

知乎

首发于
川流不息

出色不如走运 (II)?



石川

量化交易 话题的优秀回答者

已关注

babyquant 等 87 人赞同了该文章

摘要

本文指出在做因子测试时应考虑多重检验的影响、排除 data mining 造成的运气成分，从而有效的从大量因子中选出真正能够解释截面收益率的好因子；该方法也可用于基金经理或投资策略的筛选。

1 引言

两年前，我写了一篇《出色不如走运？》。该文使用顺序统计量（order statistics）解释了当很多投资者（或基金）使用相同的数据构建不同的策略时，最好的那个一定是非常优秀的，但它很有可能仅仅是因为运气好，而非真正的水平高。

如果我们**直接**从某个经济学规律中找出了一个解释股票预期收益截面差异的因子，并且该因子在统计上显著，那么它可能是真的显著；但如果我们试了 500 个因子，然后找到了一个最牛逼的，那么哪怕它的 t-statistic 非常高，我们也不能保证它就一定是个真的因子。



知乎

首发于
川流不息

对的还会有 250 人左右，但是我们会认为这些人仅仅是运气好。

这些例子背后的数学逻辑是，如果有一个因变量 Y 和一个解释变量 X ，通过回归分析后我们发现回归系数的 t -statistic 很高（比如 2.0，对应 5% 的显著性水平），那么从传统的单因素假设检验角度可以认为 X 能够显著的解释 Y 。然而，如果我们有很多个变量（比如 100 个） X_1 、 X_2 、...、 X_{100} ，我们全都试了之后发现第 55 个变量最好。这时，如果它的 t -statistic 也是 2.0，我们却不能说 X_{55} 显著的解释 Y 。这是因为仅仅靠运气，这 100 个变量（假设独立）中最好的那个的 t -statistic 大于 2.0 的概率高达 99%。

如何在层出不穷的因子中排除靠 data mining 挖掘的、而找到真正能够解释股票预期收益截面差异的？如何在大量的基金经理（或策略）中排除走运的、而找到真正能够战胜市场的？这些已成为非常迫切的问题。

在《出色不如走运》中，我们只说了仅仅凭运气就能得到非常好的结果，却没有说应该怎样排除运气，找到真正的好因子或者好策略。带着这些问题，今天就来一篇升级版——出色不如走运 (II)?

最后一点提示，本文非常 technical，建议静下心来阅读。此外，熟悉《股票多因子模型的回归检验》、《为什么要进行因子正交化处理？》、以及《用 Bootstrap 进行参数估计大有可为》对阅读本文会有帮助。

2 理论依据

既然是升级版，就不能光靠 order statistics 说事儿了，咱也得武装升级一下理论。

当学术界有大量因子来解释同一个问题——股票截面预期收益（或者有许多不同的策略在同一个市场中交易时），仅考虑单一检验（single testing，即每次检验一个 hypothesis，比如一个单因子是否有效？）就不再适合了；这时候必须要考虑 multiple testing（多重检验）造成的影响。在统计上，multiple testing 指的是同时检验多个 hypotheses。

在金融领域应用 multiple testing 在最近几年得到了飞速发展。这其中的代表人物要数杜克大学的 Campbell Harvey 教授（曾于 2016 年任美国金融协会主席），他自 2014 年以来发表了多篇文章、进行了多个演讲。其中最具代表性的文章包括：

- Harvey et al. (2016) 研究了学术界发表的 316 个显著的选股因子，在已有的多重检验修正——包括 Bonferroni adjustment、Holm adjustment 以及 Benjamini-Hochberg-Yekutieli (BHY) adjustment——的基础上，提出了一种能够利用不同因子之间相关性的全新检验框架、以排除 multiple testing 的影响，并指出只有在 single testing 中 t -statistic 超过 3 非人们传统认为的 5% 的显著性水平对应的 2) 的因子才有可能在考虑了多重检验之后依然有效。Harvey 同时也指出，3.0 其实都是非常保守的。



知乎

首发于
川流不息

50%。Harvey and Liu (2015a) 定量计算了不同大小的 Sharpe Ratio 在实盘外的“打折程度”（他们称为 haircut ratio），发现了 haircut ratio 和 Sharpe Ratio 之间的非线性关系。

除上述研究外，Harvey and Liu (2015b) 提出了一个全新的基于 regression 的检验框架排除 multiple testing 影响、解决因子挑选问题。它的优势是可以按顺序逐一挑出最显著的因子、第二显著的因子，以此类推，直到再没有显著因子。这么做的好处是可以评价每个新增加的因子在解释股票截面收益率时的增量贡献。这是传统的多重检验无法做到的。此外，该方法也可以被用来找到真正能够战胜市场的基金经理或投资策略。

