

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

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

0 引言

前些天，有个小伙伴问了这么一个问题：**现在量化选股因子泛滥，大家都用同样的数据做同样的测试，最后也得出相似的结论，导致很多策略同质化。这个问题量信怎么看？**

今天这篇文章就来回答这个问题。

首先，**一个因子能够挣钱是因为市场在该因子方面存在非有效性**。用的人越多，市场在那方面越有效，导致因子的效果越差。但就现阶段而言就说市场在哪个因子上已经完全有效为时尚早。事实上，由于 A 股市场中噪声投资者的高度参与感，市场仍远不够有效。

拿我们熟悉的价值投资来说，价值投资有效的直接原因不是相关因子的额外风险溢价补偿（事实上，价值投资构建的投资组合风险较市场往往更低），而是由于噪音投资者的存在（见《写在当越来越多人谈论价值投资时》），这是不会消失的，所以它会持续有效。

在因子仍将会有效这个大前提下，我想更系统的谈谈第二个方面：**同样是使用主流的因子（比**

值、动量、质量等），有的人赚钱了，有的人却没赚钱。虽然是同质化的策略，但因子在靠谱广

答案很简单：**细节决定成败。**

随着量化选股的普及，市面上出现了很多优秀的回测平台，能够快速测试选股因子是否有效。然而，无论测试再怎么方便快捷，一旦真要实盘时，会有一些不得不面对的问题：每个因子应该选多少股票？不同因子如何取舍？股票之间的仓位如何确定？调仓频率如何选取？交易时如何降低成本？等等。因子投资的成败取决于能否很好的应对这些问题。

本文旨探讨在因子投资实践中必须考虑的细节问题。需要说明的是，这里面的某些问题并没有标准答案、不同的选择和投资者的风险偏好有关。

1 细节一：选择合适的代理指标

在确定风格因子后，首先要决定的就是找到合适的代理指标。我们以价值 (value) 因子为例说明这个问题。

顺便提一句，现如今我们提到价值投资的时候，由于巴菲特精神深入人心，往往指的是“好公司”+“便宜”。但是，最初的价值投资就是买便宜的。便宜的定义就是公司每单位市值的基本面价值更高，而不在乎基本面的好坏，这叫做 *pure value*。

当我们使用价值这个因子来选股时，必须通过具体的指标来对股票的便宜程度排名。这就是选择因子的代理指标。在衡量便宜方面，有很多常见的 multiples (见下面的小贴士)，比如 P/E, P/B, EV/EBIT 等。

在英文的资料中涉及价值因子的指标时，*multiple* 是个非常常见的词汇，因此拿出来解释一下，方便小伙伴们以后阅读英文文献。*multiple* 一词是倍数的意思，顾名思义，指的是两个指标相除。按照惯例，一般分子上指标的数值要大于分母上指标的数值。比如市盈率 P/E，它就是每股股价和每股 *earnings* 的比值，这就是一个 *multiple*；而 EV (Enterprise Value) 和 EBIT (*earnings before interest and taxes*) 的比值也是一个 *multiple*。

面对同一因子的多个代理指标时，我们可以选择回测时效果最好的那个，也可以使用多个指标同时选股。单一指标可能由于财报数据的不准确而产生噪声；使用多个指标可以避免这个问题。从实证来看，选择多个指标通常能提高该因子的选股效果（但这不是绝对的）。不过，由于同一因子的不同指标之间相关度较高，因此也没有必要选择太多的指标。

以中证 500 为例，使用 P/B 和 EV/EBIT 两个指标作为价值因子选股。简单回测的效果如下表所示。无论是从策略的风险收益特征，还是因子本身的 IC 和 IR 来看，同时使用两个因子都优于单一因子。



知乎

首发于
川流不息

因子 IC 均值	0.056	0.047	0.064
因子 IR	0.312	0.404	0.420
年化收益	22.22%	15.56%	22.70%
最大回撤	-67.91	-74.86%	-68.27%
夏普率	0.54	0.34	0.55

因子 IC：初期股票因子排名和当期收益率排名的相关度，越大说明因子预测性越高

因子 IR：IC 均值与IC 标准差之比，越大说明因子越稳定

知乎 @石川

当然，同时使用多个代理指标虽然提高了效果，但我们也会问这里面是否有数据挖掘的成分呢？在这方面，著名的量化投资基金 AQR 告诉我们，使用多个指标并不是一种因子激增（因此没有数据挖掘问题），而是提高因子健壮性的一种方法，因为无论哪个单一指标都无法完美的代表我们的目标因子。

It is important to note that using multiple measures is not a form of factor proliferation, which can lead to concerns about data mining; instead, using additional measures leads to a more robust version of the ideas behind the factors as there isn't a single, perfect definition of each style.

