 RHS，使用资产的收益率放在 LHS，截面回归求出因子的 risk premium  $\lambda$ ；每一期得到每个因子的溢价后，最后检验每个因子溢价的均值是否显著。

上述过程的第一步是为了第二步服务，而第二步具备两个优点：（1）在截面上使用多因子回归可以考察每个因子对解释资产预期收益率差异的增量贡献；（2）“先单期截面回归、再从时序上取平均”可以排除残差收益率的截面相关性带来的影响。

然而，上述两步走的过程存在一个不可避免的问题，即第一步通过时序回归得到的因子载荷仅仅是真实但未知的  $\beta$  的估计，因而存在误差；将这个 estimate 直接作为解释变量用在第二步就引入了 **errors-in-variables 问题 (EIV)**。

Fama and MacBeth (1973) 自然意识到了这个问题。为此，在这篇以检验 CAPM 为初衷的文章中，他们并没有使用个股的收益率，而是将个股按照历史  $\beta$  的大小构成了不同的投资组合，然

消，因此对投资组合的  $\beta$  估计会更准确，从而在一定程度上降低 EIV 的影响。

自此以后，在进行 Fama-MacBeth Regression 检验因子时，使用投资组合而非个股放在 LHS 就成为了学术界的标配。但有大佬对此颇有微词，这其中就包括 Fama 的弟子 Richard Roll。

Roll 和他的 co-authors 在 Jegadeesh et al. (2019) 这篇最新发表于 Journal of Financial Economics 的文章指出**将个股按照某种属性分组实际上是一种降维处理，投资组合会丢掉很多个股截面上的特征**。如果待检验的因子和这些 LHS 组合的分组属性正交，用它们作为 LHS 进行 Fama-MacBeth Regression 是无法发现这些因子的 risk premium 的。因此，Jegadeesh et al. (2019) 建议使用个股收益率放在 LHS。

除此之外，学术界对于投资组合的过度使用还有另一个不好的 side effect。Fama and French (1993) 三因子的横空出世，不仅让人们从此接受了 SMB 和 HML 两个因子，更让使用 Size 和 B/M 进行 double sort 得到的  $5 \times 5$  一共 25 个投资组合成为了评价不同因子模型时的标配。**很多模型以能够在截面上很好的解释这 25 个资产的预期收益率为依据，说明提出的新因子是有效的**。对此，另外两位大佬 Stefan Nagel 和 Jay Shanken 认为这是极大的误解。

Nagel 和 Shanken 在 Lewellen, Nagel, and Shanken (2010) 一文中指出，使用 Size 和 B/M double sort 的这 25 个组合有非常强的 factor structure。任何一个待检验的因子，如果它们和 HML 或 SMB 的相关性大于它和这 25 个组合被 HML 和 SMB 解释后的残差的相关性的话，这个新因子就能在这些投资组合的截面上获得很高的  $R^2$ ，换句话说，这些组合的  $\alpha$  会很接近零。因此，哪怕一个因子能够很好的解释这 25 个投资组合的截面预期收益率差异，它也未必是有效的。

以上两个问题均说明**在检验因子以及因子模型时，仅使用有限个投资组合作为 test assets 不再合适**。Lewellen, Nagel, and Shanken (2010) 提出的解决方法之一正是使用更多的投资组合。比如，在 Fama and French (2019) 这篇最新的研究中，他们二位就使用不同的公司指标两两 double sort 构建了 200+ 个投资组合，检验了时序和截面因子模型的效果。

如今，距离 Fama and MacBeth (1973) 的提出已经过去了快 50 年。无论是计量经济学的发展还是学术界因子挖掘的日趋严重 (factor zoo)，都促使我们摒弃投资组合、转而使用个股收益率放在回归方程的 LHS。

**当使用个股收益率放在 LHS 时，必须尽量排除 EIV 问题、获得尽可能准确的因子载荷的估计**。为此，Jegadeesh et al. (2019) 提出在 Fama-MacBeth Regression 的第一步时序回归中引入 Instrumental Variables (IV)，它将是下一节的内容。

对 EIV 问题，业界有着不同的做法。**我们熟悉的 Barra 的纯因子模型本质上正是 Fama-MacBeth Regression**。但是它没有使用第一步时序回归计算因子载荷，而是使用了 firm characteristics 作为因子载荷，然后进行截面回归。由于不需要估计  $\beta$ ，Barra 在回归方程的 LHS 使用了个股收益率。当然，Barra 多因子模型的目的是为了计算个股的协方差矩阵，从这个意义上说，LHS 使用投资组合也没什么道理。





### 3 Instrumental Variables

在 Fama-MacBeth Regression 中，第二步截面回归的数学表达式为：

$$\mathbf{r}_t = \mathbf{B}\lambda + \xi_t$$

由于采用了个股收益率放在 LHS，因此上式中  $\mathbf{r}_t$  是个股收益率向量； $\mathbf{B}$  是第一步时序回归得到的因子载荷估计矩阵，它是一个  $N \times K$  阶矩阵（ $N$  支个股； $K$  个因子）； $\lambda$  是带求解的因子 risk premium 向量； $\xi_t$  是残差向量。

