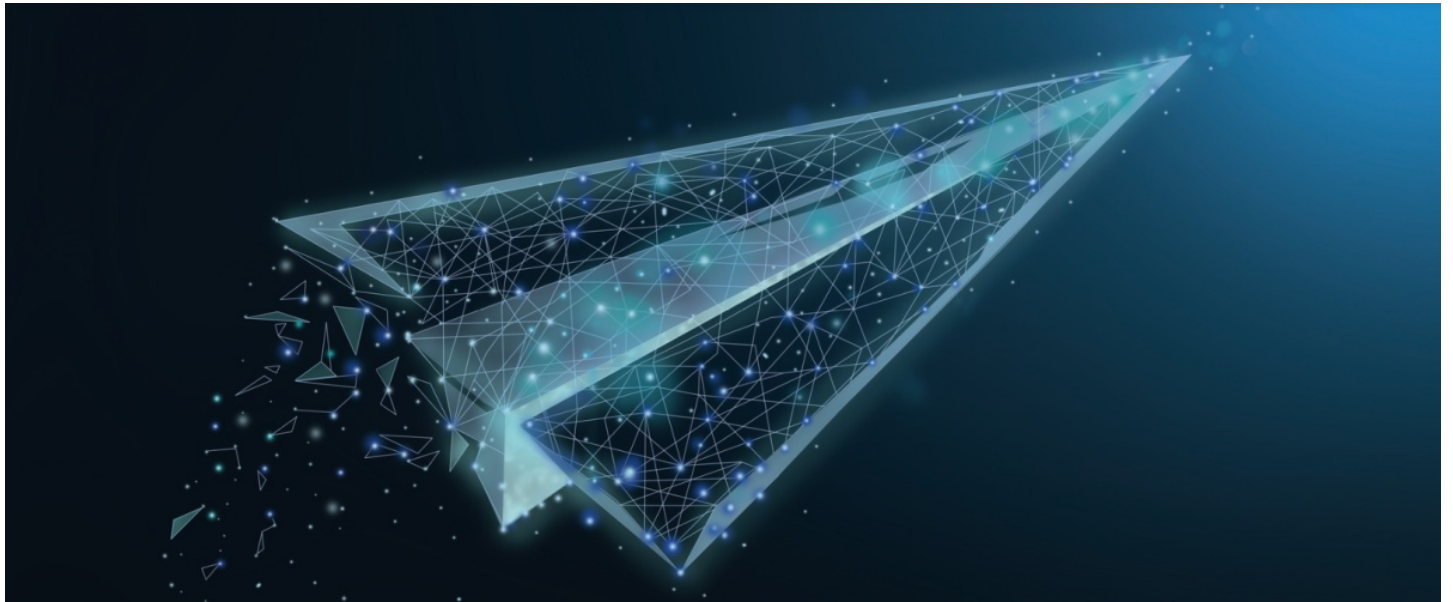


知乎

首发于
川流不息

实证研究 —— 使用正交化和自助法寻找显著因子



石川

量化交易 话题的优秀回答者

已关注

41 人赞同了该文章

摘要

本文在 A 股上复现 Harvey and Liu (2018) 提出的方法，使用正交化和自助法寻找显著的因子。

1 引言

《出色不如走运 (II)》一文曾介绍了 Harvey and Liu (2018) 提出的基于回归的因子有效性检验方法。**该方法避免了因 multiple testing 造成的运气成分、从而鉴别出能显著解释资产截面预期收益率差异的因子。**考虑到《出色不如走运 (II)》涉及的内容理论有余、实证不足，本文以中证 500 指数成分股为例，做一些实证分析，同时也指出在大 A 股中应用 Harvey and Liu (2018) 时遇到的一些坑。

首先简单回归一下 Harvey and Liu (2018) 提出的方法。

当大量因子**同时**被用做 empirical asset pricing 分析时，它们之中效果最显著的因子中一定包含了运气的成分。**Harvey and Liu (2018) 巧妙之处在于通过正交化和自助法 (Bootstrap) 得到了仅靠运气能够得到的显著性的经验分布；如果当排除了运气造成的显著性之后某个因子依然显著，那它就是真正有效的因子。**下图高度概括了该方法的逻辑（正交化和 Bootstrap 是核心）



知乎

首发于
川流不息

OX 无法解释资产（个股或投资组合）预期收益的截面差异；



使用每个正交化后的因子 OX 和一系列资产进行回归分析，找出 OX 中最能解释资产预期收益率截面差异的，这个显著性全部来自运气的成分；



通过 Bootstrap 采样进行大量上述实验、获得仅凭运气造成的显著性的经验分布，这就是在 null hypothesis（即仅靠运气）下显著性的分布；



如果来自真实数据中某原始因子 X 的解释程度在这个分布的 5%（或 10%、1%）阈值之下，则该因子真正有效。



重复上述步骤，直到找到全部有效因子。

知乎 @石川

值得一提的是，Harvey and Liu (2018) 提出的方法既可以用针对 empirical asset pricing 找出显著因子；也可以用于更一般的 predictive regression —— 即考察解释变量 X 能否预测被解释变量 Y。面对不同的问题，正交化和 Bootstrap 的核心思想是一致的，但在具体处理方法上存在差异。《出色不如走运 (II)》一文以 predictive regression 为例详细介绍了该方法，本文不再赘述。

