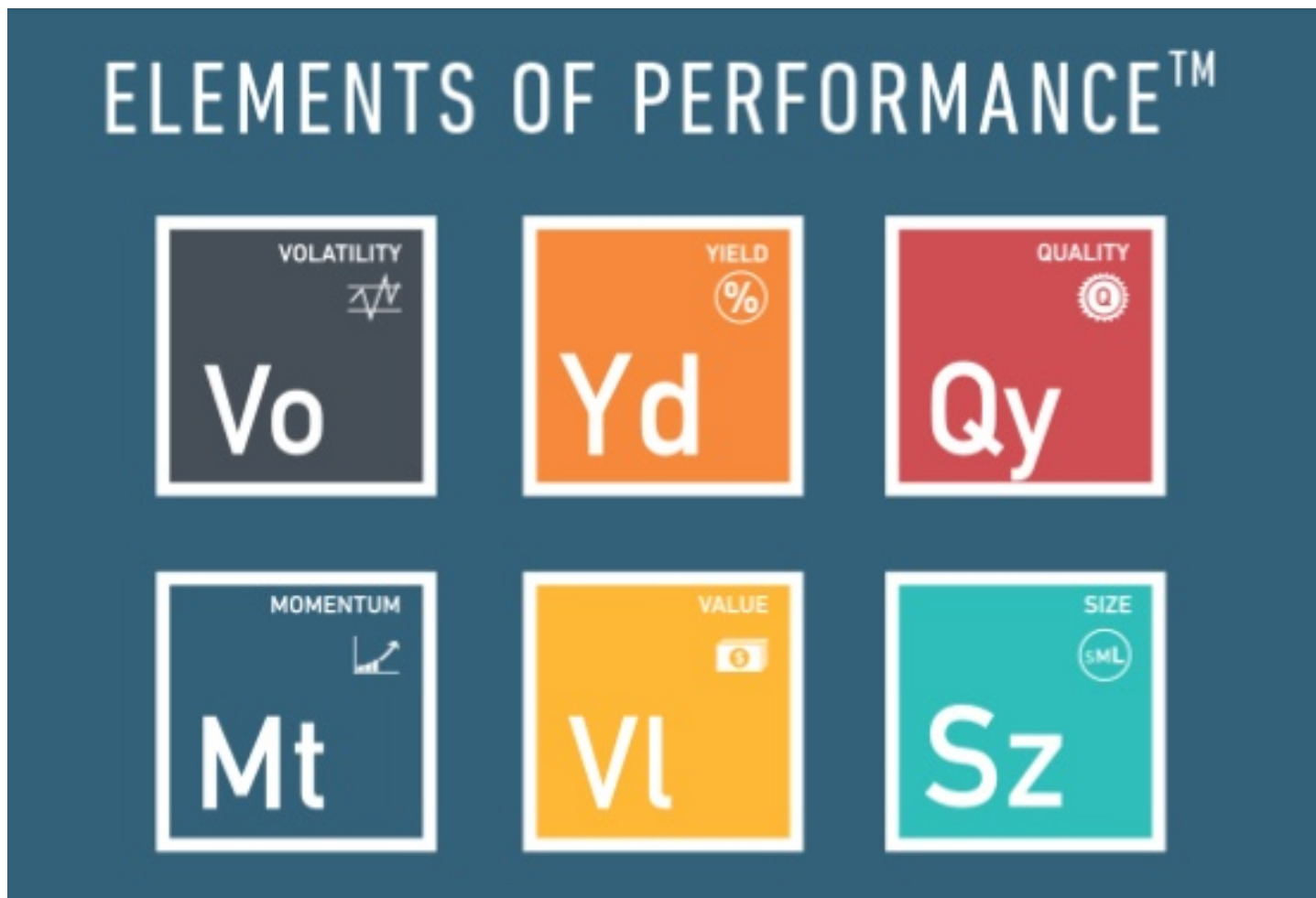


知乎

首发于
川流不息

正确理解 Barra 的纯因子模型



石川

量化交易 话题的优秀回答者

已关注

Steven Li 等 191 人赞同了该文章

1 引言

在（风险）多因子模型中，**因子暴露 (factor exposure)** 和 **因子收益率 (factor return)** 是两个核心的概念。不清楚它们的定义将影响对多因子模型的理解。**所谓因子，就是一个可以描述股票某方面特征的因素**，比如行业因子描述了股票是否属于这个行业，P/E 因子描述股票 Price-to-Earnings ratio。因子暴露就是股票在因子所代表的特征上的取值，比如一个股票的 P/E 为 15.9，那么它对 P/E 因子的因子暴露就是 15.9。**对于一个给定的因子，按照某种权重组合所有股票便形成了一个基于该因子构建的投资组合，该投资组合的收益率就被定义为这个因子的收益率。**

对于给定的因子，如何构建因子投资组合呢？常见的做法是，将所有个股在该因子上的因子暴露在截面上标准化；之后所有股票会按照因子的业务逻辑、根据因子暴露的数值从好到坏排列；最后假设做多前 10% 或者 20% 的股票，做空后 10% 或者 20% 的股票，以此来构建一个零额投资组合，它就是该因子的投资组合。这个做法在业界非常流行，但它也存在两个缺陷：