这种处理类似于机器学习中的集合学习算法，它和随机森林以及 AdaBoost 算法比单一的决策树算法分类效果更好有异曲同工之妙。

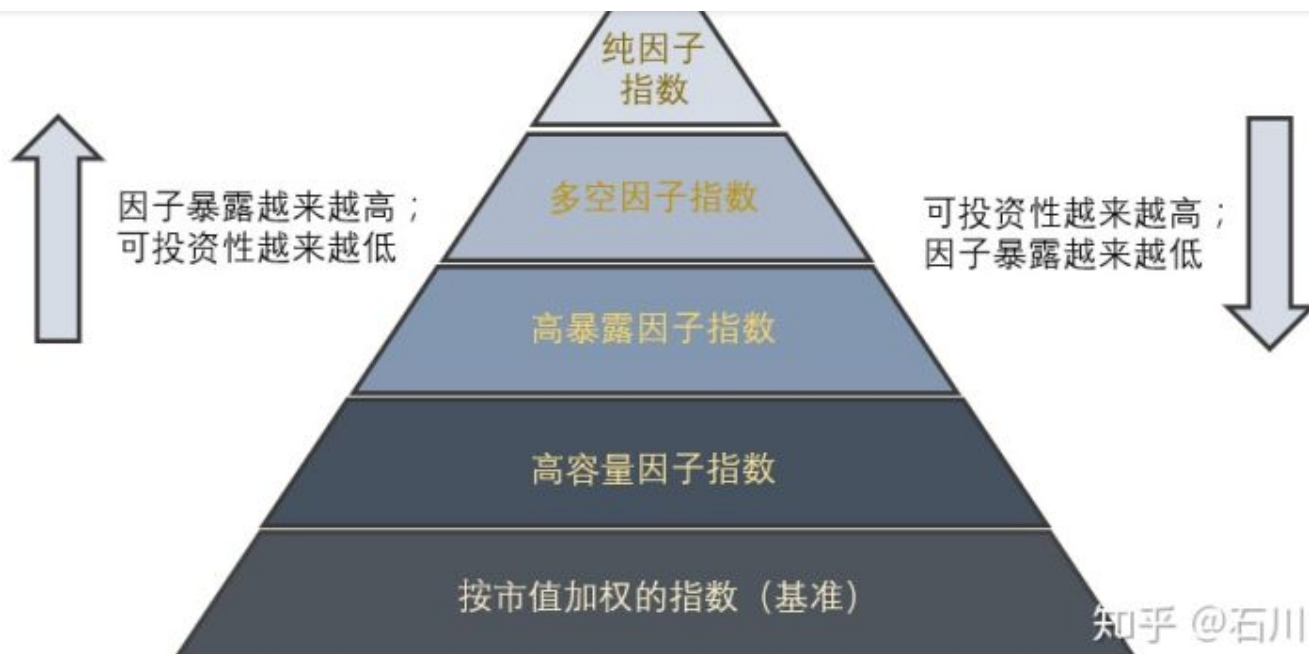
2 细节二：构建因子投资组合

在使用因子选股、构建投资组合时，**因子暴露程度**和**可投资性**是必须考虑的两个因素，然而高的因子暴露是通过牺牲可投资性得到的，我们必须在二者之间取舍。

可投资性是指投资组合中股票的仓位是否合理，该组合的换手率和交易成本是否实际，进入该组合的股票是否有足够的流动性、该投资组合能承担的资金量（即投资组合的容量）是否足够大等。

下面的金字塔图描绘了五种构建因子投资组合的方法。自下而上，它们的因子暴露越来越高，而**可投资性却越来越低**。《因子投资——“被动的”主动投资》曾对这五种方法进行了详细的阐述，本文只是简要说明一下。