以下行文将假设读者了解 Harvey and Liu (2018)。但是，我会针对 empirical asset pricing 问题，说明使用正交化和 Bootstrap 时的各种细节。这篇实证的目的更多的是介绍在 A 股上如何复现 Harvey and Liu (2018)。相反的，由于实证中的因子以及用来检验这些因子的一组资产，均无法避免的存在主观成分（本文最后一节会再聊聊这点），因此实证结果仅是示例性的。

2 Block Bootstrap

先来说说 Bootstrap。本来 Bootstrap 无需多言（需要背景知识的朋友请点[这里](#)），但是数据的特殊性决定了 Bootstrap 的特殊性。

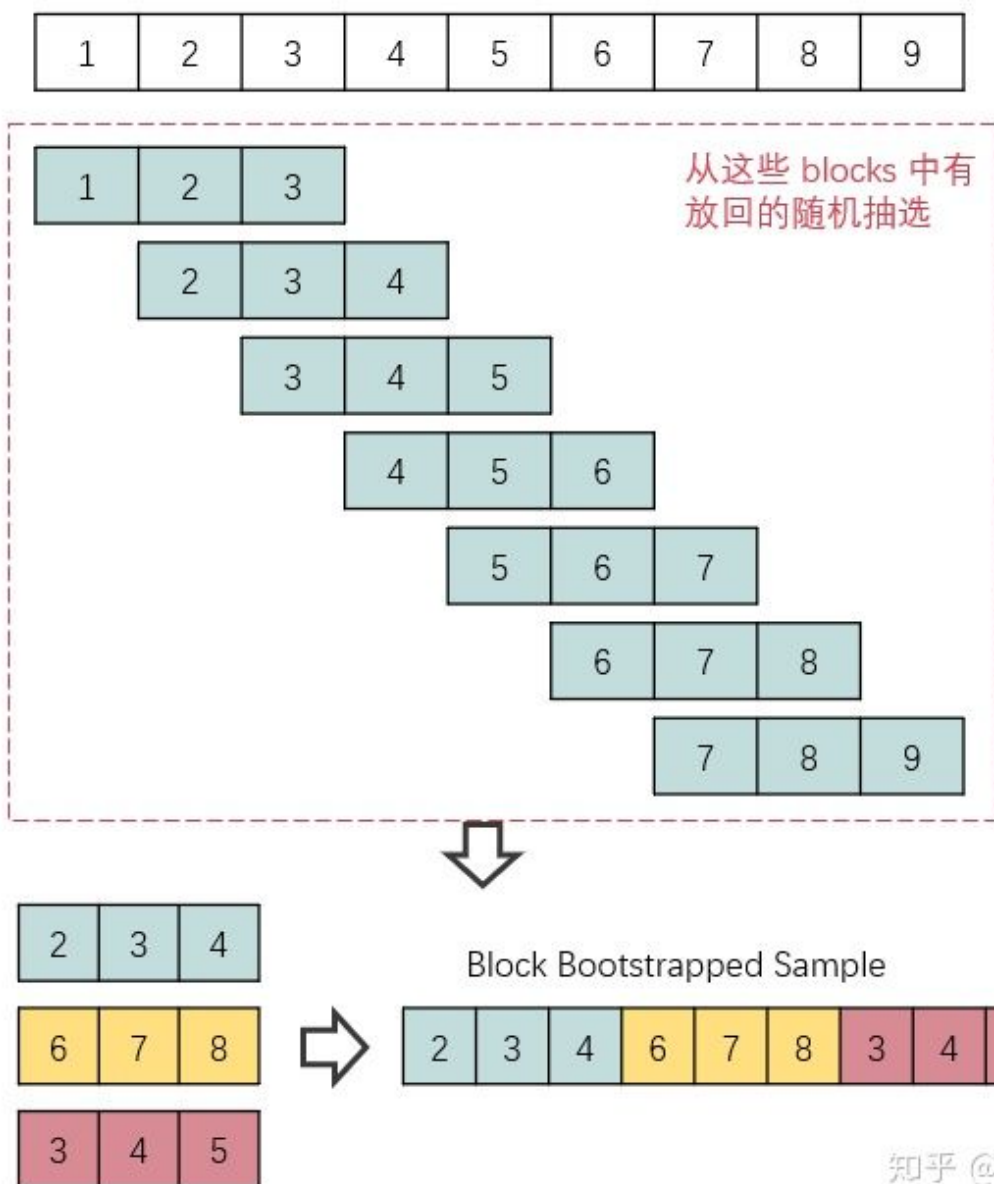


Bootstrap. 顾名思义, Block Bootstrap 就是每次从序列中有放回的抽取一个由连续 n 个相邻数据点构成的 block (大小由 block size 决定)。主流的 Block Bootstrap 算法包括以下三种:

1. Moving Block Bootstrap (Kunsch 1989, Liu and Singh 1992) ;
2. Circular Block Bootstrap (Politis and Romano 1992) ;
3. Stationary Bootstrap (Politis and Romano 1994) 。

下图说明了 Moving Block Bootstrap (MBB) 的原理。假设原始数据由 1 – 9 组成, 且令 block size = 3。MBB 依次以序列中的每个数字为起点构建长度为 3 的 blocks (本例中一共 7 个), 然后从这 7 个 blocks 中有放回的随机抽取, 直至构成和原始序列长度一样的 bootstrapped sample。

Moving Block Bootstrap



知乎 @石川

知乎

首发于
川流不息

个圆圈（Circular 一词的出处），然后再按照给定的 block size 进行重采样，避免首尾两端采样不足。

最后一种方法是 Stationary Bootstrap (SB)，它和前两者最大的区别是使用非固定的 block size。SB 中的 block size 满足几何分布；作为输入而给定的 block size 是它的期望。该方法得到的 bootstrapped 样本可以更好的满足平稳性的要求，因此当原始时间序列难以满足平稳性时有更好的效果。

本文的实证采用 stationary bootstrap，并在第四节说明原因。

3 因子模型和投资组合

Harvey and Liu (2018) 的目的是找到能够真正解释资产截面预期收益率差异的显著因子。因此在实证中，我们需要选定多因子模型，以及用来检验这些因子的一组资产。

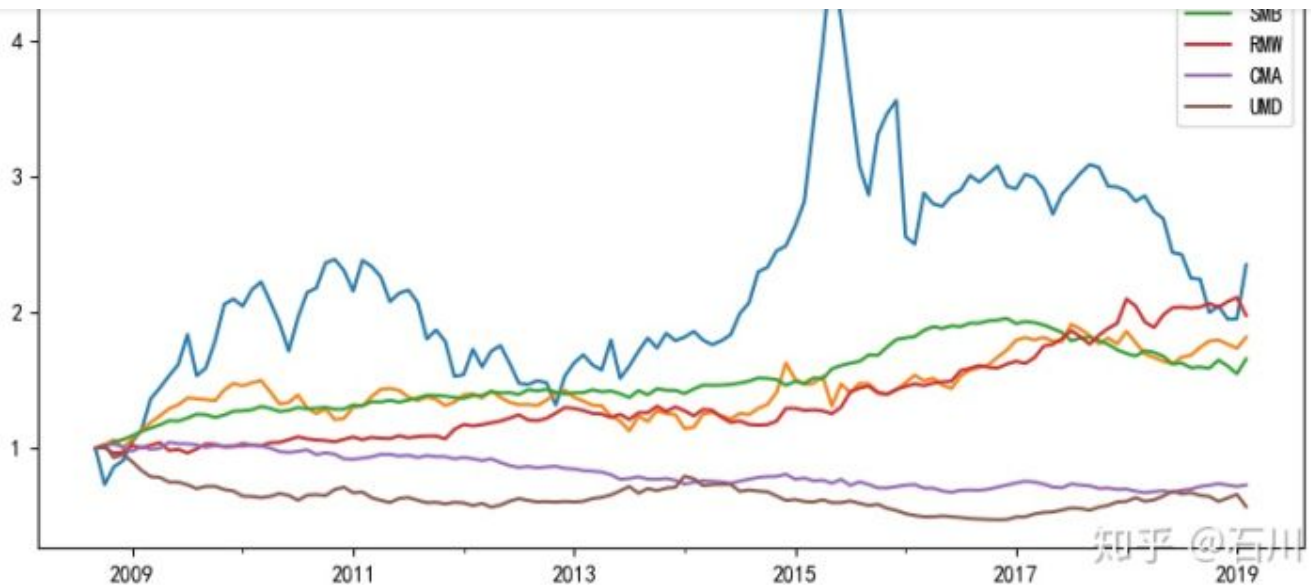
在因子方面，实证中选择 Fama and French (2015) 五因子以及 Carhart (1997) 的动量因子，一共六个因子：MKT、HML、SMB、RMW、CMA 和 UMD。构建这些因子的标的均为中证 500 的成分股；实证区间为 2008 年 10 月到 2019 年 2 月。每个月最后一个交易日排除停牌的股票，使用剩余成分股构建多空投资组合（多、空两边均等权配置），以此计算每个因子下个月的收益率，全部数据均来自 JoinQuant。具体的，这些因子定义如下：

- **MKT**：中证 500 指数的收益率，为了简化没有考虑其相对无风险利率的超额收益；
- **HML**：做多 Book-to-Price 最高的 150 支、做空 Book-to-Price 最低的 150 支；
- **SMB**：做多流通市值小的一半、做空流通市值大的一半；
- **RMW**：使用营业利润 TTM 与总市值之比为排序指标、从大到小排序，做多排名靠前的 150 支、做空排名靠后的 150 支；
- **CMA**：使用总资产增长率为排序指标、从小到大排序，做多排名靠前的 150 支、做空排名靠后的 150 支；
- **UMD**：使用 T - 12 到 T - 2 月之间（即过去 12 个月到上一个月）的累积收益率为排序指标，从大到小排序，做多排名靠前的 150 支、做空排名靠后的 150 支。