知乎

首发于
川流不息

第一个问题造成在时间序列上评价一个因子的作用时，每期的投资组合对该因子的暴露程度都不一样。举个例子，假如相邻两个月中，某因子的投资组合对该因子的暴露为 1 和 2，而相应的因子收益率为 1% 和 2%。如果我们忽视了因子暴露程度不同的话，就会得到本期因子收益率较前一期的数值有所提高的错误结论。

第二个问题更为严重。风险因子大多来源于股票的基本面数据，很多因子之间存在一定的线性相关性。**为了正确的评价一个风险因子是否有效以及在什么程度上有效，必须保证围绕该因子来构建的投资组合可以最大程度的剥离因子之间的相关性。**换句话说，针对某因子构建的投资组合应该避免在其他因子上有任何暴露。

为此，Barra（如今已被 MSCI 收购了）提出了**纯因子模型（pure factor model）**，它能够保证在截面上构建因子投资组合时，每个因子的投资组合对目标因子有 1 个单位的暴露，而对其他因子的暴露为 0*。

*** 严谨的说，根据因子的性质不同（即国家因子、行业因子、风格因子），因子的投资组合在其他因子上是否完全为 0 单位暴露略有差异（下文会具体说明）。但这不影响我们从广义上说“Barra 的模型中，因子的投资组合对目标因子有 1 个单位的暴露，对其他因子没有暴露”。这就是纯因子模型中“纯”字的含义。**

纯因子投资组合是为了正确量化因子的收益和风险而从纯数学的角度构建的。建立时没有考虑任何可投资性的要求，因此**纯因子投资组合的可投资性非常低**。它满足对目标因子有 1 个单位的暴露，对其他因子没有暴露，因此可以正确的衡量因子的有效性。

可投资性是指投资组合中股票的（多、空）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。

看到这里也许有的小伙伴会说“没有可投资性那有什么用？”。正确的解答是，**Barra 的风险因子模型的核心是做风险分析**。具体来说有两个目的：

1. **计算个股收益率之间的相关系数。**市场中个股的数量是非常多的，如果使用个股自身的收益率序列求相关系数，那么则要求收益率序列的时序长度不低于个股的数量，否则收益率矩阵就不是满秩的，因此就不可逆。由于这个要求在现实中难以实现，人们就想能不能把个股的收益率分解到一些常见的因子上，然后转而通过求解因子收益率的相关系数再推导出个股收益率的相关系数。
2. **为给定的资产或者投资组合做风险归因。**对于一个资产或投资组合，我们想要弄清楚它的收益率的波动率可以由哪些因子解释。

为了上面两个目标，构建因子的投资组合时必须能够正确计算因子收益率，这就是纯因子组合价值所在。**虽然纯因子组合可投资性低，但它在风险管理和业绩归因中有着非常重要的作用。**

2 Barra 的中国因子模型 CNE5

CNE5 是 Barra 的最新一代面向中国股票市场的多因子模型。**该模型考虑了一个国家因子、多个行业因子以及多个风格因子。**假设市场中共有 N 支股票， P 个行业，以及 Q 个风格因子。在任意给定时间点，该模型使用因子暴露和个股收益率构建**截面回归 (cross-sectional regression)** 如下：

$$\begin{bmatrix} r_1 - r_f \\ r_2 - r_f \\ \vdots \\ r_N - r_f \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} f_C + \begin{bmatrix} X_1^{I_1} \\ X_2^{I_1} \\ \vdots \\ X_N^{I_1} \end{bmatrix} f_{I_1} + \cdots + \begin{bmatrix} X_1^{I_P} \\ X_2^{I_P} \\ \vdots \\ X_N^{I_P} \end{bmatrix} f_{I_P} + \begin{bmatrix} X_1^{S_1} \\ X_2^{S_1} \\ \vdots \\ X_N^{S_1} \end{bmatrix} f_{S_1} + \cdots + \begin{bmatrix} X_1^{S_Q} \\ X_2^{S_Q} \\ \vdots \\ X_N^{S_Q} \end{bmatrix} f_{S_Q} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

其中 r_n 是第 n 支股票的收益率， r_f 是无风险收益率。 $X_n^{I_p}$ 是股票 n 在行业 I_p 的暴露，如果假设一个公司只能属于一个行业，那么 $X_n^{I_p}$ 的取值为 0（代表该股票不属于这个行业）或者 1（代表该股票属于这个行业）。 $X_n^{S_q}$ 是股票 n 在风格因子 S_q 的暴露，它的取值经过了某种标准化（标准化的方法会在下文说明）。 u_n 为股票 n 的超额收益中无法被因子解释的部分，因此也被称为该股票的特异性收益。 f_C 为国家因子的因子收益率（所有股票在国家因子上的暴露都是 1）； f_{I_p} 为行业 I_p 因子的因子收益率； f_{S_q} 为风格因子 S_q 的因子收益率。

对于给定某一期截面数据（记为 T 期），在截面回归时，Barra 采用**期初**的因子暴露取值（等价于 $T-1$ 期期末的因子暴露取值）和股票在 T 期内的**收益率**进行截面回归。在 USE4 模型中，因子收益率是**日频**的，因此截面回归也应该是日频的，所以按照上述说明，在 $T-1$ 日结束后更新因子的暴露，并利用 T 日的股票收益率和因子暴露做截面回归。以下说明来自 Barra Risk Model Handbook。

... the previous steps have defined the exposures of each asset to the factors **at the beginning of every period** in the estimation window. The factor excess returns **over the period** are then obtained via a cross-sectional regression of asset excess returns on their associated factor exposures ...