在金字塔底部是市场组合，由所有股票按市值加权构成。我们认为市场组合中因子的暴露为零。往上移动是高容量因子组合——这里的容量是就该组合中股票支撑的资金容量而言。该组合不剔除任何股票，但是它们的权重不再由市值决定，而是由股票在该因子上的暴露决定。较市场组合而言，该组合在目标因子上有更高的暴露。继续往上是高暴露组合，它会剔除部分因子暴露度低的个股而集中于剩下那些因子暴露度更高的股票，使得组合的因子暴露度更高，但牺牲了一定的可投资性。**这三种组合都是纯多头组合。**

最上面两类指的是 Fama-French 三因子模型中的多空构建方法以及 Barra 的纯因子组合（见《[正确理解 Barra 的纯因子模型](#)》），它们更多的都是为了评价因子的效果。这两种组合都是多空组合，由于做空的限制，它们的可投资性比较差。尤其是 Barra 的纯因子组合，它在构建时没有考虑任何可投资性的限制，但它在风险管理中有着非常重要的作用，在本文的第六节还会提到它。

在因子投资实务中，通常的做法是根据个股在因子上的排序选出一小部分在该因子上暴露高的个股，构建一个投资组合。这种做法相当于上述金字塔中的高暴露组合法。

在选择股票时，值得考量的标准除了股票在该因子上的强弱外，还包括股票的市值（只有大市值才能支持大的资金容量）、所属的行业等因素（想要尽量分散行业从而规避行业特有的风险），目的是尽量构建一个行业中性且投资性高的组合。

这种方法简单、透明、逻辑清晰，然而它不容易控制指数对其他风格因子的暴露。比如我们针对价值因子构建的可投资性高的投资组合难免也会在其他因子上有一定的暴露。如果我们观察到因子的投资组合表现不佳，这可能不是因为该因子失效造成的，而是由该组合在其他因子上的暴露带来的。

3 细节三：多个因子如何选股？

1. 每个因子独立选，然后把选出来的股票放在一起。这个方法称为 portfolio mix。
2. 使用所有因子给股票打分，每个股票得到一个总分，然后按照总分的高低选择。该方法称作 integrated approach。

这两种方法并无一定谁对谁错。但它们可能会选出完全不同的结果。

考虑下面这个假想的例子，它使用价值因子和盈利因子选股，股票池中的股票在每个因子上的打分从好到坏被分为 A、B、C、D 四挡。

		价值因子			
		A	B	C	D
盈利因子	A				
	B				
	C				
	D				

注释：两种方法选股，分别为单个因子选，和两个因子综合选

蓝色表示被两种方法同时选出

红色表示被综合法选出

黄色表示被单因子法选出

知乎 @石川

当我们使用 portfolio mix（单个因子独立选股）时，对于价值（盈利）因子，会把所有在该因子上得分为 A 的股票选出来然后放在一起。**该方法不看每支股票的综合实力，而更在乎它们是否有“偏科”（在特定因子上暴露高）**。按此种方法，我们会选出落入上图中蓝色和黄色格子里的那些股票。

另一方面，当使用 integrated approach（多个因子综合打分）时，只有股票在这些因子上的表现都不错时，才会被选入（比如在两个因子上的得分都不低于 B）。**该方法考察的是每支股票的综合实力，它会惩罚“偏科”选手**。使用此方法，我们会选出落入上图中蓝色和红色格子里的股票。

重申一遍：**这两种方法并无一定谁对谁错**。取舍的角度是到底想通过因子投资实现什么样的目标，以及策略的收益风险比和交易成本。以价值投资为例，它要求股票既满足盈利高又要便宜，因此在盈利和价值两个因子上都要表现突出，这时可以选择的是 integrated approach。而如果我们的目标并不是传统的价值投资，而只是希望把投资组合尽可能的暴露在盈利和价值这两个风格因子之上，那么 portfolio mix 这种方法未尝不是一个更好的选择。

4 细节四：因子择时和因子配置

以中证 500 的成分股为选股池，构建下列五个风格因子的纯多头组合：流动性因子、规模因子、成长因子、价值因子、反转因子。这五个投资组合相对中证 500 的**超额收益**的风险收益特性如下。

	年化收益率	最大回撤	平均回撤	夏普率
流动性因子	8.10%	3.80%	0.76%	1.38
规模因子	11.94%	2.86%	0.45%	1.91
成长因子	8.47%	3.26%	0.46%	1.69
价值因子	6.91%	4.31%	0.98%	1.20
反转因子	7.34%	10.67%	1.46%	0.97

