

知乎

首发于
川流不息

所有历史数据都是样本内



石川

量化交易 话题的优秀回答者

已关注

babyquant、量化投资机器学习、黑猫Q形态、王坚等 127 人赞同了该文章

摘要

我们构建策略靠的不是站在回测起点往后看的先见之明，而是站在回测终点往前看的后见之明。所有历史数据都是样本内。

1 引言

金融领域的小伙伴想必对 Institutional Investor Journals (II Journals) 并不陌生。它旗下有很多影响力很高的期刊，其中最著名的大概是 Journal of Portfolio Management。与 JF、JFE、RFS 这些纯学术类期刊不同，II Journals 的期刊更注重实践，深得业界喜爱；而 Fischer Black、Robert Engle、Daniel Kahneman、Andrew Lo、William Sharpe、Robert Shiller 等大咖也均在 II Journals 的期刊发表过研究。

2018 年 II Journals 更名为 Institutional Portfolio Research Journals (IPR Journals)，并于 2019 年为旗下的期刊系列增加了一位最新成员 —— Journal of Financial Data Science。该刊的 Editors 和 Advisory Board 可谓十分豪华（下图）。



知乎

首发于
川流不息

EDHEC Business School

AQR Capital Management

Natixis Investment Managers

MANAGING EDITOR

FRANCESCO A. FABOZZI
Frank J. Fabozzi Associates

ADVISORY BOARD

IRENE ALDRIDGE
Cornell University and AbleMarketsROBERT ARNOTT
Research AffiliatesJOSEPH CERNIGLIA
University of PennsylvaniaRAMA CONT
Oxford UniversityGERMÁN G. CREAMER
Stevens Institute of TechnologySANJIV RANJAN DAS
Santa Clara UniversityCAMPBELL HARVEY
Fiqua School of Business
Duke UniversityMICHAEL IMERMAN
Peter F. Drucker and Masatoshi Ito
Graduate School of Management
Claremont Graduate UniversityBRUCE JACOBS
Jacobs Levy
Equity ManagementRON KAHN
BlackRock
Financial ManagementKATHRYN KAMINSKI
AlphaSimplex Group LLCHOSSEIN KAZEMI
CISDM/Isenberg School of Management
and University of Massachusetts
at AmherstPETTER KOLM
Courant Institute, New York UniversityHAN LIU
Northwestern UniversityANTHONY LEDFORD
Man-AHLJIM KYUNG-SOO LIEW
Johns Hopkins UniversityPAUL F. MENDE
MIT Sloan School of ManagementWARREN PENNINGTON
VanguardSIDNEY C. PORTER
FEV Analytics

知乎@石川

随着大数据和机器学习算法在金融和投资中的潜在作用越来越大，这本期刊的诞生无疑是非常及时的。它旨在指导金融领域的实践者正确使用与日俱增的数据和日新月异的技术。如今，该刊第一卷第一期已经发行。

作为处女刊，它里面包含了很多非常有意思的文章。其中最吸引我的是一篇题为 A Backtesting Protocol in the Era of Machine Learning（机器学习时代的回测协议）。它的作者是 Rob Arnott、Campbell Harvey 以及 Harry Markowitz（三位都无需介绍）。在这篇文章中，三位作者从七个方面提出了新时代下进行策略回测时应该遵循的一些规则，从而最大化的降低数据挖掘和过拟合的影响。这七方面构成了一个完整且可操作的体系，能够帮助我们更好的规避样本内的虚假信号、找出能在样本外更有效的交易策略。

这七方面的每一块都值得好好研读一番（建议感兴趣的小伙伴找来 Arnott, Harvey, and Markowitz 2019 这篇文章读一读）。在今天的文章中，我仅想对第四部分 Cross-Validation 里面的两个论点谈一些想法。在这一部分，三位作者抛出了两个观点：

1. Out of Sample is Not Really Out of Sample.

2. Iterated Out of Sample is Not Out of Sample.

这两句话结合起来再翻译成中文就是所有历史数据都是样本内。我对此深以为然。



知乎

首发于
川流不息

针对 Out of Sample is Not Really Out of Sample 这个观点, Arnott, Harvey, and Markowitz (2019) 写到:

Researchers have lived through the hold-out sample and thus understand the history, are knowledgeable about when markets rose and fell, and associate leading variables with past experience. As such, no true out-of-sample data exist; the only true out of sample is the live trading experience.