这六个因子在实证区间内的累积净值如下图所示。

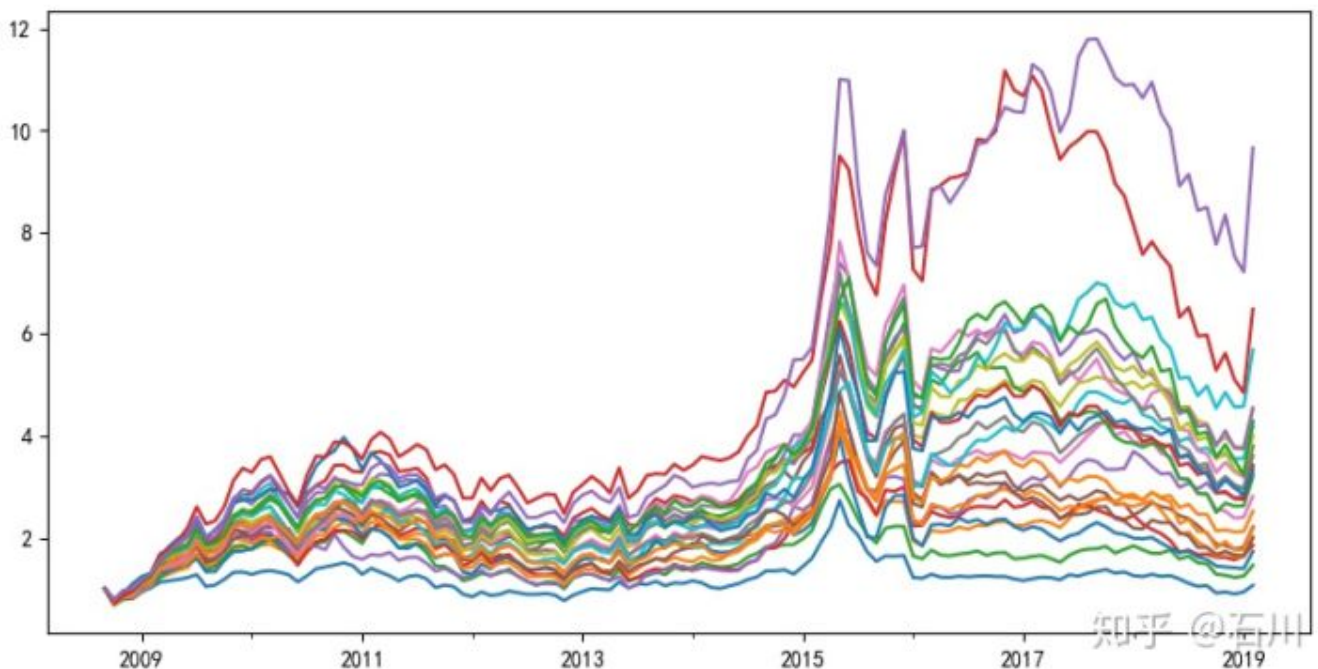


知乎

首发于
川流不息

下面再来看看用来检验因子的一组资产。由于多因子模型是用来给个股的超额收益定价的，因此最纯粹的实证是使用个股。考虑到本文更多的是为了说明如何复现 Harvey and Liu (2018)、实证结果仅是示例性的，因此我们也像 Harvey and Liu (2018) 一样，考虑使用 BP 和流通市值 double sort 得到的投资组合作为资产。

具体的，将中证 500 成分股按照 BP 和流通市值的大小各分成五档，并按照个股在两个指标上的档位取值将它们归类于 $5 \times 5 = 25$ 个投资组合中的某一个，由此构建 25 个纯多头投资组合。这 25 个投资组合在实证区间内的累积净值如下图所示。