本文的主要目标是介绍 Harvey and Liu (2015b) 提出的基于 regression 的多重检验方法。考虑到早期的多重检验方法（即 Bonferroni、Holm、BHY adjustments）也非常容易上手便捎带着加以说明。至于 Harvey et al. (2016) 提出的方法，其技术性较强，复制起来比较困难，因此我们今后找机会再聊它（倒是可以先记住它的结论，即 t-statistic 要至少大于 3 才有可能在排除了 multiple testing 影响后依然显著）。

下面首先来看容易上手的 Bonferroni、Holm 以及 BHY adjustments。

3 Bonferroni、Holm、BHY Adjustments

这三种多重检验修正可以分为两类：

- Bonferroni 和 Holm adjustments 的目的是控制 family-wise error rate（族错误率）；
- BHY adjustment 的目的是控制 false discovery rate。

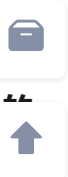
在多重检验中，family-wise error rate (FWER) 和 false discovery rate (FDR) 代表着 Type I error 的两个不同的定义。Type I error 是错误的拒绝原假设，也叫 false positive 或 false discovery。在我们的上下文中，它意味着错误的发现了一个其实没用的因子。

假设 K 个 hypotheses 的 p-value 分别为 p_1 、 p_2 、...、 p_K 。根据事先选定的显著性水平，比如 0.05，其中 R 个 hypotheses 被拒绝了。换句话说，我们有 R 个发现 (discoveries)——包括 true discoveries 和 false discoveries。令 $N_r \leq R$ 代表 false discoveries 的个数。由此，FWER 和 FDR 的定义如下：

$$\text{FWER} = \text{prob}(N_r \geq 1)$$

$$\text{FDR} = E[N_r / R]$$

从定义不难看出，FWER 是至少出现一个 false discovery 的概率，控制它对单个 hypothesis 来说是相当严格的，会大大提升 Type II Error。相比之下，FDR 控制的是 false discoveries





需要说明的是 Bonferroni、Holm 以及 BHY 这三种方法都是为了修正 single testing 得到的 p-value，修正后的 p-value 往往会大于原始的 p-value，也就意味着修正后的 t-statistic 更小，即 hypotheses 不再那么显著。

下面通过简单的例子（出自 Harvey and Liu 2015a）解释这三种方法。假设一共有六个因子，它们 single testing 的 p-value 从小到大依次是 0.005、0.009、0.0128、0.0135、0.045、0.06。按照 0.05 的显著性水平来看，前五个因子是显著的。

首先来看 Bonferroni correction（中文称作邦费罗尼校正），它对每个原始 p-value 的调整如下：

$$p_i^{\text{Bonferroni}} = \min\{Kp_i, 1\}, \quad i = 1, \dots, K$$

根据定义，这六个因子的 Bonferroni p-value 分别为 0.03、0.054、0.0768、0.081、0.27 和 0.36。经过修正后，在 0.05 的显著性水平下，仅第一个因子依然显著。

接下来看看 Holm 修正（Holm 1979）。它按照原始 p-value 从小到大依次修正，公式为：

$$p_i^{\text{Holm}} = \min \left\{ \max_{j \leq i} \{(K - j + 1)p_j\}, 1 \right\}, \quad i = 1, \dots, K$$

根据上述定义，原始 p-value 最小的因子被修正后，其 Holm p-value 为 0.03；第二个因子的 Holm p-value 为 $\max\{6 \times 0.005, 5 \times 0.009\} = 0.045$ 。以此类推就能计算出其他四个因子的 Holm p-value：

$$\begin{aligned} p_3^{\text{Holm}} &= \max\{6p_1, 5p_2, 4p_3\} = 4p_3 = 0.0512 \\ p_4^{\text{Holm}} &= \max\{6p_1, 5p_2, 4p_3, 3p_4\} = 4p_3 = 0.0512 \\ p_5^{\text{Holm}} &= \max\{6p_1, 5p_2, 4p_3, 3p_4, 2p_5\} = 2p_5 = 0.09 \\ p_6^{\text{Holm}} &= \max\{6p_1, 5p_2, 4p_3, 3p_4, 2p_5, p_6\} = 2p_5 = 0.09 \end{aligned}$$

经过 Holm 修正后，在 0.05 的显著性水平下，只有前两个因子依然显著。

最后来看看 BHY 修正（Benjamini and Hochberg 1995, Benjamini and Yekutieli 2001）。它从原始 p-value 中最大的一个开始从大到小逆向修正，公示如下：



知乎

首发于
川流不息

$$p_i^{\text{BHY}} = \begin{cases} \min \left\{ p_{i+1}^{\text{BHY}}, \frac{K \times c(K)}{i} p_i \right\} & \text{if } i \leq K - 1 \end{cases}$$