对于预留出来的样本外数据, 它们也是已经发生过的历史。尽管模型将它们视作样本外, 但由于我们经历了那段历史, 知道在那段时间市场是如何发展的(什么时候涨、什么时候跌), 会不可避免的把领先变量和这些历史经验联系起来。正是由于这个原因, 历史数据中并没有真正的样本外, 只有实盘交易才是真正的样本外。

举个例子。在因子选股中, 我们通常会用 PE、ROE、EPS 等指标来选股, 并按照各种逻辑——比如动量或者估值——进行难如登天的因子择时。在择时的时候, 我们会非常小心的使用滚动窗口外推, 利用 point in time 数据来避免 look-ahead bias (前视偏差)。但无论是在全部回测期内多因子选股, 还是使用滚动窗口在“样本外”做因子择时, 这背后无法否认的事实是我们已经知道 PE、ROE、EPS 在过去整段历史时间内是有效的!

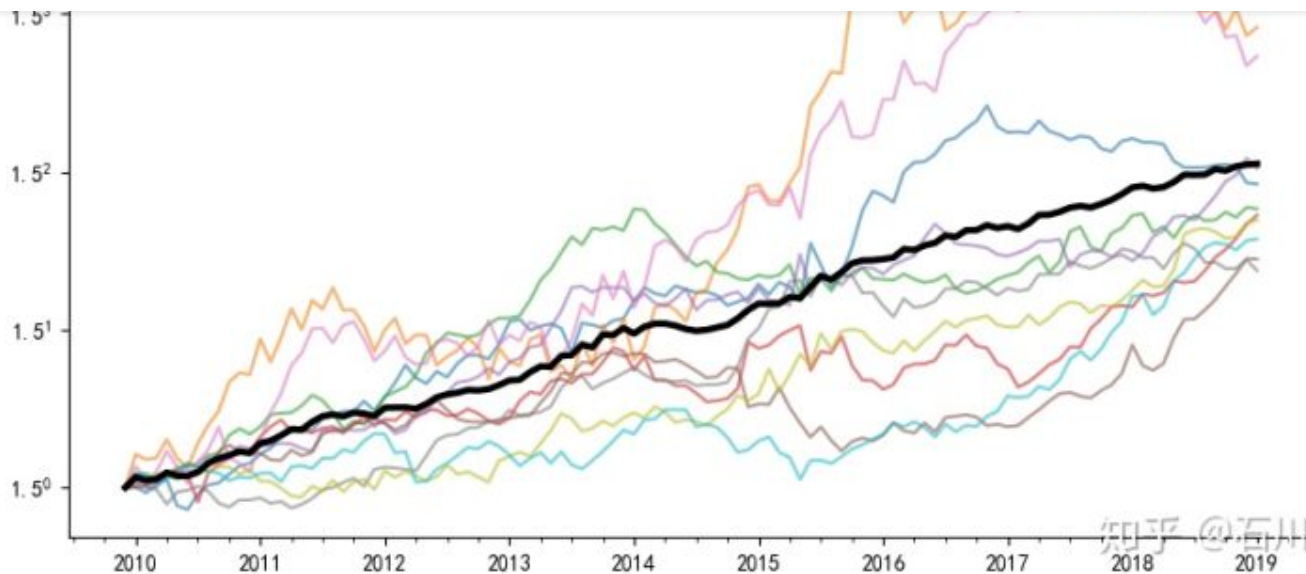
我们并不是在 10 年前就预见到了 PE、ROE、EPS 会有效并在过去的 10 年里使用它们; 而是在已经知道市场在过去 10 年是如何展开的前提下, 通过 empirical data analysis 找到了 PE、ROE、EPS 这些好使的因子然后再去对着它们回测。扪心自问一下, 如果一个因子在回测期内的表现是近似随机的, 我们会用它来做因子择时吗? 这当然不是说它们并没有业务依据, 而只是想强调正是因为历史已经发生了, 因此它不是真正的样本外。

