

知乎

首发于
川流不息

Reversion to the mean is the iron rule of the financial markets.

— John C. Bogle —

次世代均值回归策略



石川

量化交易 话题的优秀回答者

已关注

淮浩等 74 人赞同了该文章

1 金融市场的均值回归

在量化投资领域，**均值回归 (mean reversion)** 代表着一大类策略。金融市场的均值回归定义如下：

*In finance, mean reversion is the assumption that a stock's **price** will tend to move to the average price over time.*

译：在金融领域，均值回归假设随着时间的移动，股票的**价格**朝着它的均值移动。

我们可以把上述定义中的“股票”换成其他任何投资品。这个定义中最核心的两个字是“**价格**”（而不是投资品的“收益率”）：

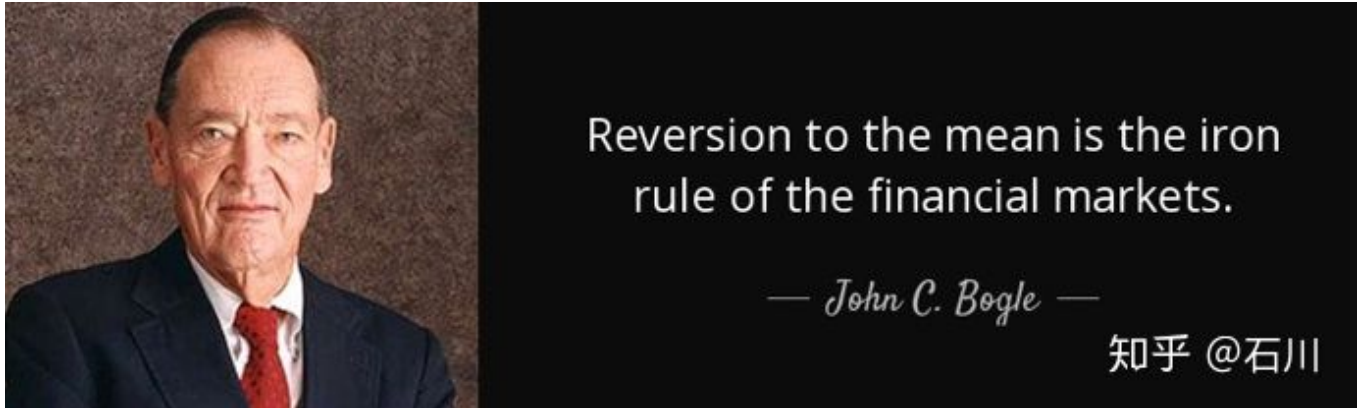
- “**价格呈现均值回归**”等价于“**收益率呈现序列负相关性**”。这是一种非常好的、可以被拿来构建优秀策略的特性。
- “**收益率呈现均值回归（即收益率围绕 0 随机的上下波动）**”等价于“**价格呈现随机游走**”。这是投资品最差的一种形态，在这种情况下不存在有效的赚钱策略（运气除外）。

请暂停一下，花几秒钟体会一下上面这两句话。它们是所有均值回归策略的核心，也是本文的核心。

知乎

首发于
川流不息

部的均值上下往复波动，呈现出围绕着均值的回归运动。但是这种程度的回归对于在统计上构建一个有效的量化策略并没有太多的帮助。



在现实中，为了构建一个均值回归策略，我们要求**价格的时间序列满足平稳性**。显然，**单一投资品的价格**是很难满足这个假设的。于是量化界的小伙伴们便开动脑筋，终于发现**虽然单一投资品的价格不满足均值回归，但是可以把多个投资品（通常是两个）线性组合在一起，使它们的价差满足均值回归**。

找到一对价差满足均值回归的投资品是早期均值回归策略的初衷，因此这类策略又有另外一个广为人知的别名——**配对交易（pairs trading）**。

配对交易就是当下市场上主流的构建均值回归策略的方法。它通常利用**协整（co-integration）**或者**价格距离法**来找到这样一对投资品，并认为它们的价差会在一定的区间内往复运动，然后基于价差的统计特性计算阈值进行交易，因此这种策略通常又叫**统计套利（statistical arbitrage）**。

在我看来，**多品种构建的价差很难满足一价定理、不易回归**；针对这类价差的统计套利策略的**风险收益比很差**。本文介绍一种新的均值回归思路。不过在此之前，还是让我们先来看看市场上流行的这两种投资品配对方法。