下面用这 25 个投资组合检验上述六个因子。



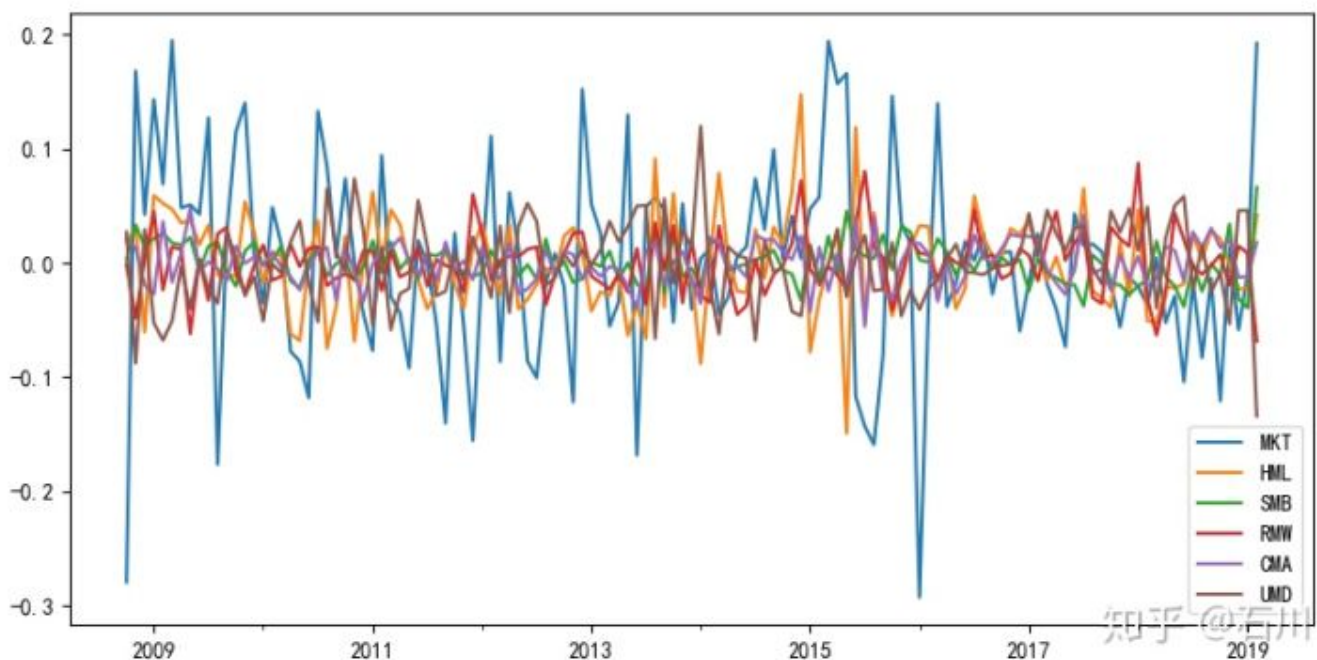
知乎

首发于
川流不息

在检验有效因子时，Harvey and Liu (2018) 的思路是反复使用正交化 + 自助法，逐一找到显著的因子，直到所有剩余因子中没有任何显著的。下面就来进行这个过程。

在第一轮中，这六个因子是否显著都是未知的。首先对它们进行正交化处理。**正交化的目的是排除因子对资产截面预期收益差异的解释能力；正交化的手段是对因子收益率序列在时序上去均值 (demean)。**

时序上 demean 不改变这些因子对于资产收益率波动的解释力度（时序回归的 R-squared 不变），但由于 demean 之后因子收益率期望为零，**使用 demean 因子对资产进行时序回归得到的截距恰好等于资产本身的时序均值，说明 demean 因子在截面上不再具备解释能力。**这六个因子 demean 之后的时序如下图所示。



由于 A 股上旗帜鲜明的牛熊市，demean 之后的 MKT 因子（以及其他一些因子）难以满足平稳性，如不加以处理，将会影响 Harvey and Liu (2018) 的使用。该方法的本质是 demean 之后的因子无法解释资产预期收益率的截面差异（能解释的都是运气），因此在 bootstrap 时应保证 bootstrapped 样本中这些 demean 之后的因子的收益率依然为零。

如果使用 MBB，由于对原始序列两端采样不足，则会使 bootstrapped 样本中 demean 后因子的收益率均值对于实证窗口的选择异常敏感。例如上图中，原始序列两端 demean 之后的 MKT 收益率一个很小、一个巨大，如果采样不足会影响 bootstrapped 样本。为此，在实证中选择将原始序列首尾相连的 Stationary Bootstrap。这是在 A 股中复现 Harvey and Liu (2018) 时的第一个坑。

下面再来说说第二个。

知乎

首发于
川流不息

味着，哪怕不 demean，由于 $E[MKT] \approx 0$ ，MKT 因子对于资产截面收益率的解释力度也十分有限。

对于一个真正有效的因子，demean 的目的是排除它的截面解释能力，把它暂时视作一个随机因子（随机因子的期望收益为零），从而仅考察由于多个因子 multiple testing 中的运气成分能造成的解释力度。不幸的是，如果选择从 2010 年 1 月开始的实证区间，我们这位 MKT 因子都不需要 demean 就 $E[MKT] \approx 0$ 了，剩下再怎么检验，它也难言有效，从而造成错误的结论。这就是第二个坑。

OK，坑挖完了也填完了，下面继续实证。

分别使用这六个因子和 25 个投资组合收益率做时序回归。对于每个因子，得到 25 个回归截距，它们是这些投资组合在使用该因子时的定价错误。使用这 25 个截距绝对值的中位数作为该因子解释这些投资组合的能力。定价错误低意味着解释力度高，因此该中位数越小说明因子越有效。

此外，为了计算仅凭运气能够获得的显著性，将原始 25 个投资组合的收益率序列和 demean 之后六个因子的收益率序列放在一起（ $T \times 31$ 阶矩阵），令 block size = 4，进行 1000 次 stationary bootstrap，得到 1000 个 bootstrapped 样本。对于每个样本，分别使用每个 demean 因子对这 25 个投资组合进行时序回归、得到 6 个截距绝对值中位数，将这 6 个中位数中的最小值作为 test statistic，它就是在这个 bootstrapped 样本中，仅靠运气能够获得的最低定价错误。