除了自身的优异表现外，这些因子之间的相关性也很低（下图）：



如同配置股票一样，**配置因子**也是量化投资中的一个重要课题。它希望把相关性低的多个因子通过**某种方式配置在一起，得到一个收益风险比更优的多因子配置组合**。

常见因子配置方法包括：简单多样化、波动率倒数、跟踪误差倒数、风险平价以及趋势追踪。的业务含义、实现方法以及相应的数学公式如下表所示。

简单多样化	不对因子的未来表现做预测，将它们等权配置在一起	赋予每个因子同样的权重	$w_i = \frac{1}{N}$
波动率倒数	绝对收益率（而非相对基准指数的相对收益率）的波动率较低的因子会带来更优的收益风险比	因子按照其绝对收益率波动率的倒数赋权	$w_i \sim \frac{1}{\sigma_i}$
跟踪误差倒数	相对基准指数的相对收益率的波动率（又称跟踪误差）较低的因子会带来更优的收益风险比	因子按照其相对收益率波动率（即跟踪误差，tracking error）的倒数赋权	$w_i \sim \frac{1}{TE_i}$
风险平价	平摊不同因子的风险贡献，从风险的角度实现多样化	按照因子对投资组合边际风险贡献率相等的基准赋权	令 σ_P 为投资组合的标准差，则所有因子的边际风险贡献（下式）相等： $w_i \frac{\partial \sigma_P}{\partial w_i}$
趋势追踪	过去一段时间表现好得因子未来也会有更优异的表现	以因子过去一段时间相对基准指数的收益率的信息率作为权重	$w_i \sim \frac{\tilde{\mu}_i}{\tilde{\sigma}_i}$ 知乎 @石川

除了简单多样化这种被动的静态配置方法外，其他四种配置方法都需要结合投资者的**主观判断**。它们从不同方面考虑因子在过去一段时间的表现，从而判断因子未来的表现，以此**动态**的配置因子在下一个配置周期的权重。

上述五种配置方法的效果如下图所示。可以看到，被动的简单多样化就非常优秀。而其他四中主动配置方法并没有带来显著的提高（有两种方法甚至不如简单多样化）。

	年化收益率	最大回撤	平均回撤	夏普率
简单多样化	8.68%	1.27%	0.13%	2.27
波动率倒数	8.70%	1.49%	0.14%	2.33
跟踪误差倒数	8.20%	1.39%	0.15%	2.17
风险平价	8.67%	1.42%	0.14%	2.28
趋势追踪	8.22%	2.28%	0.18%	2.03

单多样化非常接近，因此这两种方法的配置效果和简单多样化接近。

波动率的倒数配置策略与风险平价策略关注的都是因子**绝对收益率**的波动。反观跟踪误差倒数策略，它考量的是因子**相对收益率**的相关指标。在构建每个因子的投资组合时，我们在满足可投资性的前提下尽量地追逐投资组合在目标因子上的暴露；高的因子暴露意味着高的主动管理波动，即较大的跟踪误差。**较大的跟踪误差对因子的配置效果造成了负面的影响。**

最后来看看趋势追踪法。随着因子投资越来越流行，投资者在选择因子时往往看中的是因子最近的表现——选择那些最近几年收益率高的因子。对业绩的追逐造成大量资金涌入过去有效的因子，从而大幅**提高了该因子的估值**。这么做造成了两个问题：

1. 因子估值提升使得因子收益率被高估，容易使人们对因子的效果产生错误的预期。在任何理性投资中，哪怕一个标的再好，我们都不应该不计成本的买入。
2. 历史数据及经济规律表明，估值满足均值回归。一旦在过去表现过热的因子的估值回归，那么就会大大降低它在未来的收益率。

Arnott et al (2016) 指出因子收益率中包括很大的**估值溢价 (valuation premium)**。他们把由估值上升带来的因子收益率称为环境 α 。**当剔除了因子估值的变化后，很多因子竟然并不能获得超额收益。**Arnott et al. (2017) 的研究发现**选择那些估值处于历史低位的因子（即过去表现的不怎么好的因子）比选择那些过去一段时间过热的因子，能够在未来获得更高的收益。**这解释了为什么按照趋势追踪来配置因子效果并不好。