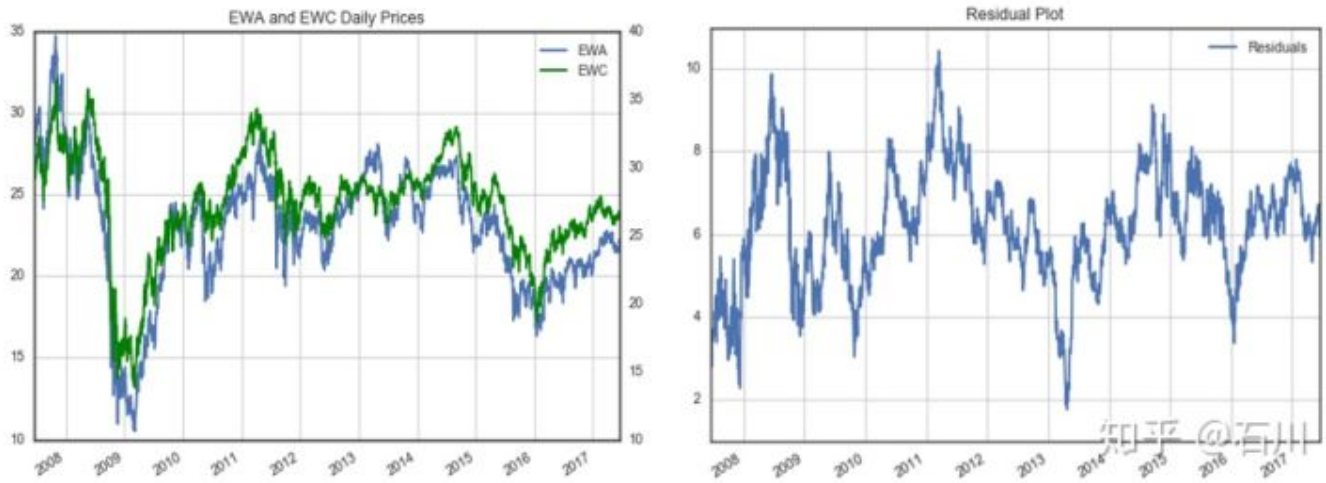
2 协整

我们都知道收益率序列满足平稳性，而价格序列不满足平稳性。收益率满足平稳性仅仅说明价格呈现随机游走，它对于构建赚钱的投资策略没什么用。我们想要的是价格序列呈现出平稳性。但凡事都有一个例外：**虽然单一投资品的价格不满足平稳性，但有时把多个投资品（通常是两个）线性组合在一起构成一个价差序列，该序列满足平稳性**。

在数学上，如果多个非平稳的时间序列通过线性组合得到一个平稳的时间序列，则把满足这种关系称为**协整（co-integration）**。我们在[《小心伪回归发现的假关系》](#)中介绍了协整发生的原因。

对于两个投资品的价格序列，它们的价格都表现出了一定的随机性。**如果它们的随机性来自同一随机过程（共同的因素），则说它们满足协整关系，并可以通过一定的线性组合把这个共同的**

EWA 和 EWC 是一对儿满足协整的经典例子。它们分别代表澳大利亚（EWA）和加拿大（EWC）股指的两个 ETFs。由于这两个国家的经济都主要依靠商品，因此可以认为这两个股指的波动来自共同的因子。下图为这两个 ETFs 的价格序列（左图）和回归得到的价差序列（右图）。



对价差进行 ADF 检验，得到的统计值为 -4.09（p-value 为 0.0065），小于显著性 1% 对应的阈值 -3.96，因此在 1% 的显著性水平下拒绝原假设。ADF 检验说明该价差序列满足平稳性，即 EWA 和 EWC 满足协整关系。

3 价格距离方法

再来看看另一种方法：价格距离法。该方法的主要参考文献是 Gatev et al. (2006)。该文的作者指出，寻找配对的一种方法是看两个投资品标准化后的价格序列是否足够接近——**即在寻找配对的形成期这两个投资品标准化后的累积价差是否足够小。**

关于为什么采用这种方法找配对，该文作者给出的解释如下：

We use this approach because it best approximates the description of how traders themselves choose pairs. Interviews with pairs traders suggest that they try to find two stocks whose prices “move together”.