其中,

$$c(K) = \sum_{j=1}^K \frac{1}{j}$$

在本例中, 因为 $K = 6$, 因此 $c(K) = 2.45$ 。由 BHY 的定义可知原始 p-value 最大的因子调整后的 BHY p-value 就是它自己。然后从第二大的开始, 依次按照上述公式计算, 最终得到了全部因子调整后的 BHY p-value, 它们是 (从小到大排列): 0.0496、0.0496、0.0496、0.0496、0.06、0.06。在 0.05 的显著性水平下, 前四个因子依然显著。

BHY 方法是以控制 false discovery rate 为目标, 它的修正比另外两种以控制 family-wise error rate 的方法更加温和。这体现出来的结果就是在 BHY 调整下, 有更多的因子依然显著。此外, BHY 方法对检验统计量之间的相关性不敏感, 它的适应性很强。

各位小伙伴不妨使用上面介绍的这三种方法对因子的 p-value 进行修正试试。

4 基于 Regression 的检验

本节介绍 Harvey and Liu (2015b) 提出的基于 regression 的检验方法, 该方法受到了 Foster et al. (1997) 以及 Fama and French (2010) 的启发, 在这二者的基础上又有不少的创新。它的目的是为了从一大堆号称显著的因子中排除 data mining、找到真正显著的; 该方法也可以被用于从一大堆基金经理或策略中找出真正能够战胜市场的。

当很多因子被用来解释截面收益时, 效果最显著 (最显著可以由最高的 t-statistic、R-squared 等指标代表) 的因子中一定包含了运气的成分。**这个方法的巧妙之处在于通过正交化和 Bootstrap 得到了仅靠运气能够得到的显著性的经验分布; 如果在排除了运气带来的显著性之后某个因子依然显著, 那它就是真正的因子, 而非 data mining 的结果。**

随着处理方式略有不同, Harvey and Liu (2015b) 这个方法可以用于 predictive regression (考察哪个 X 能预测 Y)、panel regression 以及 Fama-MacBeth regression (这两类回归可以用于挑选好因子), 但它们背后的逻辑完全一致。下面高度概括一下该方法的逻辑 (**正交化和 Bootstrap 是核心**):



知乎

首发于
川流不息

Y 的部分，从而构建 null hypothesis: X 无法解释 Y;



使用每个正交化后的因子 X 解释 Y、找出它们中最显著的，这个显著性全部来自运气的成分（因为正交化保证了 null hypothesis 是正交化后 X 无法解释 Y）；



通过 Bootstrap 采样进行大量实验就可以得到仅凭运气得到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最显著的那个因子的显著性在这个分布的 95%（或 90%、99%）分位数之上，那就说明即便排除了足够好的运气成分，该因子依然有效。

知乎 @石川

接下来以 predictive regression 为例说明该检验方法的具体步骤。Harvey and Liu (2015b) 中给出了使用 panel regression 和 Fama-MacBeth regression 时所需的改动。为了评价哪个因子有效，需要用到 panel regression，因此下一节会介绍针对 panel regression 的改动。

假设有因变量 Y 和 100 个解释变量 X 的 500 期样本数据，我们想看看哪个 X 能够预测 Y。多重检验的步骤为：

第一步：用每个 X 和 Y 回归（在我们的例子中就是 100 次回归），得到 100 个残差 OX，它们和 Y 正交。这构成了 null hypothesis: 所有 OX 对 Y 没有预测性。

第二步：以这 500 期的 Y 和正交化得到的 OX 为原始数据（ 500×101 的矩阵，每一行代表一期，第一列为 Y，第二到第 101 列为 100 个 OX 变量），使用**带放回**的 Bootstrap 重采样从这 500 行中不断的随机抽取，构建和原始长度一样的 bootstrapped 数据（也是 500×101 矩阵）。**整行抽取保留了这 100 个变量在截面上的相关性。此外 Bootstrap 的好处是不对原始数据中的概率分布做任何假设。**

第三步：使用 bootstrapped 数据，用每个 OX 和 Y 回归得到一个检验统计量（比如是 t-statistic）；找出所有 OX 中该检验统计量最大的那个值，称为 max statistic。如果我们的检验统计量是 t-statistic，那么这个 max statistic 就是 500 个 t-statistic 中最大的。



够得到的 max statistic 的分布。

第五步：比较原始数据 Y 和每个 X 回归得到的 max statistic 和第四步得到的 max statistic 的经验分布：

- **a.** 如果来自真实数据的 max statistic 超过了经验分布中的阈值（比如 95% 显著性水平对应的经验分布中 max statistic 的取值），那么真实数据中 max statistic 对应的解释变量就是真正显著的。假设这个解释变量是 X_7 。
- **b.** 如果来自真实数据的 max statistic 没有超过经验分布中的阈值，则这 100 个解释变量全都是不显著的。本过程结束，无需继续进行。