上式就是 CNE5 多因子模型。**在这个模型中，国家因子的因子暴露和 P 个行业的因子暴露之间存在共线性。**具体来说，国家因子的因子暴露向量可以表达为 P 个行业因子因子暴露向量的线性组合。这会造成上式的解不唯一。为此，对行业因子的因子收益率作如下限制：

$$s_{I_1} f_{I_1} + s_{I_2} f_{I_2} + \cdots + s_{I_P} f_{I_P} = 0$$

知乎

首发于
川流不息

通过截面回归，对 CNE5 模型求解的对象是每个因子的投资组合中所有股票的配比权重。对于因子 k 和股票 n 来说，用符号 ω_{kn} 来表示。一旦得到所有的 ω_{kn} ，便可通过下式求出当期因子的收益率 f_k ：

$$f_k = \sum_{n=1}^N \omega_{kn} r_n$$

对 CNE5 模型求解其实是一个多因子回归的求解。把多个因子放在一起回归同时求解就是为了考虑不同因子之间的相关性。这可以保证根据因子权重 ω_{kn} 来构建的因子投资组合对于这个模型所涉及的所有因子都是纯因子的投资组合。对于任何一个风格因子 S_q ，上述截面回归保证了它的投资组合仅仅在这个因子上有 1 个单位的暴露，而在其他所有因子上均没有任何暴露，即

$$\begin{cases} \sum_{n=1}^N \omega_{S_q n} X_n^k = 1, S_q = k \\ \sum_{n=1}^N \omega_{S_q n} X_n^k = 0, S_q \neq k \end{cases}$$

对于国家因子和行业因子的投资组合，“纯因子”组合的解释略有不同，我们会在第 4 节中阐述。

前面说到，在使用截面回归求解时，必须对风格因子的因子暴露进行标准化（国家和行业因子的因子暴露不需要标准化）。令 s_n 表示股票 n 的流通市值权重。对风格因子的因子暴露进行标准化的初衷是这样的：按照股票的流通市值权重构建的投资组合等同于整个市场，而市场对所有的风格因子都应该是中性的。因此，按流通市值权重构建的股票投资组合在所有风格因子上的暴露必须是 0。这意味着经过标准化后的风格因子暴露 $X_n^{S_q}$ 必须满足：

$$\sum_{n=1}^N s_n X_n^{S_q} = 0, q = 1, \dots, Q$$

此外，我们还必须对风格因子的因子暴露进行标准差的标准化，即要求对每一个风格因子 S_q ， $X_n^{S_q}$ 的标准差为 1。这样便完成了对风格因子的因子暴露的标准化。

3 国家因子的本质

相比于早期的中国股票因子模型，CNE5 中的关键变化之一是加入了国家因子。（类似的，在美国市场的最新模型 USE4 中，Barra 也加入了这一因子。）那么，国家因子投资组合的本质是什么呢？

知乎

首发于
川流不息

有点绕？没关系，一步一步来。对于国家因子，所有个股的暴露都是 1，这个组合的收益率为 f_C 。而在市场组合中，个股是按流通市值权重配置的，我们用 r_M 表示市场组合的收益。那么，只有当 f_C 近似的等于 r_M 时，上面的结论才成立。下面就来看看 f_C 和 r_M 是否近似相等。

由前文所述， s_n 是股票 n 的流通市值权重。将 $\{s_n\}, n = 1, \dots, N$ 这一组权重带入到 CNE5 的因子模型中可以得到如下关系。其中左侧就是市场收益 r_M ，右侧是使用国家因子、行业因子、风格因子、以及个股特异性收益率对 r_M 的分解。

$$\begin{aligned}
 r_M &= f_C + \sum_{n=1}^N \sum_{p=1}^P s_n X_n^{I_p} f_{I_p} + \sum_{n=1}^N \sum_{q=1}^Q s_n X_n^{S_q} f_{S_q} + \sum_{n=1}^N s_n u_n \\
 &= f_C + \sum_{p=1}^P \left(\sum_{n=1}^N s_n X_n^{I_p} \right) f_{I_p} + \sum_{q=1}^Q \left(\sum_{n=1}^N s_n X_n^{S_q} \right) f_{S_q} + \sum_{n=1}^N s_n u_n \\
 &= f_C + \sum_{p=1}^P s_{I_p} f_{I_p} + \sum_{q=1}^Q 0 \times f_{S_q} + \sum_{n=1}^N s_n u_n \\
 &= f_C + 0 + 0 + \sum_{n=1}^N s_n u_n \\
 &= f_C + \sum_{n=1}^N s_n u_n \\
 &\approx f_C
 \end{aligned}$$