译：我们使用这种方法，是因为它最接近交易员们描述的配对方法。对配对交易的交易员的采访表明，他们试图找到两支价格“一起移动”的股票。

假设形成期的长度为 T ，期内的某投资品 i 的价格序列为 $P_{i0}, P_{i1}, \dots, P_{iT}$ 。在这种方法中，首先以 P_{i0} 为基准将价格序列转化为对数价格（这相当于把 t 时刻的价格转化成 0 到 t 之间的对数收益率），然后再计算 T 期内任意两个投资品 i 和 j 的价格偏差平方和（记为 D_{ij} ）：



知乎

首发于
川流不息

$$D_{ij} = \sum_{t=1}^T (s_{it} - s_{jt})^2$$

在没有任何限制的版本中，任意两个投资品（比如两支股票或者两种不同的商品期货）都会被使用这种方法来计算它们之间价格走势的相似程度。在有限制的版本中——或者说是从业务上出于防止过拟合的考虑——上述计算仅仅被应用于同一行业的股票（比如不同的银行）或者同一种类的商品期货（比如农产品或者化工），以杜绝仅仅因为巧合而没有业务逻辑支撑的配对儿（Bianchi et al. 2009）。

当所有潜在的投资品两两计算价格距离之后，距离最小的那些配对儿就被认为是满足价格一起运动，即它们的价差会在未来一段时间内呈现均值回归，因此被用来交易。交易的规则就是经典的统计套利，当价差的取值较历史均值偏离两个标准差时，就进行做多或者做空价差的操作。下图是 Gatev et al. (2006) 给出的一个按上述方法找出的股票配对交易的例子。

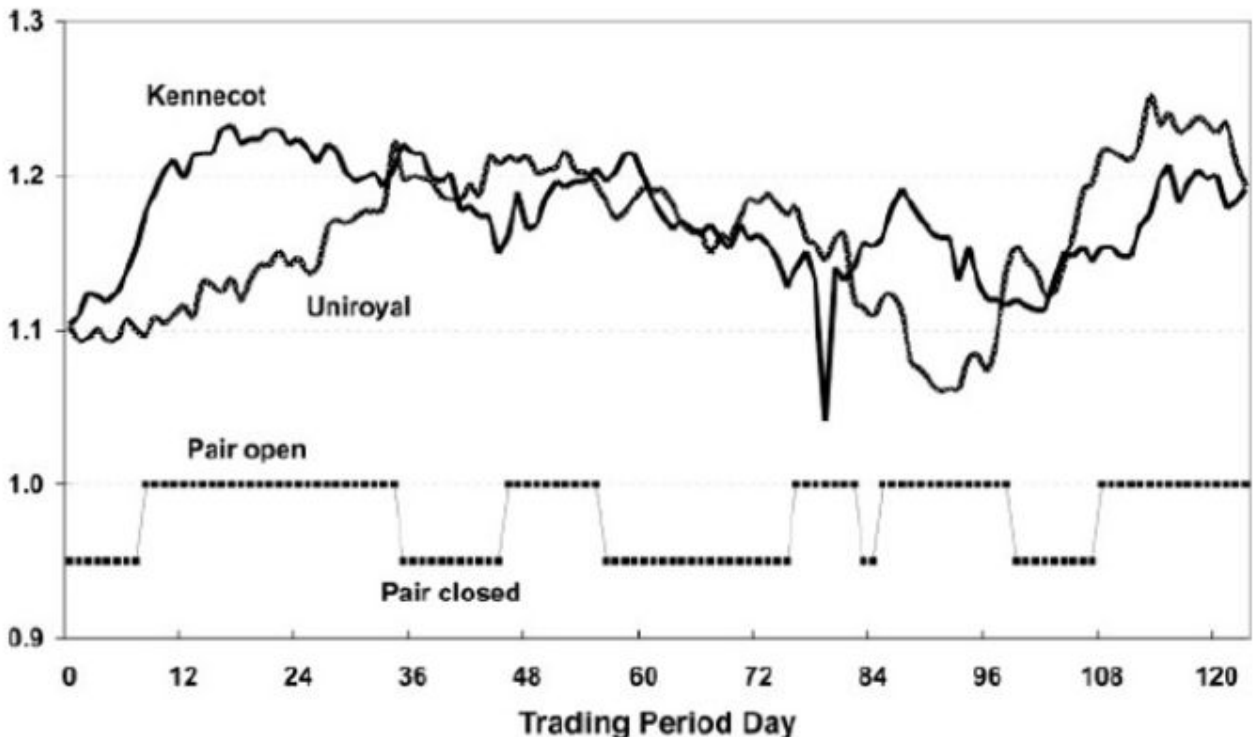


Figure 1
Daily normalized prices: Kennecott and Uniroyal (pair 5)
Trading period August 1963–January 1964.