如果在样本内找到的因子在真正的样本外无效会在多大程度上削弱策略的表现呢? 来看下面的实验。

假设回测期从 2010 年 1 月 1 日到 2019 年 1 月 31 日; 以中证 500 的成分股为股票池。针对每个因子, 使用其首尾各 10% 的股票按照等权构建多、空投资组合, 以对冲后的组合的收益率作为因子收益率, 每月末调仓且不考虑任何费用。下图显示了依照上述说明后找到 10 个回测期内年化夏普率在 0.8 到 1.0 之间的因子的累积净值(这些因子均来自基本面或量价信息、月收益率的平均相关系数为 0.1)。图中黑色粗线表示通过等权配置这些因子所构成的多因子策略的净值。该多因子策略的夏普率高达 2.56, 年化收益率 9.58%。

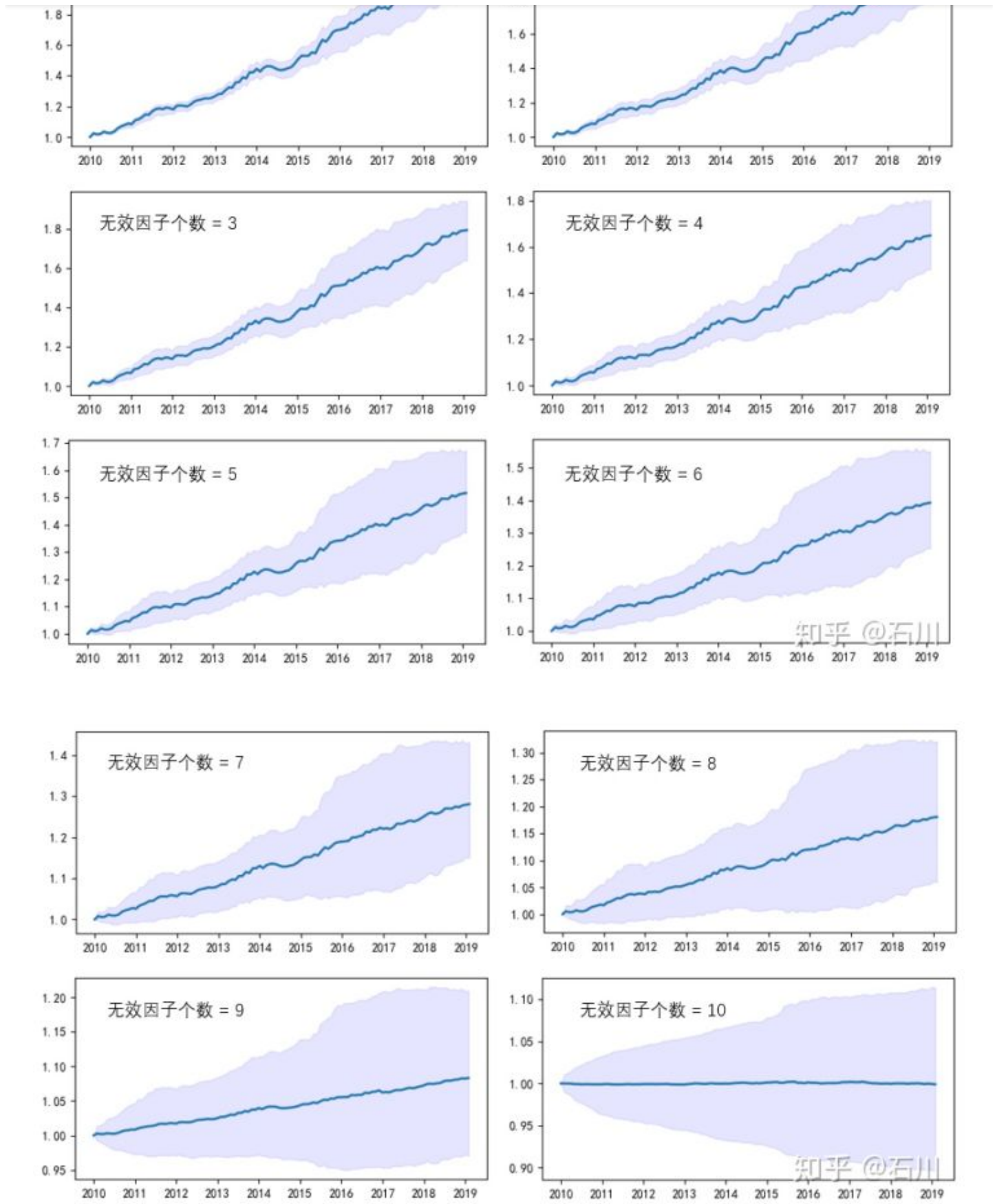


知乎

首发于
川流不息