上式中最后一项是所有股票特异性收益的和，由于它的值非常小（接近 0），因此在推导的最后一步被忽略了。推导中的核心在于倒数第三步中的中间两项如何变为 0。对于第一个 0，它用到了行业因子收益率按行业市值加权重为 0 以排除行业和国家因子之间的共线性这个约束条件。对于第二个 0，它是根据风格因子是使用流通市值权重来标准化这个定义来的。**由此可见，在 CNE5 模型的定义下， f_C 这个国家因子收益率确实近似的代表了市场组合的收益率，因此国家因子的组合就（近似地）是市场组合。在新版多因子模型中增加这一项是非常必要的。**

事实上，在对 CNE5 进行截面回归求解后可以发现，国家因子的投资组合中，个股 n 的权重 ω_{Cn} 非常接近它的流通市值权重 s_n 。



知乎

首发于
川流不息

通过上面的介绍我们已经知道，Barra 的截面回归模型针对国家因子、行业因子以及风格因子分别构建了纯因子投资组合。那么，这些组合有着怎样的性质呢？

4.1 国家纯因子投资组合

由 $f_C \approx r_M$ 可知，国家纯因子投资组合就是近似的市场组合，它是**纯多头**组合：

- **国家纯因子是满额投资的 (fully invested)**。国家纯因子的投资组合中所有股票（近似）按流通市值取权重，因此全部大于 0，即均为做多，不存在做空任何个股的情况。该投资组合使用了 100% 的资金。
- **国家纯因子投资组合对行业的暴露不为 0**。由定义可知，该投资组合在行业 I_p 的因子暴露为：
$$\sum_{n=1}^N s_n X_n^{I_p} = s_{I_p}$$
- 由于每个行业都包括一些股票（即对任何一个行业 I_p ，总有一些股票满足 $X_n^{I_p} = 1$ ），且股票的权重 $s_n > 0$ ，因此上式大于 0。事实上，国家纯因子投资组合按照行业的市值权重暴露于不同的行业之中。
- **国家纯投资组合在所有风格因子上的暴露均为 0**。

4.2 行业纯因子投资组合

行业因子的纯因子投资组合是一个**多空**组合，它满足以下特征：

- **行业纯因子投资组合是零额投资 (dollar-neutral)**。在这个投资组合中，我们做空一部分股票，然后用卖出股票的钱来做多另外一部分股票，因此整体来看我们的绝对投资额度为 0。
- **行业纯因子投资组合的本质是 100%做多该行业，并 100%做空国家纯因子组合（市场组合）**。由于国家纯因子组合对所有行业都有暴露，因此行业纯因子对自身行业有正的暴露，对其他所有行业有负的暴露。行业纯因子投资组合是 100% 做多该行业 100% 做空市场，**因此从业务上解释，这个组合就是认为该行业可以跑赢市场，该组合对应的就是该行业相对于市场的超额收益**。
- **行业纯因子投资组合对所有风格因子的暴露为 0**。该投资组合赚取的仅仅是行业相对市场的超额收益，这个超额收益不来自对任何风格因子的风险暴露（因为该组合对任何风格因子的风险暴露为 0）。

4.3 风格纯因子投资组合

风格因子的纯因子投资组合同样是一个**多空**组合，它满足下列特征：

- **风格纯因子投资组合是零额投资 (dollar-neutral)**。在这个投资组合中，我们做空一部分



知乎

首发于
川流不息

- **风格纯因子投资组合对自身风格因子外的其他所有因子、包括国家因子、行业因子和其他风格因子，的暴露都是 0。**从业务上解释，该投资组合是靠仅仅暴露于该因子来赚取这个风险因子的超额收益。

暂时抛开纯因子组合的可投资性，上面的结果对于因子投资有着非常重要的意义。它说明，如果我们从整体上看好市场，那么只需要持有国家因子的纯因子组合（即近似的市场组合）；如果我们看好了某些行业，那么只需要持有那些特定行业的行业纯因子组合，从而赚取行业相对于市场的超额收益；如果我们看好了某个风格因子（比如小市值、价值等），那么只需要持有这些因子的纯因子组合，去赚取通过暴露于这些因子的超额收益。

这就是 Barra 这个模型的最大意义 —— 它可以针对我们喜欢的因子（无论是市场、行业或是风格），构建出纯粹的仅仅针对于那些因子的投资组合，从而捕捉这些因子的风险收益。

5 如何理解风险因子收益

经过了上一节的解释，我们更愿意把因子收益理解为一篮子股票（即围绕该因子构建的投资组合）的共性收益（系统性风险溢价）。该投资组合如果赚钱，那么靠的是该投资组合在该风险因子上的单位暴露，靠的是该风险因子在时间维度上所带来的有效而稳定的风险溢价。从风险暴露的角度来说，所有股票 —— 无论被做多还是被做空 —— 在该组合中的贡献都是“一样的”，都是必不可少的。这和精选个股（通过深度研究或者靠其他歪门邪道）来赚取个股的特异性收益 u_n 是完全不同的，因为纯因子投资组合中股票的特异型收益几乎为 0。