从 1000 个 bootstrapped 样本中得到 1000 个 test statistic，便构成了它的分布，这就是仅靠运气能够得到的定价错误分布。使用该分布 5% 分位数作为阈值，如果原始因子获得的定价错误小于该阈值，我们就说该因子在 5% 的显著性水平下有效。

在第一轮中，这六个因子对 25 个投资组合定价错误绝对值的中位数、以及使用 1000 次 stationary bootstrap 获得的 test statistic 阈值如下表所示。不难看出，MKT 的解释力度最高且小于阈值，因此 MKT 被选为第一个有效的因子。第一轮正交化 + 自助法结束。



知乎

首发于
川流不息

MKT	0.3581 %
HML	1.3370 %
SMB	0.6231 %
RMW	2.0870 %
CMA	1.4430 %
UMD	1.1134 %

Bootstrap 得到的 5% 定价错误阈值: 1.2839 % 知乎 @石川

下面来看第二轮（后面的以此类推）。

在第二轮中，已选出的 MKT 将会出现在时序回归方程的右侧，这么做是为了考察剩余待检验因子对于解释资产预期收益截面差异的增量贡献。首先，对于 SMB、HML、RMW、CMA、UMD 的每一个，将它们和已经选定的 MKT 一起（如果是第 n 轮正交化 + 自助法，则使用前 $n - 1$ 轮中选出的 $n - 1$ 个因子），对那 25 个投资组合进行时序回归，得到的截距项绝对值的中位数作为这五个因子的解释力度。

接下来，使用 MKT（again，如果是第 n 轮正交化 + 自助法，则使用前 $n - 1$ 轮中选出的全部因子）对待检验因子正交化。具体做法为，将待检验因子依次放在回归方程左侧，将已经选出的因子放在回归方程的右侧，时序回归得到截距项；**使用原始待检验因子减去截距项（也是一种 demean 处理），得到的就是其正交化之后的因子。**该方法保证了使用 demean 的待检验因子和已选出因子一起对资产时序回归时得到的截距和仅使用已选出因子对资产时序回归时得到的截距相同，从而说明 demean 后的待检验因子无增量贡献。

将正交化后的五个因子和原始 MKT 因子一起，和 25 个投资组合收益率序列放在一起，构成一个 $T \times 31$ 阶矩阵。对其并进行 1000 次 stationary bootstrap 的到 1000 个 bootstrapped 样本。对于每一个样本，**逐一使用 demean 后的因子，将它和已选出的因子一起（这里是 MKT），对这 25 个投资组合回归，得到截距绝对值中位数；**然后从这五个因子中挑出中位数最低的，它就是该 bootstrapped 样本的 test statistic。这 1000 个 bootstrapped 样本的 test statistic 取值便构成了它的分布。

下表是第二轮中这五个因子的解释能力以及 test statistic 的 5% 分位数阈值。其中，SMB 的解释力度最高且小于阈值，因此 SMB 被选为第二个有效因子。第二轮正交化 + 自助法结束。



因子	25 个投资组合定价错误绝对值中位数
HML	0.2726 %
SMB	0.2419 %
RMW	0.3426 %
CMA	0.3637 %
UMD	0.3231 %

Bootstrap 得到的 5% 定价错误阈值：0.3279 % 知乎 @石川

按上述说明，反复使用正交化 + 自助法，直至没有显著的因子。在第三轮中的结果如下，RMW 因子被选出。

已入选因子 MKT、SMB

因子	25 个投资组合定价错误绝对值中位数
HML	0.2007 %
RMW	0.1913 %
CMA	0.2979 %
UMD	0.2929 %

Bootstrap 得到的 5% 定价错误阈值：0.2289 % 知乎 @石川

第四轮的结果如下，剩余三个因子的解释力度均无法超过 test statistic 阈值，因此全部被拒绝、检验结束。最终选出的因子为：MKT + SMB + RMW。

已入选因子 MKT、SMB、RMW

因子	25 个投资组合定价错误绝对值中位数
HML	0.1816 %
CMA	0.1811 %
UMD	0.1986 %

Bootstrap 得到的 5% 定价错误阈值：0.1649 % 知乎 @石川

以上就是对 Harvey and Liu (2018) 的实证研究。

知乎

首发于
川流不息

Harvey and Liu (2018) 这篇文章十分 technical。我在《出色不如走运 (II)》对它的理论部分做了介绍，但是不甚满意，因此便有了这篇实证，希望这篇没那么晦涩（但我觉着也挺晦涩的.....）。想对它进一步加深理解的小伙伴，我推荐翻墙去 YouTube 上看 Prof. Harvey 在 Jacobs Levy Center 年会上题为 Lucky Factors 的演讲，它能帮我们跳出细节、从全局掌握该方法的实质。