为减少 EIV 问题的影响，Jegadeesh et al. (2019) 在上述 OLS 回归中引入了 Instrumental Variables，因此得到  $\lambda$  的 IV estimate 为：

$$\lambda_{IV,t} = \left( \mathbf{B}'_{IV} \mathbf{B}_{EV} \right)^{-1} \left( \mathbf{B}'_{IV} \mathbf{r}_t \right)$$

上式中， $\mathbf{B}_{IV}$  和  $\mathbf{B}_{EV}$  分别为 instrumental 和 explanatory variables： $\mathbf{B}_{EV}$  是对  $\beta$  的 estimate； $\mathbf{B}_{IV}$  是  $\mathbf{B}_{EV}$  的工具变量。Jegadeesh et al. (2019) 使用互不重叠的历史数据分别进行时序回归求解  $\mathbf{B}_{IV}$  和  $\mathbf{B}_{EV}$ ，并指出正因如此，它们在截面上是不相关的，可以减少 EIV 问题。

具体来说，在每个月末，为了计算最新的  $\mathbf{B}_{IV}$  和  $\mathbf{B}_{EV}$ ，Jegadeesh et al. (2019) 使用过去三年个股的日频收益率和多因子模型的日频收益率进行时序 multivariate regression：

1. **如果当前月是偶数月**（比如二月、四月、六月等），则使用过去三年窗口内所有的**偶数月**之中个股和多因子的收益率进行回归，得到的回归系数就是  $\mathbf{B}_{EV}$ ；使用这三年窗口内所有**奇数月**之中个股和多因子的收益率进行回归，得到的回归系数作为  $\mathbf{B}_{IV}$ 。
2. **如果当前月是奇数月**（比如一月、三月、五月等），则使用过去三年窗口内所有的**奇数月**之中个股和多因子的收益率进行回归，得到的回归系数就是  $\mathbf{B}_{EV}$ ；使用这三年窗口内所有**偶数月**之中个股和多因子的收益率进行回归，得到的回归系数作为  $\mathbf{B}_{IV}$ 。

由于 EIV 问题，Fama-MacBeth Regression 中第二步截面上的 OLS 计算得到的 risk premium 往往是 biased，而多个因子一起的多元截面回归使得我们无法确定每个因子溢价上 bias 的方向。Jegadeesh et al. (2019) 指出，上述 IV estimate 可以获得 risk premium 的无偏估计。

以 CAPM 单因子为例，下图给出了 Jegadeesh et al. (2019) 的分析结果。当使用传统的 OLS 时，对 market risk premium 的估计的 bias 随着计算  $\beta$  的时序窗口增加而减少，但即便是使用了长达 10 年的数据（2500 个交易日），bias 仍然高达 -7% 左右；而采用 IV estimate，对 market risk premium 的估计是 unbiased。