第六步：使用目前为止已被挑出来的全部显著解释变量对 Y 进行正交化，得到残差 OY 。它是原始 Y 中这些变量无法解释的部分。

第七步：使用 OY 来正交化剩余的 X （已经选出来显著变量，比如 X_7 ，不再参与余下的挑选过程）。

第八步：重复上述第三步到第七步：**反复使用已挑出的显著因子来正交化 Y ，再用 OY 来正交化剩余解释变量 X ；在 Bootstrap 重采样时，使用 OY 、 k 个已经选出的 X 、和剩余 $100 - k$ 个正交化后的 OX 作为原始数据生成 bootstrapped 样本；通过大量的 Bootstrap 实验得到新的 max statistic 的经验分布，并判断剩余解释变量中是否仍然有显著的。**

第九步：当剩余解释变量的 max statistic 无法超过 null hypothesis 下 max statistic 的经验分布阈值时，整个过程结束，剩余的解释变量全都是不显著的。

以上以 predictive regression 为例介绍了 Harvey and Liu (2015b) 提出的多重检验框架。

5 用 Panel Regression 挑选好因子

在分析因子是否能显著的~~解释股票或投资组合的截面预期收益率时~~，回归方法是 panel / cross-sectional regression 而非前一节的 predictive regression。**需要说明的是，这里的选股因子都是某个投资组合的（超额）收益率，比如 MKT, HML, SMB 这种。**

在使用 panel regression 的检验中，Bootstrap 的思想和上一节介绍的完全一致，但是在正交化、回归分析、以及 max statistic 的选取有上些差异。

5.1 正交化

在挑选因子中，null hypothesis 是因子对解释预期收益率截面差异没有作用。如果能够拒绝原假设，则说明因子是有效的。但是运气的成分往往带来 false discovery，即本来这个因子没用，



hypothesis, 即因子不能解释截面收益率。

正交化的方法为：

- 在尚未选出任何显著因子时，对所有潜在因子的正交化处理方法是 **demean (去均值)**。由于每个因子都是一个收益率，因此使用原始的因子值减去它在时序上的均值就排除了它在截面上的解释性（因为 demean 后该因子在截面上的期望收益是零）。
- 如果已经选出了 k 个显著的因子，在继续挑选第 $k + 1$ 个显著因子时，正交化的方法是使用这 k 个因子作为解释变量和第 $k + 1$ 个因子在时序上回归，得到的残差就是正交化之后的待检验因子。

5.2 回归分析

在 predictive regression 中，我们会对因变量和解释变量都进行正交化。假设已经选出了 $k \geq 0$ 个显著变量。在选择第 $k + 1$ 个时，首先将 Y 投影到这 k 个变量上得到残差 OY ，这就是对 Y 的正交化。之后，再把剩余待检验的解释变量 X 逐一投影到 OY 上，得到 OX 。然后再用 OY 和每个 OX 独立回归进行后续 Bootstrap 步骤。这使得我们可以评估新加入变量 X 在预测 Y 时的增量贡献。

进行 panel regression 时，个股或者投资组合的收益率作为因变量出现在回归方程的左侧，对它们不进行正交化处理。在回归方程的右侧，使用已经选出的 k ($k \geq 0$) 个显著因子和正交化后的第 $k + 1$ 个因子（正交化方法参考 5.1 节）作为解释变量。始终将已经选出的前 k 个因子加入回归方程的右侧可保证检验第 $k + 1$ 个因子对解释截面收益率的增量贡献。将因变量和解释变量在时序上回归，得到的截距项就是这些因子无法解释的 pricing error。

上面的对比说明：在 predictive regression 中，回归方程的左侧是 OY （用已经选出的 k 个 X 正交化 Y ），而右侧只有一个 OX （每个剩余的 X 依次和 OY 回归）；而在 panel regression 中，回归方程的左侧是 Y （不正交化），而是把已经选出的 k 个 X 都放在回归方程的右侧，因此右侧为 k 个 X 以及一个新的待检验的正交化后的 OX 。不同的方法是由于这两种回归中 null hypothesis 的性质不同造成的。虽然这两种方法的略有不同，但都保证了考察待检验变量对解释 Y 的增量贡献。

在 Harvey and Liu (2015b) 的最新版本 Harvey and Liu (2018) 中对上述回归有非常详细的说明。值得一提的是，虽然作者将这个回归称为 panel regression，但 Harvey and Liu (2018) 对每个投资品单独的使用这些因子进行时序回归。因此对于 N 个投资品，一共得到了 N 个 pricing errors；如果直接使用 N 个投资品一起做 panel regression 并加入 fixed effects 也可以得到 N 个不同的截距。

5.3 “Max statistic”





好的那个仅仅靠着运气的成分也可以让 pricing error 非常接近零。为了量化并排除运气的影响，Bootstrap 的目标就是得到 null hypothesis 下 pricing error 的经验分布，即仅靠运气能够得到的 pricing error 的经验分布。