当然，在现实中，我们不得不面对“可投资性”的问题。当我们无法按照纯因子投资组合的权重来实际构建投资组合（比如无法做空一些股票）时，我们就面临两个问题：

1. 实际的投资组合已经不再是纯因子投资组合；它无法保证纯因子投资组合在因子暴露上的特点。
2. 对个股的特异性收益和其风险无法做到充分分散，所以就不能忽视股票特异性收益自身的风险。

以上两点说明，在种种限制下，如果构建的投资组合的权重和纯因子投资组合理论权重不一致时，该投资组合便没有有效的暴露在该因子之下，也没有对其他因子隔离。此外，该组合又引入了无法忽略的特异性收益的风险。

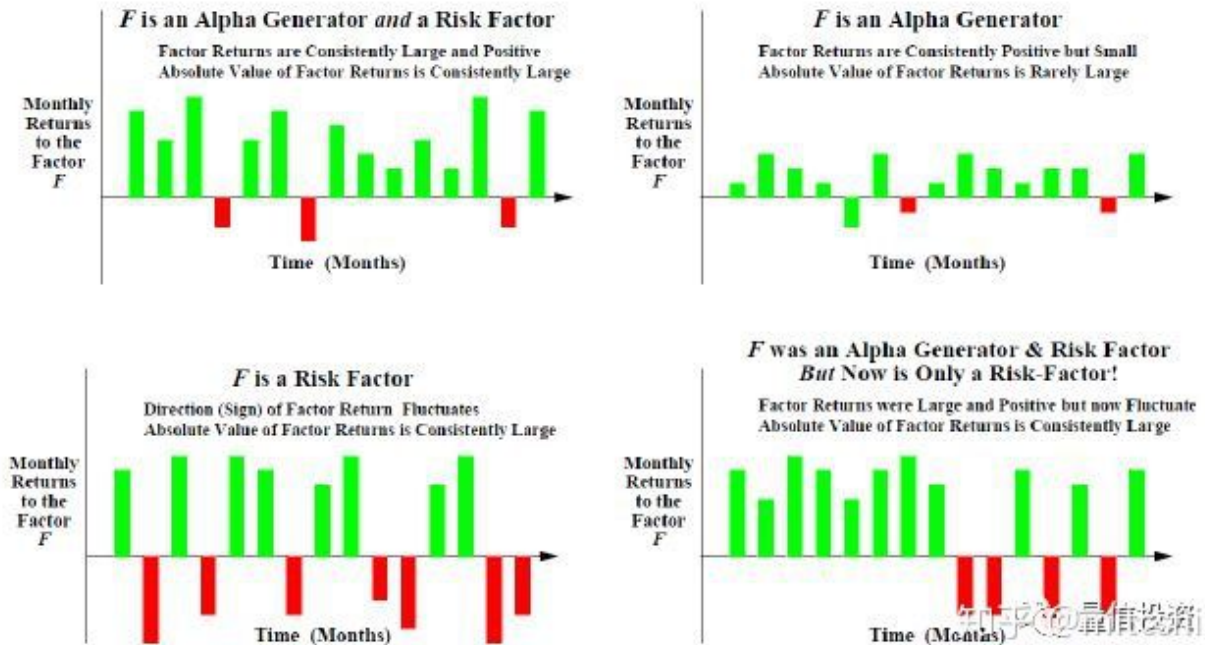
美国的 AQR 基金写过一篇文章来分析巴菲特的选股能力 (Frazzini et al. 2013)。结果显示，巴菲特选股的收益率几乎可以完全被 1 个市场因子和 5 个风格因子的收益率来解释。它说明巴菲特的投资组合能赚钱是因为它以一定的权重有效的暴露在了这 6 个因子之中，长期稳定地赚取了这 6 个因子的风险溢价。巴菲特有一个科学的价值投资框架来保证它的投资组合对最合理的风险因子有着最合理的风暴露，这些风险因子的风险溢价为带来了年复一年的优秀收益。

6 不是所有的风险因子都能带来超额收益

知乎

首发于
川流不息

超额收益，我们必须在时序上对该因子的收益率做统计分析。



上图展示了 4 种典型的因子收益率在时间维度上的统计特征：

1. 在左上角的第一幅图中，因子收益率在大部分时间为正，但波动较大。这说明该因子虽然可以贡献超额收益，但是其自身波动也带来了它对应的系统性风险。
2. 在右上角的第二幅图中，因子收益率在大部分时间为正，且波动很小。这说明该因子不但可以稳定的贡献超额收益，其自身的系统风险也非常低。**这在理论上是最优秀的收益因子。**
3. 在左下角的第三幅图中，因子收益率时正时负，波动很大，在统计上无法贡献非0的超额收益。因此，该因子无法带来超额收益，但是它可以显著的描述某种系统性风险。**因此这个因子是一个优秀的风险因子，但它不是收益因子。**
4. 在右下角的第四幅图中，因子收益率在过去显著为正，可以贡献稳定的超额收益，但是在最近不再有效，转变为纯粹的风险因子，无法贡献超额收益，仅能产生系统性风险。

在评价一个风险因子时，应按照正确的方法得到每个时间截面的纯因子投资组合，进而算出每一期的因子收益率。然后，通过对因子收益率的时间序列进行统计分析，最终判定该因子能否在长期稳定的贡献超额收益。同时，对因子收益率的统计分析也可以得到因子收益率之间的协方差矩阵，它是推导个股之间的协方差矩阵的必要条件之一。

参考文献

- Frazzini, A., D. Kabiller, and L. H. Pedersen (2013). Buffett's alpha. Working paper 1967, National Bureau of Economic Research.
- Barra Risk Model Handbook (2007). MSCI.