知乎

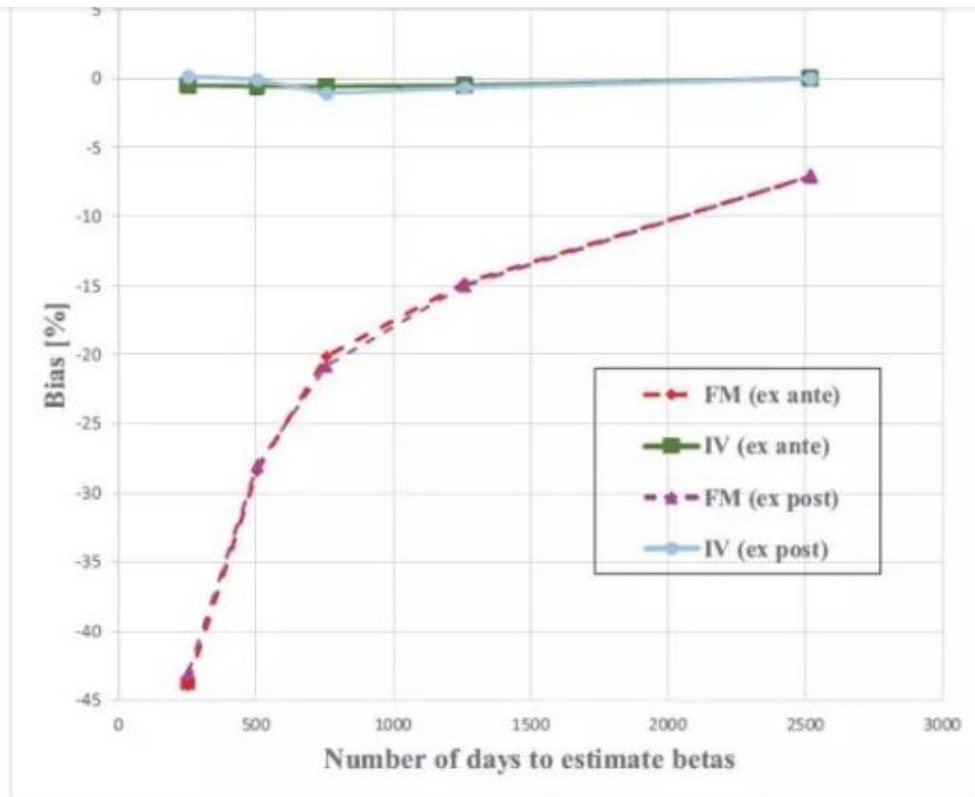
首发于  
川流不息

Fig. 1. Biases versus number of days used to estimate betas. This figure presents the ex ante and ex post biases using the ordinary least squares (OLS) and instrumental variables (IV) estimators of the market risk premium under the CAPM, as a function of the number of days in the rolling window to estimate the market betas. The simulations use the market risk premium of 5.80% per annum and 2000 individual stocks in the cross-section. The sample period for the simulations is 684 months. The y-axis is bias as a percentage of the true market risk premium, and the x-axis is the number of days in the rolling window to estimate the market betas. These results are based on 1000 repetitions for each estimation window.

出处: Jegadeesh et al. (2019)

知乎 @石川

显然，即便是对于美股，使用十年滚动窗口来估计  $\beta$  都是相当奢侈的（况且它还无法消除 bias），更不要提大 A 股了。因此，IV estimate 在计算 risk premium 时优势明显。

下面以中证 500 成分股为例（将它们放在回归方程的 LHS），检验 Fama and French (2015) 的五因子模型中的因子是否有 risk premium。实证期为 2010 年 1 月至 2019 年 4 月。在每月末，使用过去两年的日频收益率序列估计  $B_{EV}$  和  $B_{EV}$ 。作为对比，同时使用不引入 IV 的方法，即使用两年滚动窗口的全部数据估计  $\beta$  并将它用在 Fama-MacBeth Regression 的第二步。下表给出了这五个因子的检验结果（其中 Test 1 为未引入 IV 的方法；Test 2 为使用 IV estimate 的方法）。

知乎

首发于  
川流不息

- **Test 1:** 对于每只个股，使用最初四年窗口估计收益率并五因子收益率口频数据近行时序 multivariate regression 计算  $\beta$ 、并用  $\beta$  作为解释变量通过 OLS 计算 risk premium;
- **Test 2:** 在 Test 1 的基础上采用 Jegadeesh et al. (2019) 的方法引入 IV，使用  $\beta$  和 IV 计算 risk premium 的 IV estimate。
- 下表中为不同因子月频 risk premium 的均值、括号内为经 Newey and West (1987) 调整的 t-statistic。

	月收益率均值 (%)	
	Test 1: OLS	Test 2: IV
MKT	0.69 (0.94)	0.64 (0.86)
SMB	0.24 (1.59)	<b>0.58</b> <b>(1.90)</b>
HML	-0.00 (-0.02)	0.13 (0.32)
RMW	0.35 (1.28)	0.40 (0.96)
CMA	0.16 (0.90)	0.04 (0.13)

知乎 @石川

结果显示，当使用个股作为资产放在 LHS 时，由于 EIV 的问题，Fama-MacBeth Regression 给出的结论是全部五个因子均没有显著的 risk premium —— 即五因子不能有效的解释个股截面预期收益的差异。当采用 IV estimate 时，SMB 因子变得显著（其他四个因子仍然不显著）。这个简单的结果表明消除 EIV 问题，即使用更准确的  $\beta$  estimate，对于评判一个因子是否有效至关重要。

从上面的结果可以得到本文 which beta 之问的第一个观点：**在  $\beta$  估计时引入 IV 消除 EIV 的影响比直接使用  $\beta$  做 OLS 截面回归能够获得更准确的 risk premium estimate。**

如果讨论就此结束实在意犹未尽。事实上，Jegadeesh et al. (2019) 的实证分析也未止步于此；他们在  $\beta$  的基础上加入了 firm characteristics。使用时序回归系数和 firm characteristics 同时作为  $\beta$  也是这两种方法的直接 PK。

## 4 Firm Characteristics

Jegadeesh et al. (2019) 在美股上使用他们提出的 IV estimate 检验 Fama and French (1993) 三因子模型并发现 SMB 和 HML 确实有显著的 risk premium。不过，他们同时指出，以上显著的结果可能源自被忽视的变量偏差，即在截面回归时没有控制 Size 和 B/M 这些用来构建 SMB 和 HML 的 firm characteristics。为此，他们将 Size 和 B/M 加入到 Fama-MacBeth Regression 的第二步，和 HML 以及 SMB 的  $\beta$  estimate 一起进行了截面回归。



## HML 的 risk premium 不再显著。

Table 6

Risk premium estimates with individual stocks: CAPM and Fama-French three-factor model.