从 asset pricing 角度来说，如果一个因子能够解释收益率截面差异，那么回归截距应十分接近零。由于一共有 N 个投资品，使用这 N 个投资品的 pricing error 绝对值的中位数作为 “max statistic”（实际上是希望 pricing error 的绝对值越小越好，因此应称之为 min statistic；为了和前一节对应，故称之为带了引号的 “max statistic”）来评价因子。通过 Bootstrap 得到 “max statistic” 的经验分布。如果来自真实数据的最小 pricing error 绝对值的中位数小于从经验分布中得到的阈值，则它对应的因子就是真正有效的因子。

还有点乱？实在抱歉，我真的尽力了。

没关系，下面介绍 Harvey and Liu (2015b) 中的一个例子来说明挑选因子的过程。

6 一个例子

Harvey and Liu (2015b) 给出了一个示例性例子说明如何应用他们提出的多重检验框架挑选真正有效的因子。这个例子考察了学术界的 13 个 “显著” 因子。加个双引号是因为它们都在 single testing 中显著，但是新的多重检验下很多就失效了。

这 13 个因子为：

- Fama and French (1993): MKT、SMB、HML;
- Fama and French (2015): RMW、CMA;
- Hou et al. (2015): ROE、IA;
- Frazzini and Pedersen (2014): BAB;
- Novy-Marx (2013): GP;
- Pastor and Stambaugh (2003): PSL;
- Carhart (1997): MOM;
- Asness et al. (2013): QMJ;
- Harvey and Siddique (2000): SKEW。

这些因子的 single testing 结果（以因子收益率的 t-statistic 表示）以及它们之间的相关性如下图所示。从图中不难看出：（1）除了 SMB 外，所有因子的 t-statistic 都大于 2，在 0.05 的显著性水平下显著；有些因子的 t-statistic 甚至超过 5！（2）这些因子中有一些对的相关性非常高，比如 ROE 和 QMJ、CMA 和 IA（它们都是 investment 类的因子）、CMA 和 HML 等。



知乎

首发于
川流不息

Mean	0.057	0.027	0.049	0.086	0.032	0.056	0.070	0.054	0.046	0.103	0.037	0.045	0.035
t-stat	[2.28]	[1.63]	[2.96]	[3.51]	[2.34]	[2.88]	[4.89]	[5.27]	[3.36]	[5.49]	[2.88]	[4.22]	[2.88]

Panel B.2: Factor Correlation Matrix

	<i>mkt</i>	<i>smb</i>	<i>hml</i>	<i>mom</i>	<i>skew</i>	<i>psl</i>	<i>roe</i>	<i>ia</i>	<i>qmj</i>	<i>bab</i>	<i>gp</i>	<i>cma</i>	<i>rmw</i>
<i>mkt</i>	1.00												
<i>smb</i>	0.25	1.00											
<i>hml</i>	-0.32	-0.11	1.00										
<i>mom</i>	-0.14	-0.03	-0.15	1.00									
<i>skew</i>	-0.05	-0.00	0.24	0.04	1.00								
<i>psl</i>	-0.03	-0.03	0.04	-0.04	0.10	1.00							
<i>roe</i>	-0.18	-0.38	-0.09	0.50	0.20	-0.08	1.00						
<i>ia</i>	-0.37	-0.15	0.69	0.04	0.20	0.03	0.06	1.00					
<i>qmj</i>	-0.52	-0.51	0.02	0.25	0.16	0.02	0.69	0.13	1.00				
<i>bab</i>	-0.10	-0.01	0.41	0.18	0.25	0.06	0.27	0.34	0.19	1.00			
<i>gp</i>	0.07	0.02	-0.31	-0.01	0.01	-0.06	0.32	-0.22	0.48	-0.09	1.00		
<i>cma</i>	-0.40	-0.05	0.70	0.02	0.09	0.04	-0.08	0.90	0.05	0.31	-0.29	1.00	
<i>rmw</i>	-0.23	-0.39	0.15	0.09	0.29	0.02	0.67	0.09	0.78	0.29	0.47	-0.03	1.00

为了测试因子，最好的因变量应该是一揽子股票，因为我们希望考察这些因子在解释股票预期收益率截面差异上的作用。在 Harvey and Liu (2015b) 给出的例子中，二位作者使用的是 25 个投资组合，而非个股。他们强调例子的目的是为了说明多重检验的步骤。用来作为因变量的 25 个投资组合来自使用 Fama-French 三因子中的 SMB 和 HML 两个因子各自把股池分成 5 组并交叉配对，因此一共 $5 \times 5 = 25$ 个组合。

Harvey and Liu (2015b) 使用了这 25 个组合的 pricing error 绝对值的中位数作为挑选因子的指标（在文章中，这个指标被记为 m_1^a ）。除了这个指标外还有其他三个指标，这里不做讨论。