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_领先的原创内容监测、保护及快速授权平台](#)) 为进行维权行动。

编辑于 2019-07-02

量化交易 BARRA模型 多因子模型

▲ 赞同 191 ▼ 55 条评论 分享 ★ 收藏 ...

文章被以下专栏收录

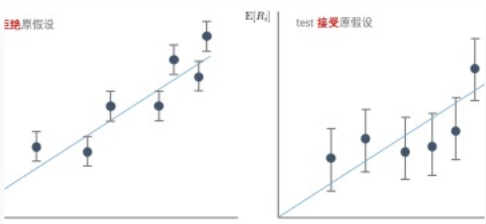


川流不息

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



股票多因子模型的回归检验

石川



Alpha调仓要如何倒腾

SuSu ... 发表于数据为剑、...





量化交易)

黄复得

55 条评论 切换为时间排序

写下你的评论...





 策商降为零

不好意思，请问引言部分第二段之后的两个点中，第一点是不是遗漏了一些内容？

1 年前

赞





哦，是的，编辑的时候遗漏了个“1”。以更正。感谢指出！

👍 赞



li lu

1 年前

请问一下

“按照股票的流通市值权重构建的投资组合等同于整个市场，而市场对所有的风格因子都应该是中性的。因此，按流通市值权重构建的股票投资组合在所有风格因子上的暴露必须是 0。这意味着经过标准化后的风格因子暴露 $X_n^{S_q}$ 必须满足：

$$\sum_{n=1}^N s_n X_n^{S_q} = 0, \quad q = 1, \dots, Q$$

此外，我们还必须对风格因子的因子暴露进行标准差的标准化，即要求对每一个风格因子 S_q , $X_n^{S_q}$ 的标准差为 1。

国泰君安报告"基于组合权重优化的风格中性多因子选股策略-五十七"中提到“在 BARRA 的风险模型中，采用根号市值加权的标准正态法，在一定程度上剔除了市值的影响，但是这样的方法会造成标准化因子截面的均值不等于 0，这在之后的股票组合权重优化的风险敞口设置时会产生一定的偏差，考虑到这点，我们仅采用简单的标准正态

$$d_{ij} = \frac{d_{ij}^{raw} - \mu_i}{\sigma_i}$$

其中， d_{ij} 为标准化因子序列， d_{ij}^{raw} 为原始因子序列， μ_i 为 d_{ij}^{raw} 的算术平均值， σ_i 为 d_{ij}^{raw} 的标准差。

和您讲是否矛盾？ 他的说法是否有误？ barra文档上的因子均值似乎是市值加权平均的

Unlike country and industry factors, which are assigned exposures of either 0 or 1, style factor exposures are continuously distributed. To facilitate comparison across style factors, they are standardized to have a mean of 0 and a standard deviation of 1. In other words, if d_{ij}^{raw} is the raw value of stock i for descriptor j , then the standardized descriptor value is given by

$$d_{ij} = \frac{d_{ij}^{raw} - \mu_j}{\sigma_j} \tag{3.2}$$

where μ_j is the cap-weighted mean of the descriptor (within the estimation universe), and σ_j is the equal-weighted standard deviation. We adopt the convention of standardizing using the cap-weighted mean so that a self-invested cap-weighted global portfolio, such as MSCI ACWI IMI, has approximately zero exposure to all style factors. For the standard deviation, however, we use the equal-weighted mean to prevent large-cap stocks from having an undue influence on the overall scale of the exposures.

非常感谢

👍 赞



石川 (作者) 回复 li lu

1 年前

我不太清楚国泰君安那句对 Barra 的评价“采用根号市值加权的标准正态法”是什么意思。你最后一张图贴的 Barra 的说明很清楚，它们从业务逻辑出发认为市场组合就应该是风格因子中性的；而市场组合中，股票是按照市值加权的，所以在标准化的时候用的

市值权重做去均值化（注意，用的市值，也不是根号市值啊。。。）。所以，在我的介绍中，只是陈述 Barra 的做法。在这个定义下，因子在截面上的均值自然也是零。在

知乎

首发于
川流不息

加精问题只限提问者和答主回答。

1



onedayyo

1 年前

因子收益算出来的数值应该是那种呢？万一是2.3这种，又是什么意义呢？

赞



石川 (作者) 回复 onedayyo

1 年前

因子收益是股票收益率按照回归出来权重的线性组合，因此数值就应该是一般收益率的取值范围。不知道您问题中的 2.3 是什么意思呢？

赞



Terry腾飞

1 年前

您好，请问一下我用某个行业指数的daily return 减去中证全指的daily return。然后 + 1) .cumprod () 做出来的曲线是不是应该跟barra模型做出来的某个行业因子收益率曲线是一样的？另外，因子用BP EP 而不是用PB PE的原因是什么呢？非常感谢