This table reports risk premium estimates, in percent per month, and their tests using individual stocks as test assets. Panel A reports the test results using the IV method in Columns (1)–(5) and those using the OLS method in Columns (6)–(10) for comparison. Panels B and C report the asset pricing test results using the IV method for two subperiods. Rows labeled MKT, SMB, and HML are the risk premiums for the market, SMB, and HML factors, respectively, and the corresponding t-statistics are in parentheses (bold if significant at the 5%). SIZE is the natural logarithm of market capitalization. BM is the book-to-market ratio at the end of the previous month and logBM is the natural logarithm of BM. Betas for each month are estimated using daily returns data over the previous 36 months and cross-sectional regressions are fitted using the IV and OLS methods. The sample period is from January 1956 to December 2012.  $N$  is the average number of stocks in monthly cross-sectional regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: 1956–2012, $N = 2425$										
	IV			OLS						
Const	<b>1.028</b> (7.89)	<b>0.725</b> (6.00)	<b>2.906</b> (4.31)	<b>2.729</b> (4.55)	<b>3.007</b> (4.92)	<b>0.951</b> (7.41)	<b>0.847</b> (7.51)	<b>3.301</b> (4.49)	<b>3.538</b> (5.66)	<b>3.695</b> (5.78)
MKT	−0.246 (−1.36)	−0.288 (−1.60)	−0.090 (−0.51)	−0.018 (−0.10)	0.019 (0.11)	−0.101 (−0.64)	−0.196 (−1.46)	0.076 (0.44)	0.120 (0.83)	0.130 (0.91)
SMB		<b>0.301</b> (2.20)		−0.043 (−0.42)	−0.019 (−0.19)		<b>0.213</b> (2.16)		−0.073 (−1.01)	−0.064 (−0.90)
HML		<b>0.344</b> (2.55)		0.242 (1.88)	0.185 (1.46)		<b>0.225</b> (2.79)		0.136 (1.73)	0.124 (1.64)
SIZE			−0.120 (−3.49)	−0.118 (−3.93)	−0.122 (−3.94)			−0.143 (−3.84)	−0.155 (−4.93)	−0.155 (−4.81)
BM			<b>0.196</b> (4.40)	<b>0.180</b> (4.50)				<b>0.186</b> (4.25)	<b>0.141</b> (3.60)	
logBM					<b>0.177</b> (4.31)					<b>0.133</b> (3.32)

出处: Jegadeesh et al. (2019)

知乎 @石川

我们同样来看看 A 股的实证结果。考虑 firm characteristics 之后，把上一节 Test 1 和 Test 2 中分别加入对应 SMB、HML、RMW、CMA 四因子的公司指标得到 Test 3 和 Test 4。在这两个 tests 中，相应的公司指标为 Size（使用的对数流通市值）、B/P、OP/TA 以及 INV。



(经标准化) 进行截面回归, 检验它们的 risk premium;

- **Test 4:** 在 Test 2 的基础上引入和因子对应的 firm characteristics; 使用因子  $\beta$ 、 $\beta$  的 IV 以及 firm characteristics 一起进行截面回归, 检验因子和 firm characteristics 的 risk premium。
- 下表中为不同因子月频 risk premium 的均值、括号内为经 Newey and West (1987) 调整的 t-statistic。

	月收益率均值 (%)	
	Test 3: OLS	Test 4: IV
MKT	0.71 (0.95)	0.70 (0.92)
SMB	0.20 (1.17)	0.43 (0.93)
HML	-0.49 (-1.38)	-0.45 (-0.92)
RMW	-0.04 (-0.20)	-0.19 (-0.41)
CMA	0.15 (0.87)	0.13 (0.38)
Size	-0.16 (-1.22)	-0.13 (-0.69)
B/P	0.42 (4.19)	0.45 (4.13)
OP/TA	0.49 (6.71)	0.53 (5.11)
INV	0.05 (1.20)	0.18 (1.55)

知乎 @石川

引入 firm characteristics 之后, Fama and French (2015) 五因子在中证 500 上均不显著; 而 B/P 和 OP/TA 这两个公司指标可以获得显著的 risk premium。Firm characteristics “战胜” 时序回归  $\beta$  的结果和 Jegadeesh et al. (2019) 在美股上的实证结果十分相似 (他们针对美股也检验了五因子模型和 Hou, Xue, and Zhang 2015 的 q-factor 模型, 均有类似的发现)。

当使用投资组合作为检验因子的资产时, 这些学术界的因子的确可以获得显著的 risk premium。然而, 当使用个股收益率、并考虑了 firm characteristics 后, 它们均不再显著。这背后的原因是什么呢? Jegadeesh et al. (2019) 给出了它们的尝试, 无奈最后的结论是 “remains a puzzle”; 我们也要给出自己的思考。

接下来从以下两个角度思考这个问题:

1. 日频收益率高噪声使得因子载荷  $\beta$  的取值在截面上非常不稳定。
2. Firm characteristics 比时序回归系数  $\beta$  是更好的因子载荷 proxy。