首先用这 13 个因子各自对这 25 个投资组合进行回归。每个因子 pricing error 绝对值的中位数如下图所示。从单个因子回归结果来看，MKT（市场）因子是最显著的（它的指标 0.285% 是所有因子中最小的），但是里面包含了运气的成分。



知乎

首发于
川流不息

Panel A: Baseline = No Factor

<i>mkt</i>	0.285	0.277	1.540	1.750
<i>smb</i>	0.539	0.513	2.851	5.032
<i>hml</i>	0.835	0.817	4.541	12.933
<i>mom</i>	0.873	0.832	4.626	13.965
<i>skew</i>	0.716	0.688	3.822	9.087
<i>psl</i>	0.726	0.699	3.887	9.548
<i>roe</i>	0.990	1.011	5.623	21.191
<i>ia</i>	1.113	1.034	5.750	21.364
<i>qmj</i>	1.174	1.172	6.516	28.427
<i>bab</i>	0.715	0.725	4.029	9.801
<i>gp</i>	0.692	0.663	3.688	8.816
<i>cma</i>	0.996	0.956	5.318	17.915
<i>rmw</i>	0.896	0.881	4.900	15.647

知乎 @ 百川

下面应用多重检验来排除运气的成分。对这 13 个因子分别正交化 (demean)，然后使用 Bootstrap 重采样进行反复多次的大量实验。每个实验中，单独使用 13 个正交化后的因子和 25 个投资组合收益率回归，得到每个因子的 pricing error 绝对值中位数的最小值（我们的 “max statistic”）。大量 Bootstrap 实验便得到了 “max statistic” 的经验分布。MKT 因子的取值 (0.285%) 在这个分布下出现的概率仅为 3.9%，即 p-value = 3.9%，小于常用的 5% 的阈值。因此我们说即便考虑了运气成分后，MKT 因子依然是显著的。市场因子是第一个被选出来的显著因子，这多少符合预期。

在接下来的步骤中，使用 MKT 因子正交化其余 12 个因子。然后用 MKT 因子和正交化之后的每个剩余因子独立对这 25 个投资组合进行回归分析，得到考虑了每个剩余因子的 pricing error 绝对值的中位数，如下图所示。不难看出，在剩余的 12 个因子中，CMA 是最好的（它的 pricing error 最低），但是 HML 和 BAB 和它也难分伯仲！因此，在真实数据中，“max statistic” 的取值为 0.112%（来自 CMA）。



知乎

首发于
川流不息

Panel B: Baseline = mkt

<i>smb</i>	0.225	0.243	1.348	1.633
<i>hml</i>	0.120	0.150	0.836	0.341
<i>mom</i>	0.301	0.328	1.825	2.469
<i>skew</i>	0.239	0.236	1.314	1.292
<i>psl</i>	0.258	0.265	1.474	1.611
<i>roe</i>	0.332	0.363	2.020	3.846
<i>ia</i>	0.166	0.163	0.907	0.358
<i>qmj</i>	0.344	0.398	2.213	4.615
<i>bab</i>	0.121	0.152	0.844	0.382
<i>gp</i>	0.305	0.314	1.745	2.148
<i>cma</i>	0.112	0.130	0.721	0.153
<i>rmw</i>	0.225	0.285	1.586	2.204

知乎@石川

再一次，使用 Bootstrap 重采样进行反复多次的大量实验得到 “max statistic” 的经验分布。CMA 因子的取值 (0.112%) 在这个分布下出现的概率仅为 2.2%，依然小于常用的 5% 的阈值。**在考虑了运气以及 MKT 因子之后，CMA 因子依然是显著的。**如果不选 CMA 作为第二个，也可以选 HML (**价值投资!**) 或 BAB 作为第二个显著的因子。

如上所述，重复这个过程就可以一直分析下去。在找到了最有效的两个因子 —— MKT 和 CMA —— 之后，剩余 11 个因子中第三个最显著的因子是 SMB，它的 pricing error 是 0.074%。然而，使用 Bootstrap 得到 “max statistic” 的经验分布后发现，SMB 因子的取值 (0.074%) 在这个分布下出现的概率高达 13.9%，大于常用的 5% 的阈值，因此认为 **SMB 以及其他 10 个因子在进一步解释截面收益率差异时都不显著。**



知乎

首发于
川流不息

Panel C: Baseline = mkt + cma

<i>smb</i>	0.074	0.102	0.569	0.319
<i>hml</i>	0.109	0.130	0.720	0.188
<i>mom</i>	0.112	0.141	0.785	0.253
<i>skew</i>	0.124	0.126	0.702	0.123
<i>psl</i>	0.113	0.128	0.711	0.128
<i>roe</i>	0.194	0.249	1.382	1.584
<i>ia</i>	0.167	0.168	0.936	0.432
<i>qmj</i>	0.327	0.307	1.707	2.494
<i>bab</i>	0.147	0.129	0.715	0.114
<i>gp</i>	0.084	0.105	0.581	-0.031
<i>rmw</i>	0.181	0.197	1.093	0.813