下面假设这 10 个因子中的 n 个 (n 取值从 1 到 10) 在真实样本外完全失效——即样本外该因子实际上无法预测收益率。为了模拟“真实样本外完全失效”，我们从上述 10 个因子中随机选出 $10 - n$ 个作为有效因子，并随机生成 n 个因子来模拟那些失效的因子，以此进行一次完整回测。为了得到这 10 个混合因子策略表现的分布，将上述过程进行 3000 次仿真，并考察回测期内每个时间点该策略累积净值的中位数、5% 分位数以及 95% 分位数。下图依次展示了当 n 从 1 取到 10 时，混合多因子策略的累积净值。



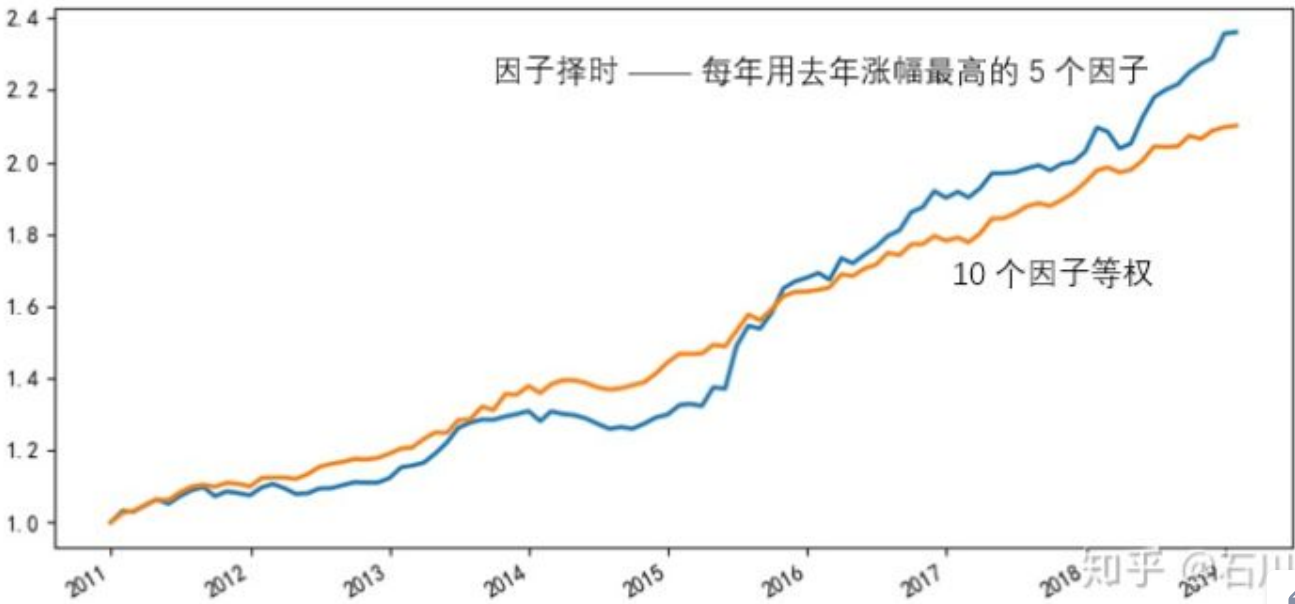


从上述结果不难看出，随着无效因子个数的增加，策略表现的“中枢”（中位数）逐渐下移，确定性（体现为 5% 和 95% 分位数的净值）逐渐增大。在整个回测期末，上述 10 种 n 取值的策略表现总结于下表。对于样本内有效的 10 个因子，如果它们在真正的样本外全部失效（即无