## 5 时序回归 vs 公司指标



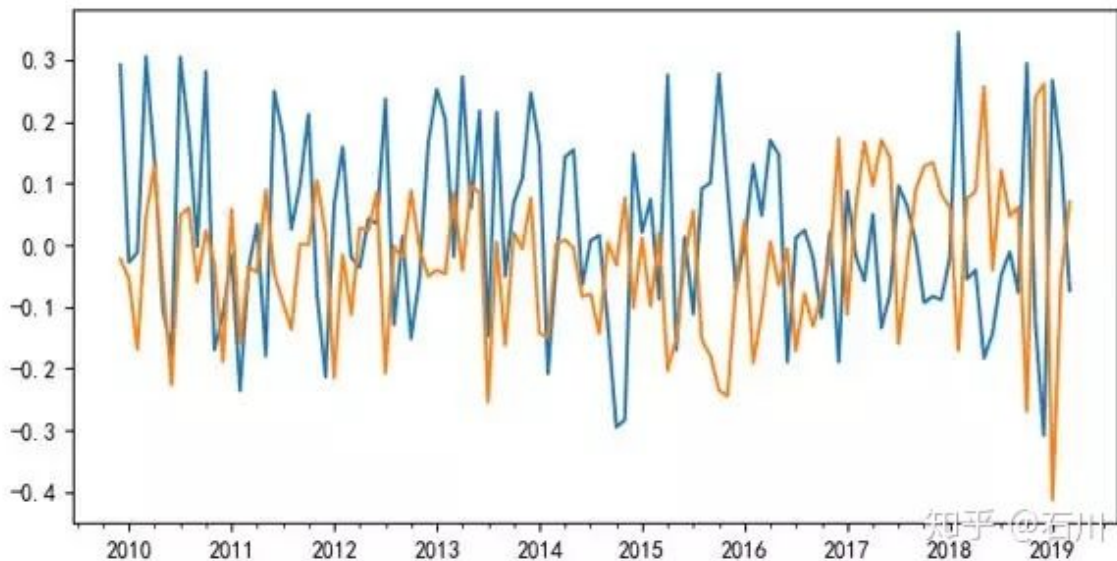


可以猜测时序回归得到的  $\beta$  在截面上非常不稳定，使得这些因子的行为就像随机因子一样。这可能是由于日频收益率数据的高噪音和多元回归中因子之间的相互影响所致。

另一方面，使用 firm characteristics 比使用时序回归  $\beta$  能获得更显著的 risk premium，这说明 firm characteristics 对应的 pure factor portfolio 能获得更高的收益率。这意味着与**时序回归的  $\beta$  相比，当期 firm characteristics 和下一期个股收益率之间的 IC 以及 ICIR 更高**。下列实证结果也证实了这个猜想。

下图是 SMB 因子的时序回归  $\beta$  和 Size 这两个解释变量和个股未来收益率之间的 rank IC 时序图（蓝色为时序回归  $\beta$ ；黄色为 Size），图中同时给出了这两个变量的 IC 均值以及 ICIR（Size 的 IC 均值及 ICIR 为负说明小市值更好）。

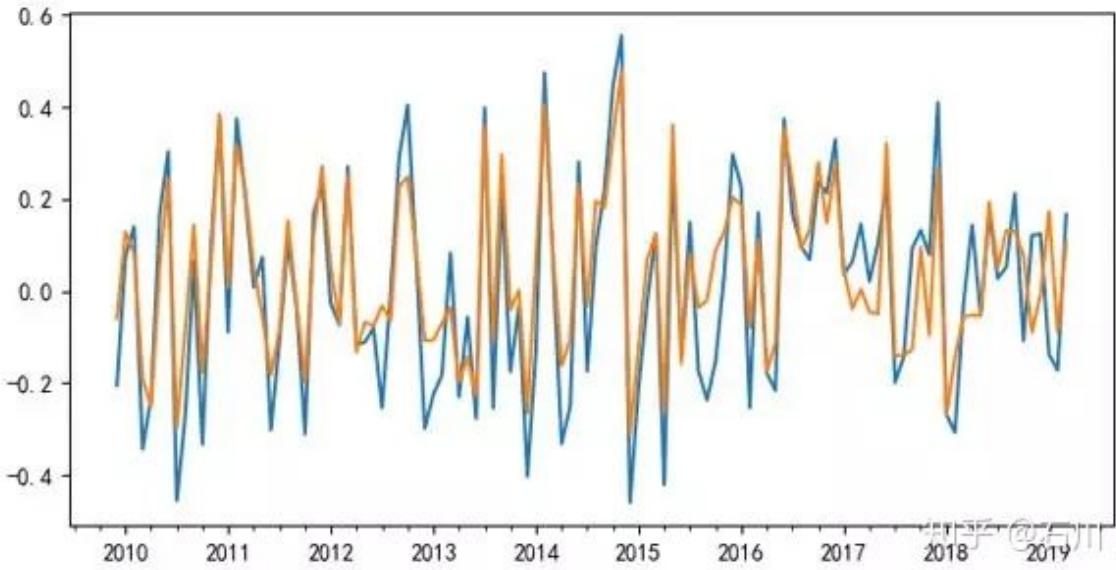
	IC	ICIR
SMB $\beta$	0.023	0.151
Size	-0.025	-0.208