出处：Gatev et al. (2006)

知乎 @石川

我们使用国内商品期货数据简单测试过这种配对儿方法。结论是，它在形成期内确实能找到价格走势非常接近的商品配对儿，但是这些商品在交易期内（即相对配对儿来说是样本外）的走势相当不一致。事实上，我们发现没有任何业务层面的机制来约束它们继续一起移动，因此其价差也就根本不满足均值回归，上述方法在国内市场上的有效性仍然是个大大的问号。

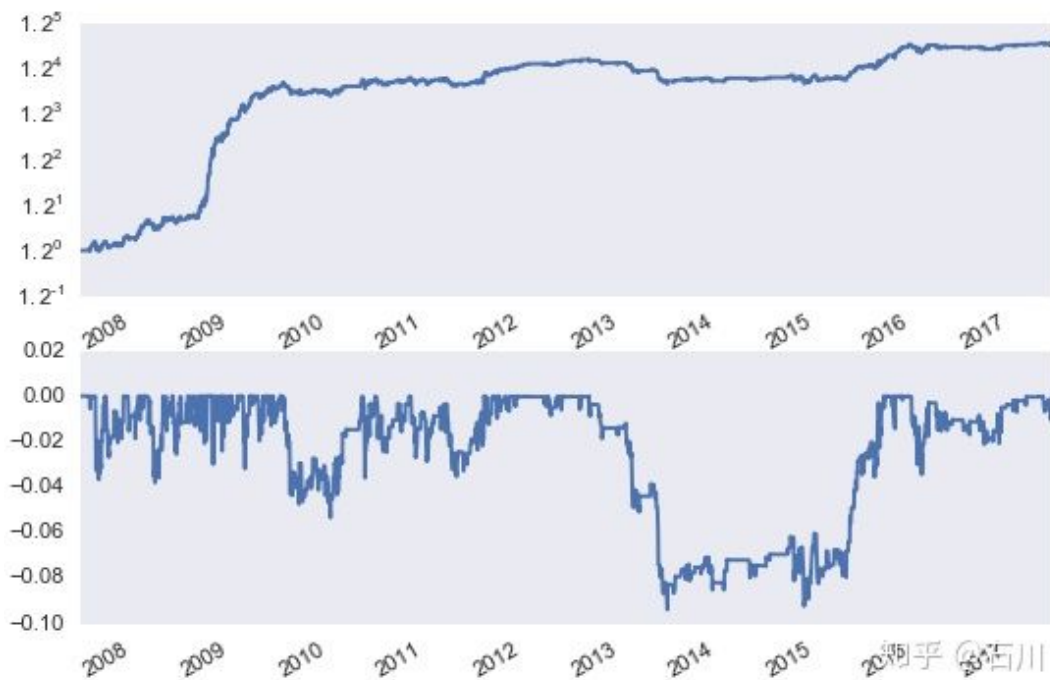
4 配对交易的不足

知乎

首发于
川流不息

来看前半句，即便是能找到两个靠谱的投资品，使它们的价差呈现稳定的均值回归特性，基于这个价差的策略的交易次数也会非常少。这是因为只有当价差偏离到一定的程度（比如 2 个标准差之外），策略才可能进行交易，而这种偏离发生的频率非常低。所以，这类策略通常在很长时间内都没有任何交易。

下图是利用 EWA 和 EWC 这两个 ETFs 构建的配对交易。该策略的年化连续收益率为 8.72%，最大回撤 -9.38%。从净值和最大回撤曲线中看出很大一部分收益来自 2009 年；在 2013 年到 2015 年间，策略发生了长达 758 个自然日的回撤。



再来看看后半句。由于被用来配对的投资品之间很难满足一价定理，所以没有任何金融业务上的核心逻辑来保证价差会一直满足均值回归。比如 Brent 和 WTI 这两种原油，它们的价格走势应该非常接近。而这二者的价差（下图）在 2011 年之前也确实呈现出稳定的均值回归特性。但是从 2011 年开始风云突变，这两种油的走势就不再一致，它们的价差也几乎没有表现出任何回归的现象。不难想象一个交易该价差的策略在价差突破历史阈值之后不但没有回归、反而持续迅速扩大时的绝望。



知乎

首发于
川流不息

Sep 07 2017, 12:48PM EDT. Powered by YCHARTS

基于上述原因，大量的实证显示，这类策略的收益风险比较差。

既然找到多个投资品并构建一个满足均值回归的价差不是特别靠谱，那么有没有更好的办法呢？究其核心，构建一个均值回归策略需要的是找到一个**具有负的自相关性的收益率序列**。正是由于投资品的**绝对收益率**几乎是随机的（没有统计上显著的负相关性），人们才退而求其用不同的投资品构建能够回归的价差（使价差收益率满足负相关性）。