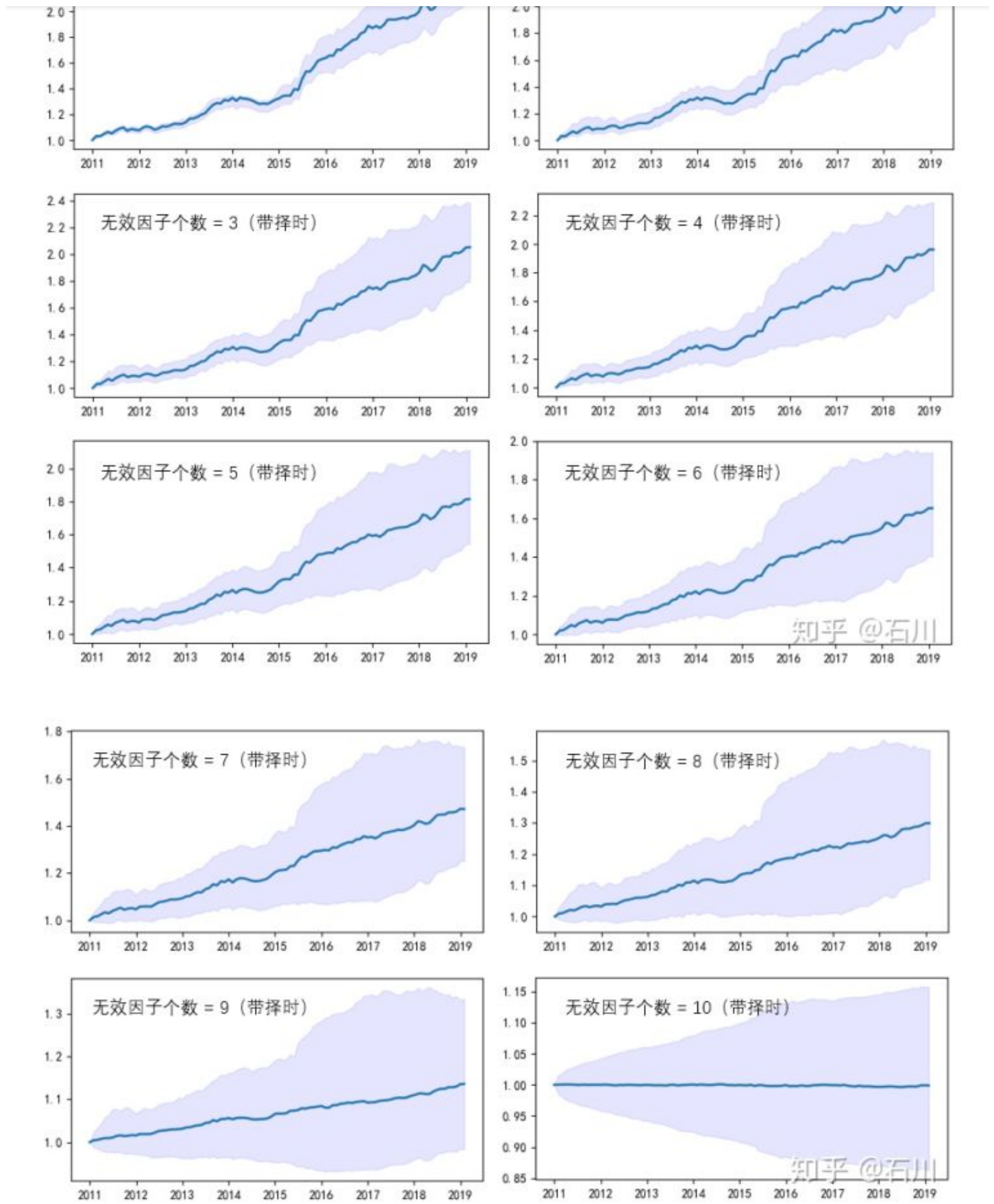
整个回测期内混合 10 因子策略的年化收益率 (%)