赞



石川 (作者) 回复 Terry腾飞

1 年前

您好，对于你的第一个问题，应该是接近的。但坦白说我真没这么验证过。可以用回归分析出来的行业因子组合中个股的权重和它们在行业内的市值权重比较一下，应该非常接近。第二个问题就更好了，Barra CNE5 中用的是 BP，券商报告中也是如此，所以我之前也没仔细思考过这个问题。现在我是这么想的，供你参考：我们找因子，看中的是它的对于股票收益率的 explanatory power，而且多因子模型是个线性模型。那么，一个因子 X，如果它的解释力度不错，就能尽可能的解释截面收益率的差别。那么对它取倒数得到 $1/X$ 就引入了新的误差，其线性解释力度肯定不如 X 本身。从这个角度出发，猜想 BP 应该比 PB 更好的解释了个股截面收益率的差别。等我回头做些实证，有了发现，再和你分享。谢谢。

赞



Terry腾飞 回复 石川 (作者)

1 年前

好的 谢谢回复：) 也很期待您实证后的结果

赞

[查看全部 10 条回复](#)

shiqiang

1 年前

才看到回归的时候收益率和因子暴露是“同期”，barra的文档中只说到了做“截面回归”，之前看的时候认为是“同期”。不过后来看东方财富的研报，发现他们应该是用的“滞后一期”

知乎

首发于
川流不息

知乎 J 主即截图展示，然给出了个问的似里(似里可能为0)，开又有以又还 1 个页。

所以现在更加不确定应该用“同期”，还是“滞后一期”了.....

不过没有关系，可以试着算一下，看看解释度啊，因子收益啊之类的就可以大致感觉一下了。

👍 赞



shiqiang

1 年前

刚刚看了一下barra的handbook, 应该还是有“滞后”一期的，原文如下：

Exposures:

Barra Updates the security exposures of most equity models on a monthly basis, using the last trading day's information to compute exposures for the coming month.

Factor Returns:

Factor Returns are pure measures of the factor's actual performance net of any other effects. Since factor returns are not readily observable, we must estimate them. Recall that asset exposures are computed at the end of each month. Using our multiple-factor model framework and the observed asset returns over the next month, we can estimate factor returns over the month. This is done with a cross-sectional regression of asset returns over the month on the exposures of the assets to the factors.

👍 1



石川 (作者) 回复 shiqiang

1 年前

感谢！我确认一下。

👍 赞



石川 (作者) 回复 shiqiang

1 年前

嗯，看了 Barra Risk Model Handbook，里面有 “The previous steps have defined the exposures of each asset to the factors at the beginning of every period in the estimation window. The factor excess returns over the period are then obtained via a cross-sectional regression of asset excess returns on their associated factor exposures”。所以对于 T 期，应该是用期初的因子暴露取值和 T 期内的收益率做回归。说“同期”是不妥当的。我去文中改一下。

另外，在 USE4 模型里，Barra 已经开始每天回归因子模型了：“The first step is to compute the factor correlation matrix from the set of daily factor returns.” 所

知乎

首发于
川流不息

再次感谢。

👍 2

查看全部 6 条回复



Beef Stew

1 年前

有两个问题请教您一下。因为国家因子和行业因子存在共线性，所以加了一个限制条件以求出唯一解，但这个限制条件会导致某些行业因子收益为负，如果在所有行业前景看好的情况下，这样的结果是否合理呢？还有一个问题，理论上讲，行业纯因子组合和国家纯因子组合也是可以通过求逆矩阵将其他所有因子暴露（国家，行业，风格）变为0的吧？

👍 1



石川 (作者)

1 年前

行业因子组合是做多行业同时做空国家因子（市场），所以算出来的是行业相对市场的超额收益，因此有正有负。行业和国家组合在风格因子上的暴露为零。国家因子在所有行业上都有暴露；行业组合因为做多行业、做空国家（而国家在所有行业都有暴露），所以行业纯因子组合在其他行业上也有暴露（很低）。具体的求解例子请参看 [Barra 因子模型截面回归求解](#) 这里面给出了一个求解实例，并解释了求出的因子收益率。

👍 1



小赵同学

11 个月前

您好，请问下为什么回归求解后就对于风格因子就可以得到只有自身暴露为1其他暴露为0的纯因子组合，而对于国家行业因子达不到这种性质，是因为求解的约束吗？这块看了好多研报感觉都写的很模糊，您解释的部分可以理解，但是从数学角度上来讲为什么国家行业因子无法有纯因子的性质呀

👍 赞



石川 (作者) 回复 小赵同学

11 个月前

您好。我从数学角度思考如下，供您参考。假设最简单的情况（因子没有共线性问题），则纯因子组合的权重矩阵 $\Omega = (X'X)^{-1}X'$ ，用 Ω 乘以 X 就得到一个矩阵，它的每一行都是对应的因子在所有因子上（列上的取值）的暴露。显然，在前面这种 Ω 的情况下， Ω 乘以 X 等于单位矩阵 I ，说明所有的因子组合都是纯因子。下面考虑复杂的情况，即国家和行业因子有共线性，因此考虑约束矩阵 R （无需考虑权重矩阵 V ）。这时， $\Omega = R(R'X'XR)^{-1}R'X'$ 。同样，用 Ω 乘以 X ，就得到每个因子的纯因子组合在所有因子上的暴露。显然，对于风格因子，上述新的 Ω 乘以 X 和最开始那种最简单的情况是一样的（可以考虑 X 的一个子矩阵代表风格因子）。而对于国家因子和行业因子，它们的纯因子组合在其他因子上的暴露显然受到共线性，即约束矩阵 R 的影响。下面这篇文章中有一个例子，可以进一步参考：[Barra 因子模型截面回归求解](#)。