虽然投资品的绝对收益率几乎是随机的，但是在**正确的方法下**，**投资品的残差收益率是可以用来构建均值回归策略的**。这就是均值回归的新思路。

5 因子残差法

Yeo and Papanicolaou (2017) 提出了针对投资品（这里特指股票）残差收益率的均值回归策略（该研究很大程度上受到了 Avellaneda and Lee 2010 的启发），用几句话来高度概括一下就是：

个股的收益率中有能够被公共因子解释的部分（比如市场因子、盈利因子、规模因子、估值因子等），把这些公共部分刨去之后剩下的就是残差收益率。对所有个股的累积残差收益率做随机分析，找到残差收益率的变化过程中满足均值回归的那些个股，进行交易。

下面我们结合上述过程中涉及到的量化模型来对每一步做一个简要介绍。

首先是通过多因子模型来求出个股的残差对数收益率（Yeo and Papanicolaou 2017 使用了统计因子，即认为因子本身是观测不到的，而是从股票的收益率中直接提取的——比如利用主成分分析）。这里使用对数收益率是为了方便之后把残差收益率直接求和得到“残差对数价格”的随机过程（下文略去“对数”二字）。假设 u_{it} 是股 i 在第 t 期的残差对数收益率，将 0 到 t 期内残差收益率加起来就得到 t 时刻的累积对数残差价格：



接下来，对 X_{it} 使用 **Ornstein-Uhlenbeck (OU) 过程**建模。OU 过程由 Leonard Ornstein 和 George Eugene Uhlenbeck 提出，该随机过程满足如下的随机微分方程：

$$dX_{it} = \theta_i(\mu_i - X_{it})dt + \sigma_i dW_{it}$$

其中， θ_i ， μ_i ，以及 σ_i 分别是针对个股 i 的模型参数， W_i 是布朗运动。和标准的布朗运动相比，这个模型的不同之处在于上式右侧的第一项，其中 μ_i 表示 X_{it} 的长均值。这一项说明，如果 X_{it} 大于（小于） μ_i ，那么在下一时刻的增量 dX_{it} 会是负的（正的），从而使 $t + 1$ 时刻的 X_i 倾向于向均值移动，而 θ_i 则代表像均值移动的速度，因此 **OU 过程满足均值回归特性**。

当然，众多个股之中，哪些的残差价格更好的满足 OU 过程需要通过参数估计来确定。为此，Yeo and Papanicolaou (2017) 考虑了三个要求：

1. 入选个股的 θ_i 必须足够大，代表着它的残差价格序列的均值回归速度很快。
2. 入选个股的模型参数的误差必须足够小，这样才能保证统计上的可靠性，如果估计误差很大，那么参数是不可信的。
3. 必须随时间的变化滚动建模，因为没有任何机制来保证一支个股的残差价格能够一直满足均值回归。

满足上述条件的个股则脱颖而出，它们将被用来构建最终的投资组合。

在构建投资组合时，最核心的一点是：**整个策略针对的都是个股残差价格的均值回归。因此，在投资组合中必须让入选的个股满足市场中性，更精确地说是满足计算残差收益率时使用的风格因子中性**。如果无法保证这一点，即便残差价差回归了，但是因为在个别因子上仍有暴露，导致的股价走势可能仍然和策略的建仓方向（多、空）不同、产生亏损。

具体的，在构建最优投资组合时，Yeo and Papanicolaou (2017) 考虑了如下因素，其中第一个就是因子中性（假设一共有 p 个因子），此外还要求了 dollar 中性（即多空对应的资金量相同），以及总杠杆的限制。