使用哪个因子是个因子择时问题；如何将资产分配到不同因子上是个因子配置问题。在实际投资中，任何主观的择时、配置方法都会比简单多样化这种被动方法带来更多的主动误差；这增加了交易难度、并有可能提高换手率、增加交易成本。从这个意义上说，被动的简单多样化足够优秀，可以胜任配置多个因子的重任。

5 细节五：调仓频率和交易

在确定了使用的因子和如何根据因子挑选股票之后，紧接着的环节就是交易了。**交易涉及调仓的频率以及交易日内买卖如何成交。**

在调仓频率方面，可以采用定期（每隔固定的一段时间）和每日调仓。从大量实证经验来看，如果把调仓的粒度（周期）定为大于 1 个交易日，即**不是每天调仓，那么策略的回测效果受实际的调仓日期影响较大（因为在一些特定的日期能恰好选出一些特别优秀的股票，而在其他日期却没选出）**。这一方面自然说明如此的多因子策略是有问题的（对输入太敏感），但同时也说明这种较低调仓周期也有自身的问题。

虽然财务因子的变化频率很低、只有当新的财报被披露时才会更新，但是对于和价格有关的因子——比如价值因子或者规模因子，由于价格每天都会变化，它们的数值也会频繁变化。每天刷新因

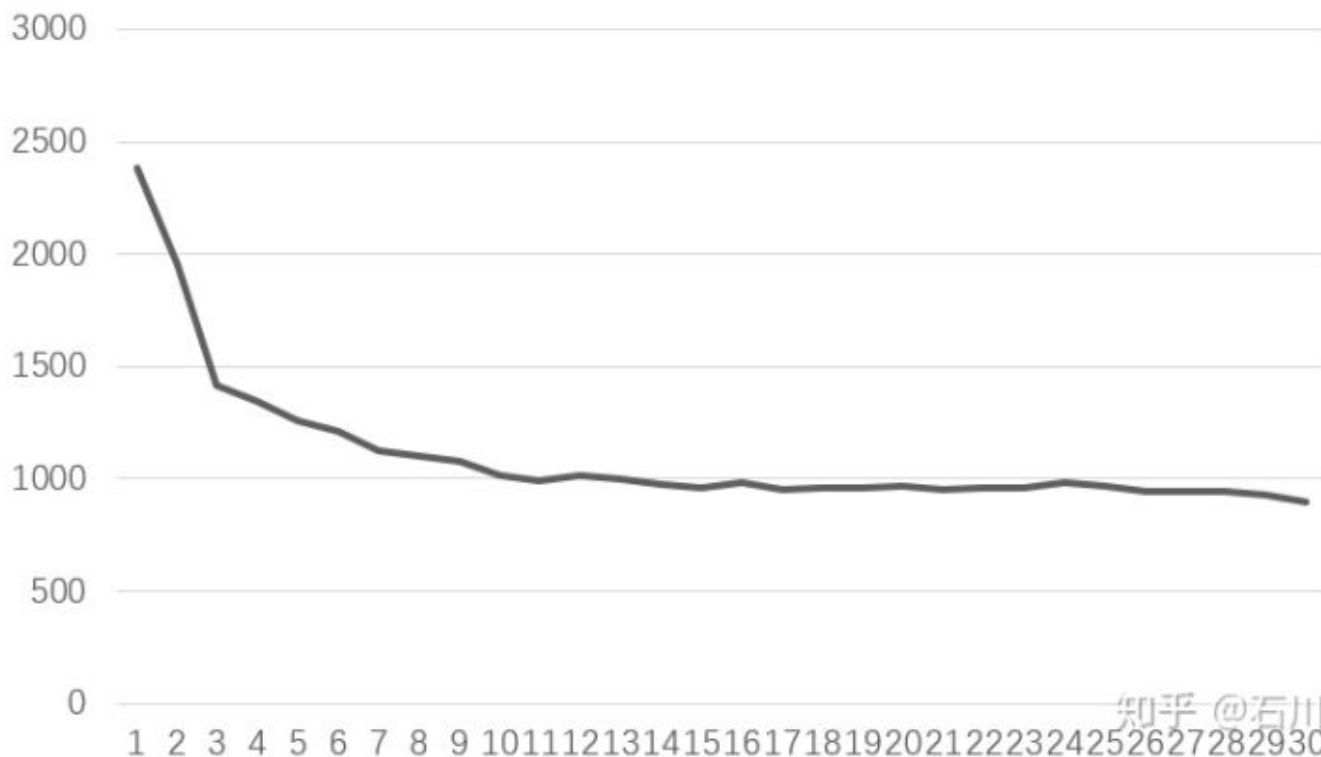


知乎

首发于
川流不息

关系。可以看到，当持股个数过少时 (≤ 5 支)，该策略的换手率非常高。当持股个数在 10 以上时，策略的换手率就比较稳定了（注：上述 10 这个数字和具体实证有关，不同的投资者应根据自己的多因子策略来寻找换手率稳定时持股个数的最小阈值）。