从我的实证体会来说，最难的不是方法本身，而是如何尽可能客观的选择多因子模型以及用来检验该模型的资产。

在 empirical asset pricing 的研究中，最著名的文章当属 Fama 和 French 的一系列文章。无疑，这和他们开创性的研究密不可分，但不容忽视的一点是，人家 marketing 做的好！Ken French 在其网站上定期更新因子收益率序列供所有人免费下载。**这些高质量且被市场广泛认可的数据为美股研究的蓬勃发展奠定了坚实的基础。**



知乎

首发于
川流不息

Changes in CRSP Data

Fama/French 3 Factors [TXT](#) [CSV](#) [Details](#)Fama/French 3 Factors [Weekly] [TXT](#) [CSV](#) [Details](#)Fama/French 3 Factors [Daily] [TXT](#) [CSV](#) [Details](#)Fama/French 5 Factors (2x3) [TXT](#) [CSV](#) [Details](#)Fama/French 5 Factors (2x3) [Daily] [TXT](#) [CSV](#) [Details](#)

Univariate sorts on Size, B/M, OP, and Inv

Portfolios Formed on Size [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Size [ex.Dividends] [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Size [Daily] [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Book-to-Market [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Book-to-Market [ex. Dividends] [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Book-to-Market [Daily] [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Operating Profitability [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Operating Profitability [ex. Dividends] [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Investment [TXT](#) [CSV](#) [Details](#)Portfolios Formed on Investment [ex. Dividends] [TXT](#) [CSV](#) [Details](#)

Bivariate sorts on Size, B/M, OP, and Inv

6 Portfolios Formed on Size and Book-to-Market (2 x 3) [TXT](#) [CSV](#) [Details](#)6 Portfolios Formed on Size and Book-to-Market (2 x 3) [ex. Dividends] [TXT](#) [CSV](#) [Details](#)6 Portfolios Formed on Size and Book-to-Market (2 x 3) [Weekly] [TXT](#) [CSV](#) [Details](#)6 Portfolios Formed on Size and Book-to-Market (2 x 3) [Daily] [TXT](#) [CSV](#) [Details](#)25 Portfolios Formed on Size and Book-to-Market (5 x 5) [TXT](#) [CSV](#) [Details](#)25 Portfolios Formed on Size and Book-to-Market (5 x 5) [ex. Dividends] [TXT](#) [CSV](#) [Details](#)25 Portfolios Formed on Size and Book-to-Market (5 x 5) [Daily] [TXT](#) [CSV](#) [Details](#)100 Portfolios Formed on Size and Book-to-Market (10 x 10) [TXT](#) [CSV](#) [Details](#)100 Portfolios Formed on Size and Book-to-Market (10 x 10) [ex. Dividends] [TXT](#) [CSV](#) [Details](#)100 Portfolios Formed on Size and Book-to-Market (10 x 10) [Daily] [TXT](#) [CSV](#) [Details](#)

反观 A 股，很难找到大家都认可的标准化因子收益率数据，这造成不同人对于哪怕是对同一因子的研究结果都有可能大相径庭。（就我有限的所知，中央财经大学维护了针对全 A 股的 Fama-French 五因子 + Carhart 动量因子数据。）究其原因是因为处理因子数据背后的细节太过纷繁——按什么频率调仓；具体使用什么指标（比如当期还是 TTM）；依何种权重配置多空组合；如何处理 ST、涨跌停、停牌、退市等股票。

但即便有各种困难，为了在 A 股上得到更加客观、公允的实证分析结果（无论某个因子是有效还是无效都是有价值的发现），构建标准化因子数据的努力都必不可少。（最近，我在构思写一篇比较学术界主流多因子模型在 A 股上的效果分析文章，但是第一步就是要获得客观、标准的因子收益率。）希望未来能有更多的你、我（特别是市面上的各种量化平台）能为此努力，这将是造福量化多因子研究的大功绩。

参考文献

知乎

首发于
川流不息

- Fama, E. F. and K. R. French (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, Vol. 116(1), 1 – 22.
- Harvey, C. R. and Y. Liu (2018). Lucky Factors. Working paper, available at SSRN: ssrn.com/abstract=25287...
- Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, Vol. 17(3), 1217 – 1241.
- Liu, R. Y. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, eds.), 225 – 248. Wiley, New York.
- Politis, D. N. and J. P. Romano (1992). A circular block resampling procedure for stationary data. In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, eds.), 263 – 270. Wiley, New York.
- Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association*, Vol. 89(428), 1303 – 1313.

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/版权分发](#)）为进行维权行动。

编辑于 2019-07-03

[多因子模型](#)[资产定价模型 \(CAPM\)](#)[资产定价](#)[▲ 赞同 41 ▼](#)[● 25 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



量信 ↑ 14



【量化策略】基于Optimal Control Theory的高频做市



《高频因子掘金》——开盘收盘成交量占比因子

底是什么？有什么作用？顾老师

BigQuant

25 条评论

切换为时间排序

写下你的评论...