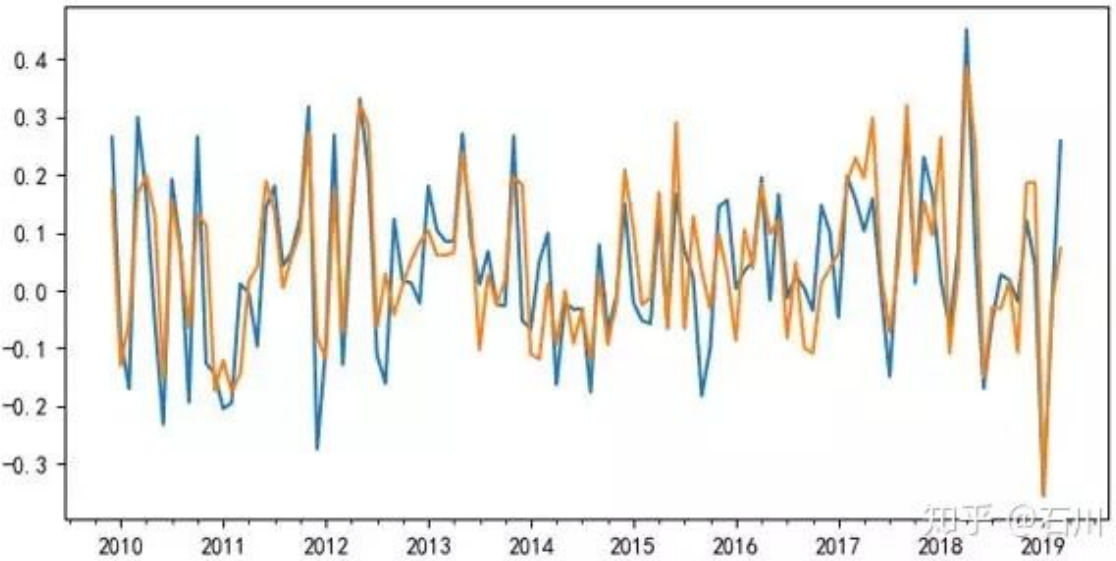
结果显示，时序回归得到的 SMB 的  $\beta$  和个股收益率的 IC 均值（绝对值）没有 Size 高；说明 Size 比 SMB  $\beta$  对于未来收益率更具有预测性。此外 Size 的 ICIR（绝对值）更高也说明它作为因子载荷，要比 SMB 的  $\beta$  更加稳定。

在其他三个因子 HML、RMW 以及 CMA 上也能观察到类似的结果，这里不再赘述，结果汇总于以下三张图。

HML $\beta$	0.010	0.043
B/P	0.035	0.196



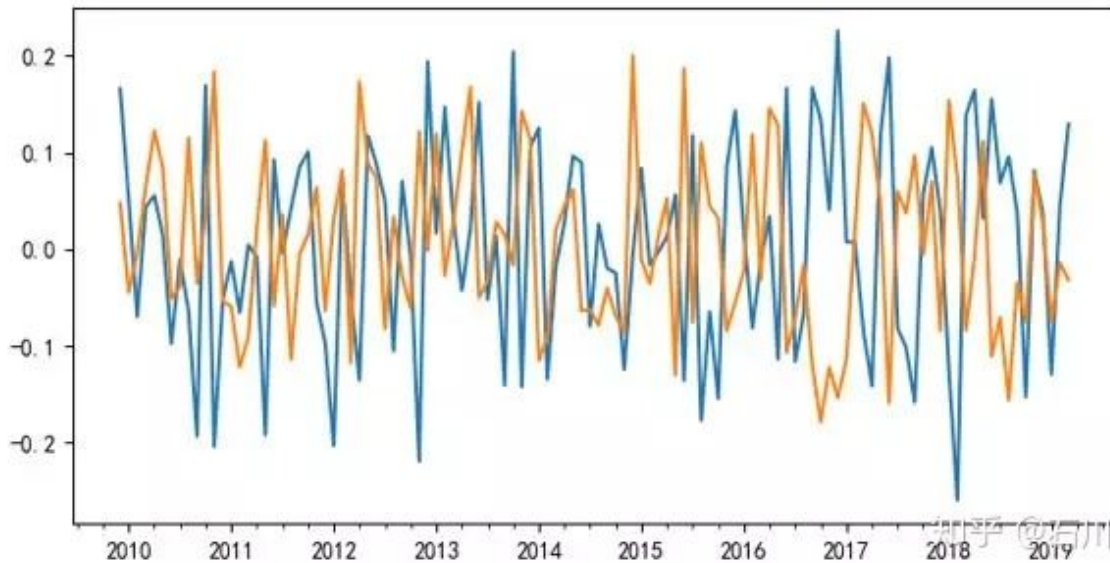
	IC	ICIR
RMW $\beta$	0.038	0.263
OP/TA	0.047	0.353



知乎

首发于  
川流不息

CMA $\beta$	0.004	0.037
Inv	0.004	0.041



以上实证结果说明，作为因子载荷，firm characteristics 确实比它们对应的因子  $\beta$  有更好的预测性。但这背后的原因是什么呢？近日，Richard Roll 和他的 co-authors 在 Review of Financial Studies 上发表了一篇题为 A protocol for factor identification 的文章 (Pukthuanthong, Roll, and Subrahmanyam 2019)，阐述了一个系统识别因子的规程。这篇文章本身非常值得一读，不过我提它是因为一个 side note。