年化换手率 (%) 和持股个数的关系



除了换手率之外，每天交易需要交易者承受更大的心理压力。在当前的 A 股市场，自动化交易的限制还比较大，更多的是靠人来交易。人类的情绪、认知偏差等弱点不可避免的会给策略带来难以评估的不确定性，尤其是在交易频率高的时候。华尔街的前辈在谈及人对量化投资策略的影响时指出：**严格由机器执行得到的效果是量化策略的上限而非下限，任何人为的干预从长期来看都只能削弱策略的表现。**

选股个数太少的另外一个问题是风险无法充分分散。因子投资看中的是一揽子股票在风格因子上的共性。如果仅选择太少的股票，则和这个初衷背道而驰，而更多的是暴露在这些个股的特异性收益率上。此外，如果个股太少，会导致我们把资金重仓在几支个股上面。一旦持仓中的个别股票发生了较大的亏损，这会给交易者造成巨大的心理压力。

交易中的第二个问题是交易的执行，它涉及的是在交易日应该如何交易来减少滑点和交易费用。**对因子选股来说，通常的做法是在 T 日收盘后，根据最新的价格和财务数据计算出最新的股票仓单，然后在 T + 1 日择机交易。**那么我们是否应该 T + 1 日开盘交易？收盘交易？或者固定时点（比如上午 10 点）交易？亦或是按照 TWAP（时间平均）交易？一般来说，市场的成交量在交易日内呈现 U 型——在开盘和收盘波动较大，因此滑点较高。

下图显示了某个因子选股策略根据 T + 1 日不同价格交易的效果（均假设千一的单边交易费用）从收益率和夏普率来看，开盘价效果 > 10 点价格效果 > 中午收盘价效果 > 日均价效果 > 收盘价效果

	开盘价	收盘价	10 点价格	中午收盘价	日均价
年化收益率	28.97%	27.24%	28.75%	28.72%	27.75%
最大回撤	-39.15%	-36.86%	-38.65%	-37.43%	-37.61%
夏普率	0.98	0.91	0.97	0.97	0.93

知乎 @石川

6 细节六：风险管理

本文讨论的最后一个细节是**风险管理**。在**市场状态发生转换时会造成风格因子失效**。然而，更加困难的是想要判断出一个因子有效或者失效需要较长的周期。

比如规模因子在 2016 年之前一直有效（挑小市值），而在 2016 年之后则失效了。但我们之所以说它失效，是因为站在两年后的 2018 年回顾过去两年小市值的表现时发现了基于该因子的策略是持续下跌的。但显然在现实中，历经两年实打实的亏损（而且还亏的很惨）才得出因子失效的结论是很痛苦的。

前文提到，为了构建具有可投资性高的因子投资组合，该组合一定不可避免的在其他风格因子上有暴露。它们会对最终的选股带来额外的风险。**我们应该定期评测投资组合在不同风格因子上的动态暴露，进行风险归因和业绩归因。**

为此，Barra 的纯因子模型闪亮登场。纯因子模型是从数学角度出发构建的完美暴露于单一风格因子的投资组合，它虽然几乎不具备任何可投资性，但是在风险归因上具有重要作用。将选股投资组合的收益率在时序上用这些纯因子组合的收益率进行回归，就可以分析出选股投资组合的收益和风险成分中由哪些风格因子组成。