知乎 @ 宿州

经过多重检验发现，MKT 和 CMA（也可以选 HML 或 BAB）是两个显著的因子，其他因子均不显著，均为 data mining 的产物。

以上便实现了从一揽子所谓显著的因子中提出运气成分、找到真正有效的因子。这就是这套多重检验体系最大的价值。这套体系也可以用于基金经理的筛选，具体的例子见 Harvey and Liu (2015b)。

7 结语

2015 年，Harvey 教授在 Jacobs Levy Center' s Conference 上进行了题为 Lucky Factors 的演讲。在演讲的开篇，他从生物进化的角度指出人类可能有 overfitting 或者 data mining 的倾向。

假设一只机警的羚羊在草原中听到了沙沙响声。如果它开始奔跑，但事后发现响声只是由于一阵微风造成的（即没有威胁），那么它无疑犯了 Type I error，为此付出的代价是消耗一定的能量；但是如果它不奔跑，但事后发现响声是因为一只猎豹冲向它造成的，那么它则犯了 Type II error，为此则付出了生命的代价。可见，从 cost 的角度，它必须选择奔跑。



知乎

首发于
川流不息

Type I error

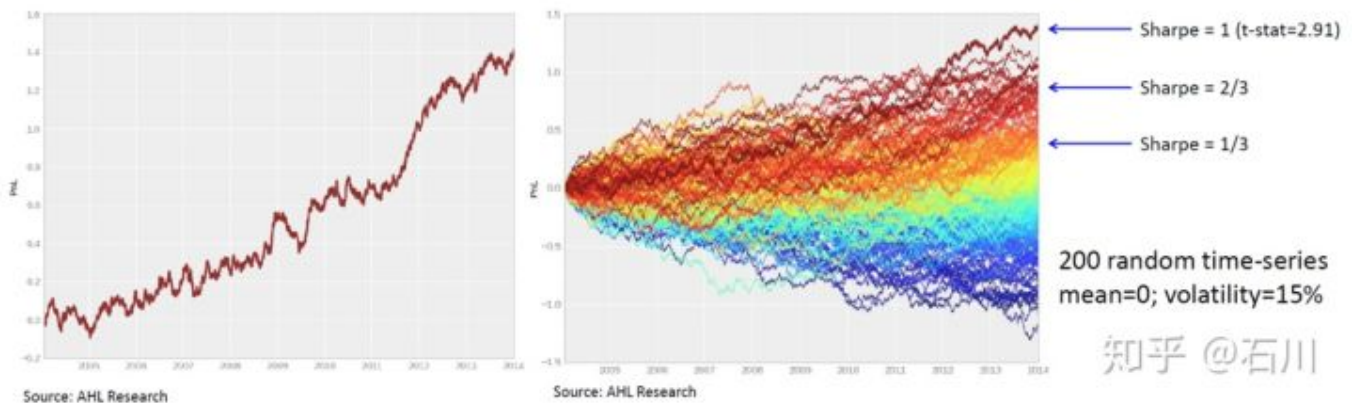


Type II error

知乎 @石川

这个故事告诉我们，动物想生存，就必须控制 Type II error，而可以允许更高的 Type I error (false discovery)。这种倾向在进化中被一代代传下来。因此，人类在分析问题时允许更高的 Type I error、存在 overfitting 或者 data mining 的倾向。

下图左侧是一个假想的策略净值曲线，它持续上涨，回撤可控，Sharpe Ratio 理想。然而，它仅仅是下图右侧中展示的 200 个使用零均值纯随机生成的策略净值中表现最好的那个。换句话说，它的表现完全来自运气。



出色还是走运？回答这个问题刻不容缓。

S&P Capital IQ 有一个 Alpha Factor Library (α 因子库)，宣称有 500 个 α 因子！这里面有多少是运气？有多少是真正的 α ？本文介绍的几种方法是为了回答这个问题所做的努力。



知乎

首发于
川流不息

Stock Selection Signals on Over 40,000 Companies in 90 Countries Around the World

Enhancements Include 115 Signals for 14 Key Industries

NEWS PROVIDED BY
S&P Capital IQ and SNL Financial
Dec 08, 2015, 02:50 EST

SHARE THIS ARTICLE



NEW YORK, Dec. 8, 2015 /PRNewswire/ -- S&P Capital IQ and SNL, a leading provider of market research, data and analytics, announced today that its Alpha Factor Library, an extensive and proprietary collection of global equity selection signals, or alpha factors, used by fundamental and quantitative investors, Fund of Funds, ETF providers and other investment professionals, now includes more than 500 alpha factors for over 40,000 companies in 90 countries.

Identify new methods to generate alpha potential.

Identify new sources of alpha with our Alpha Factor Library—the first of its kind—using our global point-in-time databases.