这篇文章最早写于 2015 年并获得了 The Q Group 的 Jack Treynor 奖；Richard Roll 也于 2016 年应邀做了该文的报告。在报告中，他指出**因子本身的运动是难以预测的，因此像 firm characteristics 这些可以提前知道的指标不能成为因子。它们能够获得风险溢价可能有两个原因：（1）它们是未知因子载荷的 good proxy；（2）它们代表了套利机会。**

## Principles that Apply

- Factor movements should not be easily predictable
- A characteristic such as firm size, or anything else known in advance, cannot be a factor
- However, characteristics can be related to mean returns either
  - because they happen to align with factor loadings
  - or because they represent arbitrage opportunities



代表了因子载荷。

需要强调的是，“firm characteristics 是更好的因子载荷”这个观点仍然是一种猜测。Jegadeesh et al. (2019) 就这一点构建了一个假设，但很遗憾实证数据并没有支持这个观点。

我们在此给出 which beta 之问的第二个观点：**同时考虑 firm characteristics 和时序回归系数时，后者无法获得显著 risk premium；从实证结果来看，firm characteristics 确实是更好使的因子载荷，但背后的原因依然未知。**

## 6 结语

如今，无论是检验单个因子还是因子模型，仅使用有限个投资组合作为 LHS 都难以令人满意。正因如此，学者们提出使用个股的收益率作为 LHS。但由于 EIV 的问题，如何计算个股的因子载荷  $\beta$  至关重要。

本文就 which beta 问题进行了系统的阐述。

文章从 Jegadeesh et al. (2019) 提出的 IV estimate 出发，考虑了时序回归估计  $\beta$ ，并引入 firm characteristics 作为控制变量，构建了一共 4 个 tests 检验了 Fama and French (2015) 五因子模型在中证 500 上是否能够获得显著 risk premium。作为回顾，这 4 个 tests 的设置总结于下表。

Test	说明
1	<ul style="list-style-type: none"><li>使用 time-series multivariate regression 计算因子载荷 beta</li><li>使用 beta 作为解释变量求 risk premium</li></ul>
2	<ul style="list-style-type: none"><li>使用 time-series multivariate regression 计算因子载荷 beta <b>并引入 IV</b></li><li>使用 beta 和 IV 一起进行截面回归求 risk premium</li></ul>
3	<ul style="list-style-type: none"><li>使用 time-series multivariate regression 计算因子载荷 beta</li><li><b>加入和 beta 对应的 firm characteristics 作为控制变量</b></li><li>使用 beta 和 firm characteristic 作为解释变量求 risk premium</li></ul>
4	<ul style="list-style-type: none"><li>使用 time-series multivariate regression 计算因子载荷 beta <b>并引入 IV</b></li><li><b>加入和 beta 对应的 firm characteristics 作为控制变量</b></li><li>使用 beta、IV 以及 firm characteristics 一起进行截面回归求 risk premium</li></ul>

综合本文的实证结果，关于 which beta 的探讨得到以下两个观点：

1. 在  $\beta$  估计时引入 IV 消除 EIV 的影响比直接使用  $\beta$  做 OLS 截面回归能够获得更准确的 risk premium estimate;





正确计算因子载荷至关重要。例如，在业绩归因时，如果不知道具体的持仓明细，就只能把待归因的收益率作为被解释变量、把因子收益率作为解释变量进行时序回归，以回归系数  $\beta$  作为评价的依据（AQR 用自家的多个风格因子对美股上多位传奇大佬的收益进行过业绩归因）。又如，在构建宏观因子模型（macroeconomic factor models）时，由于宏观经济数据不是公司财务指标，因此只能通过时序回归计算股票在这些因子上的因子载荷（比如常见的对 GDP、利率的敏感度等）。

在 Fama and French (1996) 这篇著名的解读三因子模型的文章中，两位作者指出，解释一支股票的截面收益应该关注的是它和因子之间的  $\beta$ ，而非公司指标上的取值——比如 B/P 小的公司如果在 HML 上的暴露大，那么该公司实际上会被认为是价值股，而非成长股。

以上例子都说明，使用时序回归  $\beta$  是符合我们认知的。但实证数据显示的却又是另一个故事。当个股收益率被放在 LHS 时，firm characteristics 完胜时序回归系数：前者能获得 risk premium，但后者却不行。而 Jegadeesh et al. (2019) 的分析表明，**firm characteristics 作为因子载荷时，它们和未来因子载荷之间的相关性甚至还低于使用历史数据时序回归的  $\beta$  和未来因子载荷的相关性**。这就是为什么他们把这个现象称作未解之谜。

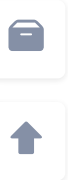
对于到底该使用时序回归系数还是 firm characteristics 作为因子载荷，《股票多因子模型的回归检验》的第六节曾有过简单讨论（今天算是个进阶版）。尽管在实证中 firm characteristics 获得了显著的 risk premium，用它作为因子载荷也不能说 100% 有道理。从多因子模型来说，股票在因子上的收益正比于它们的因子载荷。但像对数市值或 P/B 这些指标，即使公司 A 的指标是公司 B 的两倍，也不能说 A 在该因子上获得的收益率就是 B 的两倍；从因子模型的本质来说，收益率时序回归系数作为因子载荷更加合理。但另一方面，当公司基本面发生较大变化时，回归系数的变化又不能及时、准确的反映股票在因子上暴露的变化。无论如何，选择正确的因子载荷是我们需要解决的问题。

从“真香”的角度来说，似乎使用 firm characteristics 无可厚非；但是从真相的角度来说，我们还有很长的路要走。

So, which beta?