对于风险来说，Davis and Menchero (2010) 指出：

$$\sigma = \text{Exposure} \times \text{Volatility} \times \text{Correlation}$$

它的具体表现为下面这个数学关系：

$$\sigma(R) = \sum_m x_m \sigma(r_m) \rho(r_m, R)$$

其中 $\sigma(R)$ 是选股投资组合收益率的标准差，而等式右侧的三要素分别为：

- **Exposure (暴露)**：即 x_m ，它衡量投资组合对每个风格因子的暴露大小。在风险归因时



- **Volatility (波动性)**：即 $\sigma(r_m)$ ，它代表的是第 m 个风格因子收益率 r_m 的标准差。由于投资组合是暴露在不同的风格因子中，那么风格因子收益率的波动越大，它对投资组合的风险的贡献程度也越大。
- **Correlation (相关性)**：即 $\rho(r_m, R)$ ，它是风格因子 m 和选股投资组合收益率之间的相关系数。这种相关性越高，投资组合收益率受风格因子影响的确定性越显著。

定期对选股投资组合做风险归因有助于判断不理想的选股结果是来自选定的风格因子本身（比如风格切换了，风格因子赚不到钱了），还是因为投资组合暴露在其他因子上的造成的。对于后一种情况，则需要考虑重新构建投资组合从而尽量降低其在不必要因子上的暴露。

7 结语

本文抛砖引玉，介绍了一些在因子投资实践中必须要考虑的细节。不可否认，在大家都没有独门数据的今天，因子投资变得越来越同质化。但是，只要市场是非有效的，因子就还会有效，而因子投资的关键在于执行时对各种细节的精确处理。在这方面，量化投资无疑有着巨大的优势。

量化投资基金 AQR 针对因子投资中的这些问题也展开过相应的讨论，并把通过良好执行而真正实现因子投资收益这个过程称为“**craftsmanship alpha (手艺 alpha)**”，意思是只有那些专注于做好每个细节的优秀基金经理才能把因子投资在理论上的超额收益转换成现实。

虽然在这些细节中并不总是有一定正确或者错误的决策，但靠谱的基金经理应该能够为自己的选择辩护并深谙每个决策对于投资收益和风险的影响。在这方面，基于经济原则和经验证据的细节处理决策将在大概率上为因子投资带来更好的结果。

The devil is in the details.

参考文献

- Arnott, R. D., N. Beck, V. Kalesnik, and J. West (2016). How Can 'Smart Beta' Go Horribly Wrong? *SSRN*: ssrn.com/abstract=30409...
- Arnott, R. D., N. Beck, V. Kalesnik (2017). Forecasting Factor and Smart Beta Returns (Hint: History Is Worse than Useless). *SSRN*: ssrn.com/abstract=30409...
- Davis, B. and J. Menchero (2010). *Risk Contribution is Exposure times Volatility times Correlation*. Technical Report. MSCI Barra Research.

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。



编辑于 2019-07-02

多因子模型 量化交易 金融

▲ 赞同 74 ▼ 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



量化哥

一起学习量化分析，共享智慧与金融在大数据时代的红利！

关注专栏

推荐阅读



Alpha系列 (五) ——因子模型

BigQuant

Beta
Smart

从Seeking Alpha 到 Smart Beta (上)

王小仙




多因子模型子部分)

江嘉健

9 条评论 切换为时间排序

写下你的评论...



智商降为零

可以请问一下您在这篇文章里面用的回测区间吗?谢谢!

赞

1 年前

知乎

首发于
川流不息

09年1月到18年6月

赞



韩佳成

1 年前

非常感谢，最近在读您的每一篇文章，感觉每一篇的收获都很大！

赞



飞飞

1 年前

请问组合中的股票权重是怎么确定的？

赞



石川 (作者) 回复 飞飞

1 年前

我个人经验平均配然后定期再平衡就挺好，有助于样本外熵最大，即适应性强.....

赞



王二二

11 个月前

非常感谢带来这么精彩的文章。文中提到的因子配置方法是基于portfolio mix的吧，但是这些配置方法是否也可以用来为integrated approach计算权重呢？

赞



石川 (作者) 回复 王二二

11 个月前

可以，比如你可以用这几个因子等权给股票打分，然后选出分数最高的那些股票。有一篇最新的研究 (papers.ssrn.com/sol3/pa...)，它认为靠单因子选出的组合总有一部分（20%左右）因为在其他因子上的意外暴露而实际上是跑输基准的，这就削弱了因子的表现。它的论点是integrated approach更好。可以看看。

2 回复 踩 举报



王二二 回复 石川 (作者)

11 个月前

好的，非常感谢回复~

赞

展开其他 1 条回复