知乎

首发于
川流不息

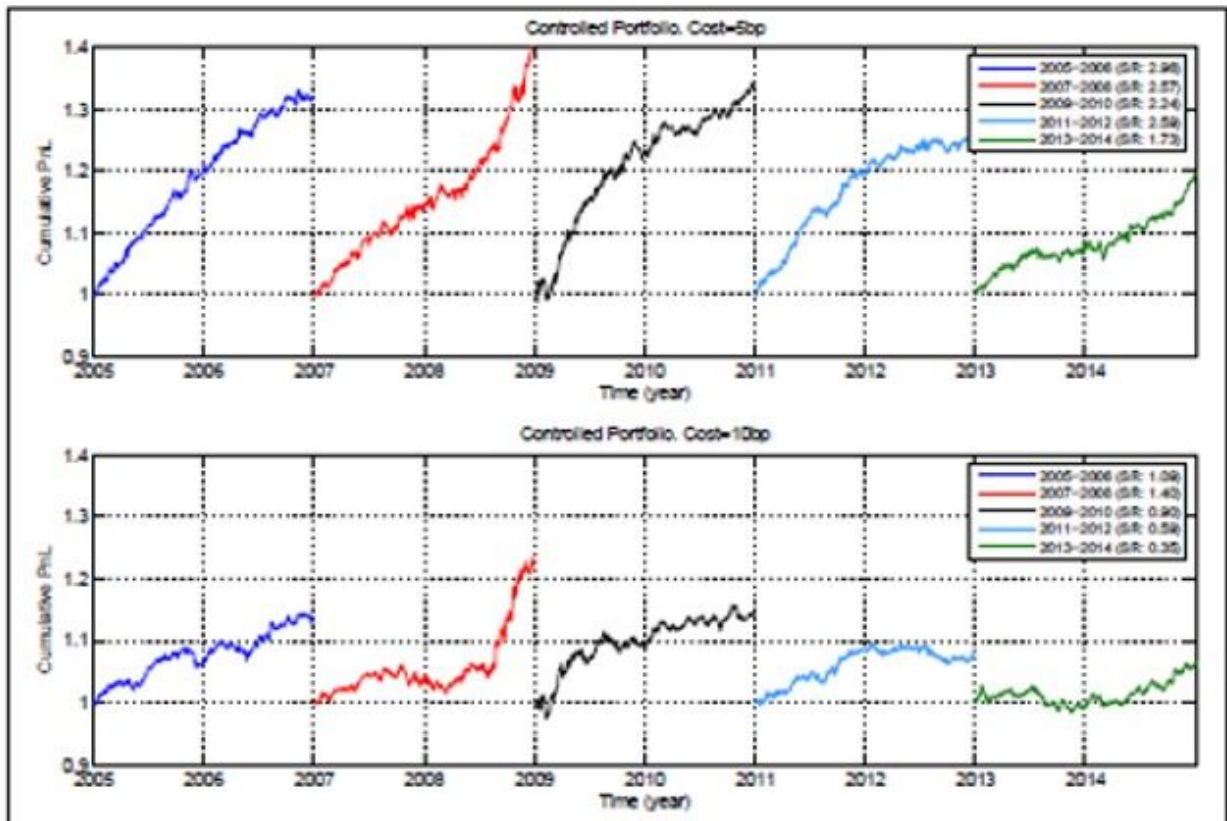
$$\begin{aligned}
 \text{Dollar-neutrality:} \quad & \sum_i q_i = 0 \\
 \text{Total Leverage:} \quad & \sum_i \|q_i\| = I \\
 \text{Long/Short:} \quad & \text{sgn}(q_i) \text{ is given by trading signals}
 \end{aligned}$$

上面的最后一条是开仓的方向（多、空）。在计算开仓信号时，Yeo and Papanicolaou (2017) 采用了传统的统计套利的思路，即计算信号 S_{it} :

$$S_{it} = \frac{X_{it} - \mu_i}{\sigma_i}$$

当 $S_{it} > 1.25$ 时，做空该个股，之后当 S_{it} 回归到小于 0.5 时平仓；当 $S_{it} < -1.25$ 时，做多该个股，之后当 S_{it} 回归到大于 -0.5 时平仓。

最后来看看效果。下图给出了这个均值回归策略在不同时间段的净值曲线。在 5 basis points 和 10 basis points 的交易成本假设下，该策略在不同的时间段（包括金融危机）均取得了正收益，效果还是很不错的。



出处：Yeo and Papanicolaou (2017)