## 参考文献

- Black, F., M. C. Jensen, and M. Scholes (1972). The Capital Asset Pricing Model: Some Empirical Tests. In *Studies in the Theory of Capital Markets*. M. C. Jensen (ed), New York: Praeger, 79 – 121.
- 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 (1996). Multifactor explanations of asset pricing



知乎

首发于  
川流不息

*Financial Economics*, Vol. 116(1), 1 – 22.

- Fama, E. F. and K. R. French (2019). Comparing Cross-Section and Time-Series Factor Models. Chicago Booth Research Paper No. 18-08; Fama-Miller Working Paper. SSRN: [ssrn.com/abstract=32557](https://ssrn.com/abstract=32557)....
- Fama, E. F. and J. D. MacBeth (1973). Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy*, Vol. 81(3), 607 – 636.
- Hou, K., C. Xue, L. Zhang (2015). Digesting anomalies: an investment approach. *Review of Financial Studies*, Vol. 28(3), 650 – 705.
- Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. Wang (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics*, Vol. 133(2), 273 – 298.
- Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, Vol. 96(2), 175 – 194.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, Vol. 55(3), 703 – 708.
- Pukthuanthong, K., R. Roll, and A. Subrahmanyam (2019). A protocol for factor identification. *Review of Financial Studies*, Vol. 32(4), 1573 – 1607.

**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。

原创不易，请保护版权。如需转载，请联系获得授权，并注明出处，谢谢。已委托“维权骑士”  
([维权骑士\\_免费版权监测/版权保护/内容多平台分发](#)) 为进行维权行动。

编辑于 2019-07-03

多因子模型    资产定价    计量经济学

▲ 赞同 51    ▼    9 条评论    ➤ 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

知乎

首发于  
川流不息**你用因子，他也用因子；你没赚钱，他却赚钱了**

石川

**出色不如走运 (III)?**

石川

**简单粗暴地  
资产定价**

=====

更新

=====

石川: Anor Multi-Factor  
写了这样一

Reinhardt J

## 9 条评论

⇌ 切换为时间排序

写下你的评论...



mathquant

1 个月前

所以Fama挖的坑，大家什么时候才能填完。已经没法相信了都。

解释股票间cross sectional的因子更应该包括是公司管理团队、公司所处行业竞争程度及行业地位、行业政策环境等等。然而很多数据比较难以获得。

构建组合因子收益率 在我A 因证券化程度不高，样本选入大概率有偏，用这个估计beta，想想都觉得不太靠谱。这个因子收益率代理变量如果长期“有效”则说明市场存在明显套利机会，如果不存在这种套利，该值估计接近于一个“随机值”。则股票相对于它的暴露值 估计没什么可解释性。

👍 1    ↩ 回复    🗑 踩    🚩 举报



石川 (作者) 回复 mathquant

1 个月前

手动点赞

👍 赞



吾与呜啦啦孰美

1 个月前

很明显firm characteristic啊，“即使公司 A 的指标是公司 B 的两倍，也不能说 A 在该因子上获得的收益率就是 B 的两倍”，这个反问根本没问到点啊，firm characteristic是拿来排序的不是看绝对值的。

👍 赞



石川 (作者) 回复 吾与呜啦啦孰美

1 个月前

请先了解线性多因子模型里 factor loading 的概念.....



知乎

首发于  
川流不息

wxdf45119c3d056f8c 回复 石川 (作者)

26 天前

墨子9现在9

赞



吃不瘦的橘猫

1 个月前

川总问个问题，你的实证里第二步Fama-MacBeth回归的Rt，是下期收益还是本期的收益，另外收益是用日度收益还是月度收益呢？我记得Bara是用下一期的日度收益，但是cov会scale到月度。

赞



石川 (作者) 回复 吃不瘦的橘猫

1 个月前

下个月的月收益率。

赞



吃不瘦的橘猫 回复 石川 (作者)

1 个月前

谢谢，尝试复现一下

赞



刘玄

1 个月前

“对数市值或 P/B 这些指标，即使公司 A 的指标是公司 B 的两倍，也不能说 A 在该因子上获得的收益率就是 B 的两倍”。

不好意思，这段我也没看懂。如果这些指标本身就是因子荷载，且风险溢价是固定的，那么A在该因子上获得的超额收益是B的两倍，应该是对的。问题在于，风险溢价不是固定的，比如说在按照P/B的一个分段函数，那么说两倍就是错的。

赞