无效因子数	5% 分位数	中位数	95% 分位数	5% 和 95% 分位数的收益率 spread
1	7.86	8.63	9.14	1.28
2	6.70	7.63	8.41	1.71
3	5.58	6.64	7.59	2.01
4	4.58	5.66	6.70	2.12
5	3.53	4.69	5.83	2.30
6	2.51	3.71	4.95	2.44
7	1.55	2.77	4.04	2.49
8	0.63	1.84	3.11	2.48
9	-0.33	0.88	2.12	2.45
10	-1.20	-0.01	1.19	2.39

依照上述实验思路也可以对一些常见的因子择时思路进行验证。下面以主流的按因子动量择时，每年配置过去一年涨幅最高的五个因子。使用全部 10 个样本内有效的因子，加之上述择时策略，得到的效果如下，因子择时年化收益率为 11.21%，超过了 10 个因子等权配（夏普率没有超过）。



同样的，假设 n 个因子在真实样本外失效，来看看对该择时策略的影响。



知乎

首发于
川流不息

无效因子数	5% 分位数	中位数	95% 分位数	5% 和 95% 分位数的收益率 spread
1	9.44	10.39	11.64	2.20
2	8.43	9.87	11.32	2.89
3	7.49	9.28	11.34	3.84
4	6.58	8.69	10.79	4.20
5	5.50	7.64	9.67	4.17
6	4.27	6.40	8.54	4.27
7	2.80	4.89	7.05	4.25
8	1.40	3.29	5.44	4.04
9	-0.21	1.59	3.64	3.85
10	-1.81	-0.01	1.81	3.62

知乎@石川

和不择时相似，随着失效因子个数的增加，策略中位数表现下移、不确定性加大。如果把择时策略和之前的不择时策略横向比较一下（即考虑给定失效因子个数 n 下两个策略的表现），可以发现有意思的结果：样本内有效的择时策略在样本外的中枢同样优于不择时的策略，但代价是更高的波动；这个高波动是由因子失效和择时失效两方面的不确定性共同贡献的。

3 Iterated Out of Sample is NOT Out of Sample

再看看 Iterated Out of Sample is Not Out of Sample。对此，Arnott, Harvey, and Markowitz (2019) 写到：

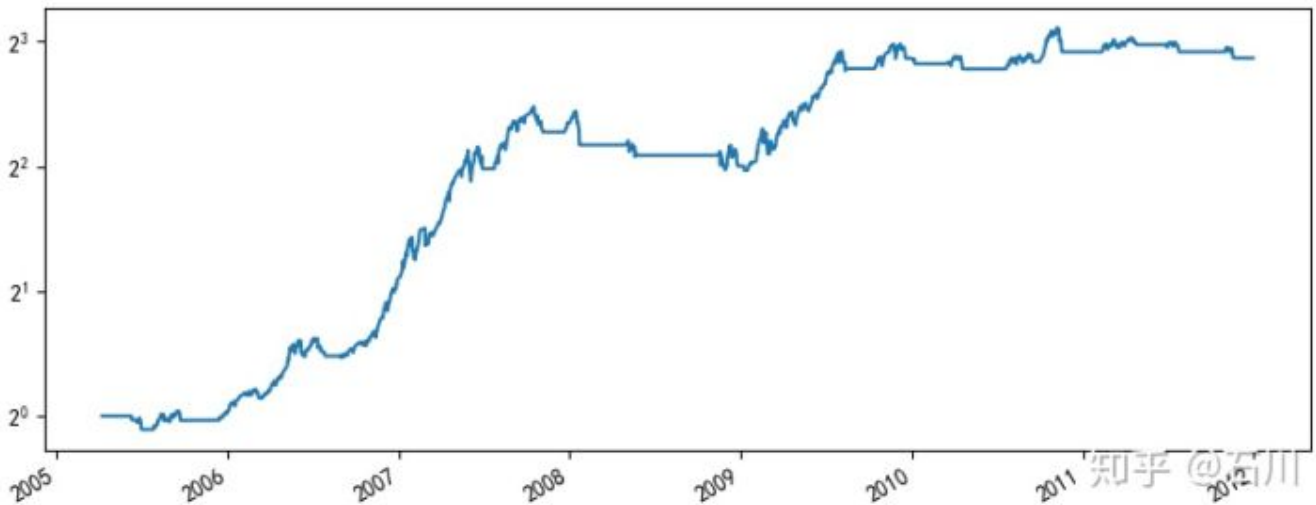
*Suppose a model is successful in the in-sample period but fails out of sample. The researcher observes that the model fails for a particular reason. The researcher modifies the initial model so it then works both in sample and out of sample. **This is no longer an out-of-sample test. It is overfitting.***