美国统计协会 (American Statistical Association) 的 Ethical Guidelines for Statistical Practice 中, 有这样一句话, 发人深省:

Selecting the one "significant" result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion. Failure to disclose the full extent of tests and their results in such a case would be highly misleading.

译: 从多项检验中挑出 “最重要” 结果很有可能造成不正确的结论。在这种情况下, 如不披露检验的全部内容及其结果, 便会造成极大的误导。

哪有那么多阿尔法? !

感谢阅读, 祝各位国庆节快乐。

参考文献

- Asness, C. S., A. Frazzini, and L. H. Pedersen (2013). Quality minus junk. AQR Capital Management working paper.
- Benjamini, Y. and Y. Hochberg (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B*, Vol. 57, 289 – 300.
- 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.
- 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 (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, Vol. 65(5), 1915 – 1947.
- Fama, E. F. and K. R. French (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, Vol. 116(1), 1 – 22.



知乎

首发于
川流不息

607.

- Harvey, C.R. and A. Siddique (2000). Conditional skewness in asset pricing tests. *Journal of Finance*, Vol. 55(3), 1263 – 1295.
- Harvey, C. R. and Y. Liu (2015a). Backtesting. *Journal of Portfolio Management*, Vol. 42(1), 13 – 28.
- Harvey, C. R. and Y. Liu (2015b). Lucky Factors. Working paper, available at jacobslevycenter.wharton.upenn.edu...
- Harvey, C. R. and Y. Liu (2018). Lucky Factors. Working paper, available at SSRN: ssrn.com/abstract=25287....
- 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.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, Vol. 6, 65 – 70.
- Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, Vol. 28(3), 650 – 705.
- Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, Vol. 108 (1), 1 – 28.
- Pastor, L. and R.F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, Vol. 111(3), 642 – 685.
- en.wikipedia.org/wiki/B...

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

数据挖掘

多因子模型

金融学

▲ 赞同 87 ▼

● 11 条评论

➤ 分享

★ 收藏

...

文章被以下专栏收录



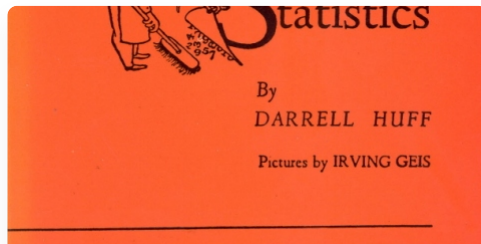
川流不息

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



知乎

首发于
川流不息

《统计数字会撒谎》How to lie with statistics【读书笔...

魔数

发表于伪极简主义

《Machine Learning: Regression》课程第6章KN...

1.KNN Regression的过程? 第一步: 找出和x最相近的K个点第二步: \tilde{y} 就是这K个点y的均值
2.KNN Regression的缺点是什么? 第一个缺点是对于样本少的区域容易overfitting。在边界处有很...

衫秋南

发表于地球派



机器学习 learning)

钱粮胡同

11 条评论

⇌ 切换为时间排序

写下你的评论...



哎哟 apj

10 个月前

假设是说一个因子利用和y呈现完全的线性回归的关系，所以用残差的方式可以得到“运气”残差，但是很多因子的有效性并不是一回归这么简单剔除的，这点您怎么看？

👍 赞



石川 (作者) 回复 哎哟 apj

10 个月前

我建议看看Harvey, C. R. and Y. Liu (2018). Lucky Factors. Working paper, available at SSRN: ssrn.com/abstract=25287... 这篇文章，里面对通过panel regression 分析因子有非常详尽的论述。谢谢

👍 赞



吃不瘦的橘猫

10 个月前

最后藏羚羊的例子很精髓！还有，第四部分，第8步，“使用 OY、k 个已经选出的 X、和剩余 500 - k 个正交化后的 OX 作为原始数据生成 bootstrapped 样本”，500是不是应该是100啊？

👍 赞



石川 (作者) 回复 吃不瘦的橘猫

10 个月前

对，笔误，100！谢谢。

👍 赞



Natedog

7 个月前

Holm修正中第一个因子修正后应该是0.03而非0.06：)





石川 (作者) 回复 Natedog

7 个月前

谢谢，改过来了。:)

 赞



教父

6 个月前

这方法和我live里讲到的一个方法有类似之处

 赞



石川 (作者) 回复 教父

6 个月前

厉害厉害

 赞



TerenceD

3 个月前

建议在方法解释部分增添必要的公式，思路应该会更清晰！

 赞  回复  踩  举报



言西早木杉tb

3 个月前

请问single test的p-value是原始的没有考虑异方差和自相关的t-statistic对应的p-value，还是可以是经过newey-west调整的t-statistic对应的p-value？？

 赞



石川 (作者) 回复 言西早木杉tb

3 个月前

学术论文中通常都是经nw调整后的

 赞