知乎 @石川

知乎

首发于
川流不息

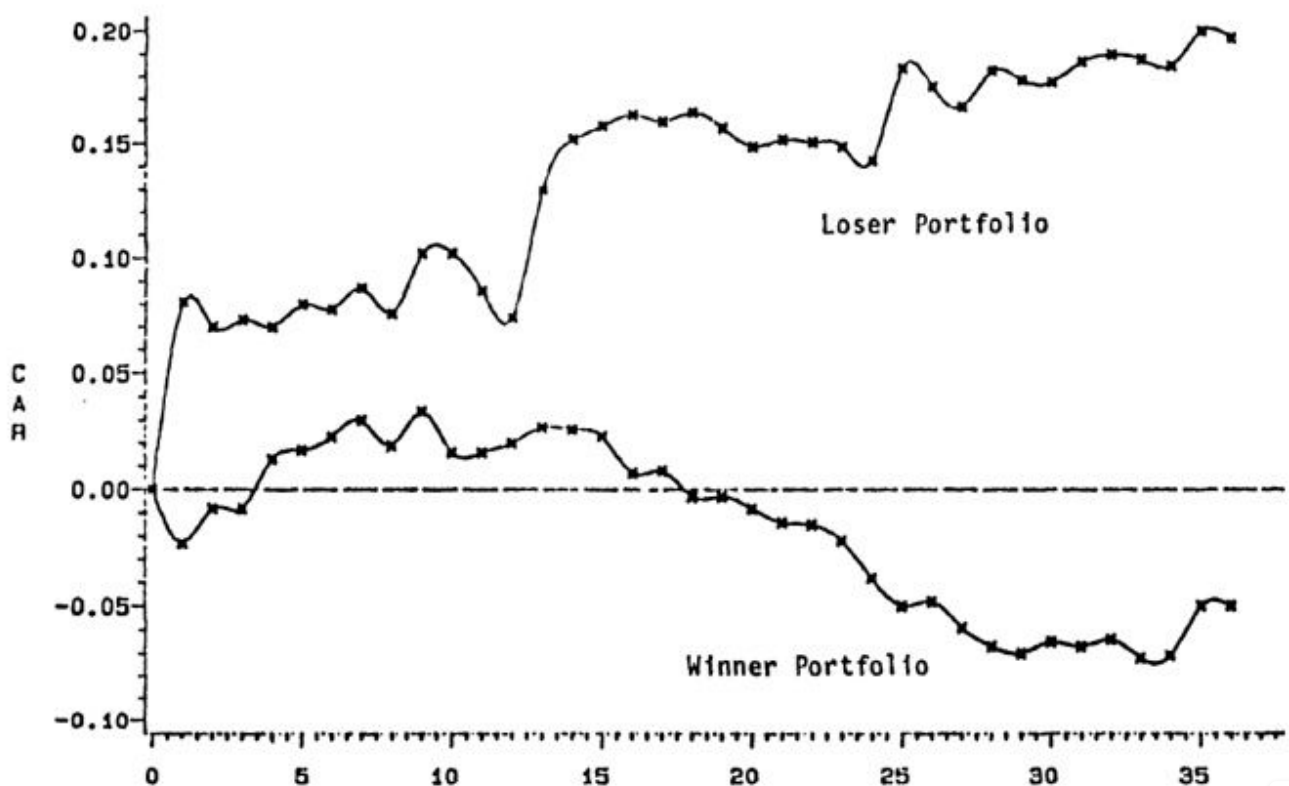
1. 优异的回测结果是否多少受益于回测期内的 data mining? 毕竟这么多数据一通优化, 而在现实中没有必然的因素来保证残差价格一定会实现均值回归;
2. 实证中的投资组合同时多、空开仓, 而在国内市场做空受限的前提下, 这个策略想要原封不动移植到 A 股上几乎不可能。

不要忙着悲观, 也许 Yeo and Papanicolaou (2017) 文中的结果有一定 data mining 的成分, **但是残差价格这背后的回归是有合理的解释的 —— 来自行为金融学的解释**。残差收益率的负相关性是由于投资者的过度反应造成的, 而从某种意义上说, Yeo and Papanicolaou (2017) 是针对残差收益率的潜在负相关性做了更加精密的量化分析。

下面就来看看来自行为金融学的例子。

6 来自行为金融学的实证

早在 1984 年, Richard Thaler (2017 年诺贝尔经济学获得者, 以研究行为经济学而闻名) 和 Werner De Bondt 写了一篇影响深远的文章, 题为 “Does the stock market overreact?” (De Bondt and Thaler 1984, 被引用近 8300 次)。文中, 他们使用**股票相对于市场的超额 (残差) 收益率**构建了一个**赢家组合 (残差收益率为正)**和一个**输家组合 (残差收益率为负)**。数据显示, 这个输家组合在未来取得了比赢家组合 (定期调仓) 更高的收益, 如下图所示。



出处: De Bondt and Thaler (1984)

知乎 @石

在计算残差收益时，DeBondt and Thaler (1984) 仅仅使用个股的收益率减去市场的收益率，并没有考虑其他因子甚至是个股的 β 。造成这种现象的原因正是投资者的过度反应。

这个简单的策略无需使用上一节介绍的那个复杂量化框架，但它利用了同样的实质——**个股的残差收益率具备负相关性**。在 A 股上，也很容易做简单的验证。以沪深 300 的成分股为例，按照个股相对于沪深 300 指数的残差收益率的 20 日均值从小到大排序分成 5 组、定期调仓，这 5 个投资组合的收益率很好的满足单调性（下图）——**残差收益率均值最低的组合收益最高**。