在《科学回测中的大学问》一文中，我们表达过完全一致的观点：

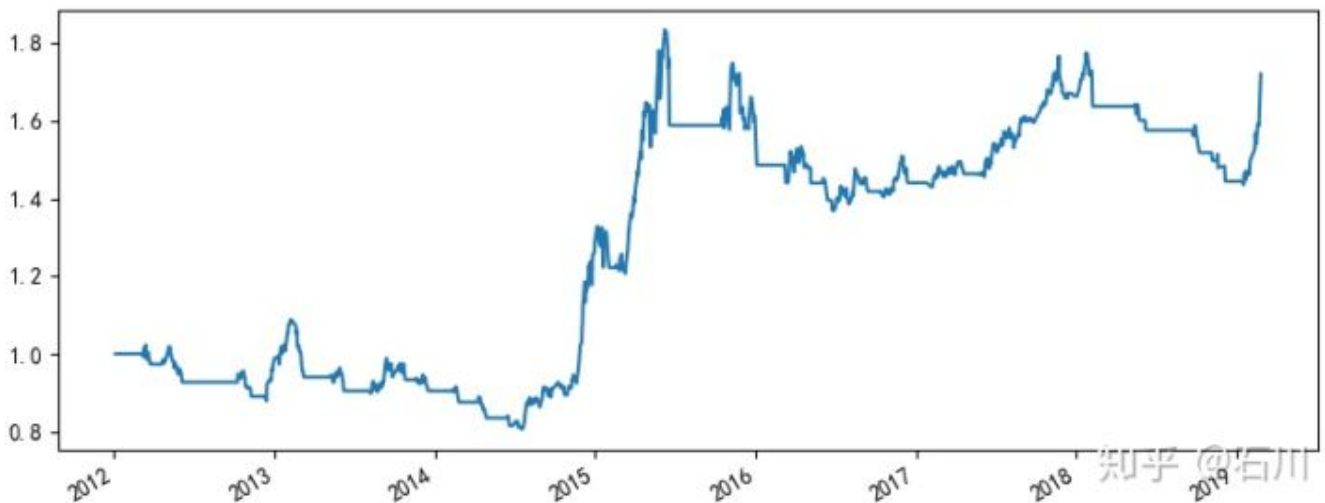
在可交易特征方面，满足一个策略假设的样本其实很少。举个极端的例子，比如 A 股中追踪大牛市的趋势类策略。在过去 20 年中，也仅有三波牛市，而且它们表现出来的市场特征均不同（比如以 2007 年大牛市构建的趋势追踪策略在 2015 年大牛市中的逃顶效果并不好）。在这种情况下，如果还把数据分成训练集和测试集会怎样呢？**我们一定会把策略在测试集中的体现**

来看一个简单的例子 —— 针对沪深 300 指数使用双均线择时。

假设使用 2005 年 1 月 1 日到 2011 年 12 月 31 日作为样本内数据；使用 2012 年 1 月 1 日到 2019 年 2 月 25 日为“样本外”数据。在样本内，使用 2 天窗口计算快均线、40 天窗口计算慢均线效果非常理想（下图），其年化收益率超过 30%，夏普率高达 1.3。