文心

5 个月前

你这篇逻辑很清晰，写的很好呀。

👍 赞



石川 (作者)

5 个月前

谢谢亲哈。我怕我写着写着就 get lost。。。

👍 赞



万善乐

5 个月前

其实有想过一个因吹思丁的事情就是

$y = \beta_1 * f_1 + res_1$

$res_1 = \beta_2 * f_2 + res_2...$

使得每个beta都显著直到 $res_x \sim N(0,1)$ 233333

👍 1



吴二真

5 个月前

没太看懂 你这个对X正交化后的OX怎么算出来的？能不能配上公式或者代码...？

👍 赞



石川 (作者) 回复 吴二真

5 个月前

回归，"Demean"

👍 赞



吴二真 回复 石川 (作者)

5 个月前

demean的中文是中心化，回归是指对下期收益率做回归吗？然后回归之后是取残差？

👍 赞

知乎

首发于
川流不息

这篇working paper 14年就读过了，当时主要的篇幅是在multiple testing上，还算是在金融领域的第一次尝试。现在文章已经被改成这个样子了...

赞



石川 (作者) 回复 lakylu

5 个月前

[飙泪笑]

赞



docdukun

5 个月前

川总您好，关于这篇文章有两个问题：

1 统计结果对Block Size的不同取值敏感吗？

2 对于MBB法举例中Block Size=3占到样本总量9的三分之一；后面的实证中用SB法，Block Size=4是不是太小了？还是我理解错了，实证中的Block Size不是指收益率时间序列的长度？

赞



石川 (作者) 回复 docdukun

5 个月前

好问题。就实证来看，因为A股太短且波动太大这个尿性，实证结果对参数还是挺敏感，选不好就会得到所有因子都不显著的结论，所以这篇文章最大的目的是讲怎么复现这个方法。对于block size，指的是每次抽取取几个数（普通自助法每次抽取从原来序列中抽一个数，相当于block size = 1）。我计算月频收益率的自相关性其实没那么高，因此block size就不需要太大。

赞



石川 (作者) 回复 docdukun

5 个月前

最敏感的是实证区间内无论是因子还是检验因子的资产的收益率均值是不是不为零.....如果大家都接近零，那么无法区分有效因子或随机因子，因为随机因子天生就是零。另外就是数据真的是越长越好，十年回测太短了。多少是一种无奈吧。

赞

展开其他 1 条回复



孟祥胜

5 个月前

哥们说了这么多有没有产品能买的啊！

赞



墨碎

4 个月前

求问中财的FF数据库网址在哪~

知乎

首发于
川流不息

石川 (作者) 回复 墨碎

4 个月前

现在不知为何404了。。。原来的地址是：sf.cufe.edu.cn/kxyj/kyj...

赞



墨碎 回复 石川 (作者)

4 个月前

多谢，我找到了这个sf.cufe.edu.cn/info/119...

赞



梁山伯

4 个月前

看了文章感觉很受用，非常感谢。文章中提到采用中证500的成分股，实证区间为 2008 年 10 月到 2019 年 2 月。我粗略算了一下2008年的成分股与2019年的成分股只有97只是重叠的，换句话说大部分的成分股都发生了变化，您在计算过程中这部分是怎么考虑的？

赞



石川 (作者) 回复 梁山伯

4 个月前

每个月末挑当期成分股计算下个月的投资组合。

1



梁山伯

4 个月前

文章中提到“从 1000 个 bootstrapped 样本中得到 1000 个 test statistic，便构成了它的分布，这就是仅靠运气能够得到的定价错误分布。**使用该分布 5% 分位数作为阈值，如果原始因子获得的定价错误小于该阈值，我们就说该因子在 5% 的显著性水平下有效。**”但是我阅读了原文，并没有看到这个提法，而且在Table 3中只有Bootstrap后 Median of min的值，是不是原文的意思是小于最小值的中位数，而不是5%的阈值，作为判断依据。

赞



石川 (作者) 回复 梁山伯

4 个月前

最新 2018 版文章里 34 页有这么一段：Panel A of Table II shows the results on multiple testing. In particular, the bootstrapped 5th percentile of SI_{ew}^{med} for the minimum statistic is -14.7%. By evaluating the 20.6% reduction against the empirical distribution of the minimum statistic for multiple testing, the p-value is 0.002. Therefore, the multiple-test p-value is below the 5% cutoff. We therefore also declare the market factor significant from a multiple testing perspective.

1



梁山伯 回复 石川 (作者)

4 个月前

使用该分布 5% 分位数作为阈值,始终得不到您文章中提到的1.2839%，比这个值要低得多。能不能再详细说明下计算方法？





首发于
川流不息



什么是当当当当当

4 个月前

这个方法其实就是White's Reality Check，只是多加了一步中性化吧？利用bootstrap来判断是否是lucky在很多论文中看到过。不过这个方法来挑选已知的定价因子还行 筛选挖掘出来的alpha因子的话 不停的对前面的因子中性化得到的残差有点没逻辑 会选出一些奇怪的因子 不知道有什么改进的方法

👍 赞 ↩ 回复 🦶 踩 🚩 举报



石川 (作者) 回复 什么是当当当当当

4 个月前

厉害了！

👍 赞



什么是当当当当当 回复 石川 (作者)

4 个月前

不知道川总有什么建议[思考]

👍 赞