上面的结果说明，个股残差收益率在一定程度上确实有负相关性。此外，残差最小的组合也显著跑赢了沪深 300 指数，说明纯多头的策略也是有价值的。所以，即便无法轻易做空个股，Yeo and Papanicolaou (2017) 提出的这个构建均值回归策略的量化思路仍然可以被借鉴。

7 结语

知乎

首发于
川流不息

搭感觉怪怪的，因此还是把英文拿掉换成中文，并蹭热点（你懂的）用了个“次世代”。

至少在我看来，Yeo and Papanicolaou (2017) 这个模型没有重新定义“次世代”，也没有重新定义“重新定义”。

一个均值回归策略能否有效取决于它是否能够保证价格时间序列回归的机制。大量的实证显示，纯粹基于不同投资品的价格数据来找这种回归是不太靠谱的。而行为金融学提供了全新的视角。**流水的投资者，铁打的认知偏差**。这些根深蒂固的认知偏差产期存在于市场之中，使得投资品的价格和收益率表现出特定的性质。如果这些行为偏差能够被我们所用，从行为金融学的角度构建均值回归策略也许会大有可为。

以某个显著的市场特性作为切入点，使用最恰当的数学工具，更加精确的实现一个交易目标、并获取优秀的风险收益，这就是量化投资的最大意义。

参考文献

- Avellaneda, M. and J-H. Lee (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, Vol. 10(7), 761 – 782.
- Bianchi, R., D. Michael, and R. Zhu (2009). *Pairs trading profits in commodity futures markets*. In: Proceedings of Asian Finance Association 2009 International Conference, June 30 – July 3, 2009, Brisbane, Queensland.
- De Bondt, W. F. M. and R. Thaler (1984). Does the stock market overreact? *The Journal of Finance*, Vol. 40(3), 793 – 805.
- Gatev, E., W. N. Goetzmann, and K. G. Rouwenhorst (2006). Pairs trading: performance of a relative-value arbitrage rule. *The Review of Financial Studies*, Vol. 19(3), 797 – 827.
- Yeo, J. and G. Papanicolaou (2017). Risk control of mean-reversion time in statistical arbitrage. *Risk and Decision Analysis*, Vol. 6(4), 263 – 290.

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 2019-07-02

金融 统计套利 量化交易

▲ 赞同 74 ▼ 16 条评论 分享 收藏 ...



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



策略不给力？来一发卡尔曼滤波

Fitz Hoo



统计套利 Review and Outlook

郭嘉公元



TIDE 数字

徐江流

16 条评论

⇌ 切换为时间排序

写下你的评论...



陈陈陈

1 年前

很奇怪为什么08年已经被证伪的东西还是那么多人趋之若鹜。。。

👍 1



石川 (作者) 回复 陈陈陈

1 年前

愿闻其详，请不吝赐教。

👍 1



潇洒的一把梭 回复 陈陈陈

1 年前

08年不是就单边的惨嘛

👍 赞

展开其他 2 条回复



知乎

首发于
川流不息

不相悖吗？

赞



石川 (作者) 回复 飞飞

1 年前

Momentum是用绝对收益率看动量强弱，不是残差收益率。另外，动量和反转策略一般的周期是不一样的。即便看绝对收益率，它也在不同时间尺度上表现出不同的动量或者反转特点。可以参考：mp.weixin.qq.com/s/XYKj...

赞



飞飞

1 年前

同意二者的周期不同。但是不太理解你说的momentum是看绝对收益的，如果用momentum factor来构建long short portfolio的话，不是也看的residual吗？

赞



石川 (作者)

1 年前

理论上是这样，但是构建动量因子的时候要非常精细，比如用 barra 那套。用多空排序构建出的动量因子难免在其他因子上面有暴露，所以不经过仔细的分析，就不知道这个组合在其他因子或者市场上是否还有暴露（从而造成动量因子有效），也就很难说是不是残差的部分在起作用。从我接触的资料和实证来看，我更倾向于相信残差收益率是满足均值回归的。我们以后再讨论哈。

1



飞飞

1 年前

谢谢

赞



石川 (作者)

1 年前

这篇文章因为一点侵权误会被知乎关小黑屋 1 天，终于又放出来了。感谢知乎。。。

1



stardust

1 年前

好文章。

赞



吴二真

10 个月前

这是在wq被玩烂了的alpha...

赞



石川 (作者) 回复 吴二真

10 个月前

谢谢 受教啦！\



黑猫Q形态

10 个月前

所谓残差就是波动

👍 1



黑暗没有骑士

4 个月前

你好，特异收益率真的满足mean reverting吗？这里的特异收益率是不是必须把momentum的exposure给去掉之后才会比较容易满足stationary

👍 赞