再来看看这组参数在“样本外”的表现。不幸的是，该策略在样本外的年化收益不到 8%、最大回撤 -25.95%、夏普率仅有 0.51。



如果对着“样本外”搜参数，则最优的快、慢均线窗口参数是 20 和 24，其夏普率也仅有 0.70，远远差于 2012 年之前的表现。考虑到这种情况，很自然的想法就是添加更复杂的逻辑来提高 2012 年之后的表现。显然，这不是样本外测试，而是过拟合。

无论从独立性还是可交易特征而言，交易数据其实都是十分匮乏的。从构建策略的角度来说，要求策略能够适应不同的市场状态，那么基于历史数据不断迭代的样本外并不是真正的样本外。

知乎

首发于
川流不息

所有的历史数据都是样本内。

我们构建策略靠的不是站在回测起点往后看的先见之明，而是站在回测终点往前看的后见之明。诚然，这并不意味着策略就不靠谱或者一定是数据挖掘的产物。但是，意识到这一点会让我们更加相信来自金融学或经济学先验的指标或因果关系，并防止我们沉迷于纯粹基于数据的交叉验证。

AQR 曾写过一篇题为 Buffett' s Alpha 的文章分析巴菲特的投资组合。结果显示，巴菲特投资组合的收益几乎可以被市场因子和五个如今我们这些后见之明的凡夫俗子能够理解的风格因子的 β 来解释。而反观巴菲特，他是在几十年前就按照这些风格因子来投资了，那是真正的样本外。

Buffett figured out things we think are genius many years before we all did. — Cliff Asness

参考文献

- Arnott, R., C. R. Harvey, and H. Markowitz (2019). A backtesting protocol in the era of machine learning. *Journal of Financial Data Science*, Vol. 1(1), 64 – 74.

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

量化交易 算法交易 回测框架

▲ 赞同 127 ▼ 8 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](#)

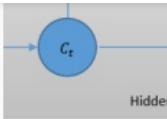
推荐阅读

读金融专业的，更加没有做过交易，彻彻底底的是一个新手，但是我觉得交易很有意思。所以想学，想尝试做一名交易员，以后往量化交易的方向走，，但是我想没有...

天才交易员... 发表于从零开始学...

到php、java一直到现在的python，后来做技术管理，但是对金融很感兴趣，感觉能用上数学知识，为此考了一个金融学硕士，所以总想试一把自己的水平。而炒...

green banbu



从语音识别
到科夫模型

江乙文

8 条评论

⇌ 切换为时间排序

写下你的评论...



石之龙

5 个月前

很多依据因果逻辑进行交易的 α ，会慢慢被量化分拆破解为 β ，随着时间的推移，要获得超过量化的 α 会变得越来越难。

👍 赞



Kingls

5 个月前

马科维茨90多了还在发论文啊.....

👍 2



石川 (作者) 回复 Kingls

5 个月前

我第一反应也很吃惊！查了一下确实是那个马科维茨：)

👍 3



lufengbo

5 个月前

个人理解：机器学习把数据集划分为训练集，验证集和测试集，前提条件是数据来之相同的分布，而股票价格时间序列显然不满足这个假设，因此通过验证集调模型超参数难以解决过拟合问题。核心还是数据的问题

👍 1



Guo Yu 回复 lufengbo

5 个月前

还是样本量不足，A股市场的历史太短，如果有100年的大样本，可能建模的效果会好很多

👍 赞



苏小叫化

5 个月前



郭嘉公元

5 个月前

那什么是alpha?

 赞



OldMonster

5 个月前

量化从业人员都相信历史规律会重现，只要这个规律统计上是站的住脚的，alpha策略之所以存在这个问题，恰恰是因为其结果在统计上没有显著性，其实入行周得实习生都能意识到这个问题了，这样的文章能发出来，多半是作者名气很大。

 赞