知乎

首发于
川流不息

颜瑞

11 个月前

大佬您好，第二部分有一点没看懂：因子收益 f 为什么要用 ω 乘 r 再求和来算呢？不是横截面回归就直接得到 f 了吗？并且 ω 在回归中也算不出来吧。。

👍 赞



Running 回复 颜瑞

10 个月前

这一部分我也没理解，回归得到的结果应该只有 f_i ，没有 ω_i, n 才对，请问作者能解释一下嘛石川

👍 赞



石川 (作者) 回复 颜瑞

10 个月前

建议看一下 OLS 的数学求解。假设模型是 $y = Xf + u$ ，OLS 可以得到 $f = (X'X)^{-1}X'y$ ，上式右侧把 y 去掉其实就是 ω ，即 $\omega = (X'X)^{-1}X'$ 。这是最简单的 OLS 的情况，barra 的回归涉及了约束条件和系数矩阵，比最简单的 OLS 复杂一点。关于求解，在 [Barra 因子模型截面回归求解](#) 这篇文章里有更详尽的说明。谢谢。

👍 2

展开其他 1 条回复



彭程

11 个月前

特异性收益率怎么算呢？多元回归预测值减实际值吗

👍 赞



石川 (作者)

11 个月前

嗯，用超额收益减去因子能解释的部分。

👍 赞



石三石

10 个月前

请问用多元回归的话，是不是就相当于剔除了自变量即因子间的共线性问题？

👍 赞



石川 (作者) 回复 石三石

10 个月前

请看 [为什么要进行因子正交化处理？](#)。

👍 赞



石川 (作者) 回复 石三石

10 个月前

不过在金融里面，包括因子选股，最核心的是 prediction 是否准确，而不是 parameter estimate 是否准确。所以即便一些因子之间高度相关，这并不意味着它们的预测能力会下降。只有当回归的残差与因子高度相关时，才会影响因子收益的估计。一个非常不错的解释



知乎

首发于
川流不息

estimate 有缺陷下降。至于为什么因子之间怕大抵相同的时候会造成 parameter

estimate 不准，请看那篇为什么要进行因子正交化处理。举个最极端的情况，有一个因子完全是其他因子的线性组合，那么因子载荷矩阵 X 就不是列满秩的，这种情况下 X 都不可逆，OLS 的解也就不存在，所以多元回归也没有解决共线性的问题。

👍 2

展开其他 2 条回复



xuyao

7 个月前

如果不加入截距项而采用全行业dummy variable会有什么区别呢？

👍 赞



我是真彩笔

5 个月前

石老师您好，我想请问下为什么求解对象是每个因子投资组合中所有股票的权重，为什么不是通过WLS求得因子收益率呢

👍 赞



石川 (作者) 回复 我是真彩笔

5 个月前

求解目标是收益率。得到纯因子组合权重是中间过程。且纯因子组合本身也是有益的信息。

👍 赞



xfcy

5 个月前

在文中“上述截面回归保证了它的投资组合仅仅在这个因子上有 1 个单位的暴露，而在其他所有因子上均没有任何暴露”。

假如有3000只股票，而只有100个因子，则计算配比权重 ω_{kn} ，有100个约束，此时解不唯一，我该如何求得配比权重？加入正则项么？

👍 赞



见习WindControl

4 个月前

为什么“通过截面回归，对 CNE5 模型求解的对象是每个因子的投资组合中所有股票的配比权重。”？求解的不是线性回归系数 f 吗？

👍 赞



石川 (作者) 回复 见习WindControl

4 个月前

您说的对，求解的是纯因子组合的收益率，只不过在这个过程中先求出了每个因子的纯因子组合中股票的权重，这往往更重要（而有了权重矩阵再乘以股票收益率向量就是因

子收益率），所以行文就那么说了.....

👍 2

知乎

首发于
川流不息

感谢耐心回复。看了下其实前面已经有回复问到了，我应该先看一下的，不好意思。



👍 赞



PCMondrian

4 个月前

你好，请问风格纯因子组合是做多因子值大的，做空因子值小的吗？因子收益率是正的，说明因子值大的更好吗？

👍 赞



石川 (作者) 回复 PCMondrian

4 个月前

可以这么理解

👍 赞



Niko

3 个月前

您好，请问这一步是怎么得出的： $\sum_{n=1}^N s_n X_n^{l_p} = s_{l_p}$ ？也就是国家因子收益率近似于市场收益率的推导过程中的第一个sum

👍 赞



石川 (作者) 回复 Niko

3 个月前

考虑简化的情况， $X_n^{l_p}$ 取值非零即一，代表股票 n 是否属于行业 l_p 。 s_n 是股票 n 的市值权重。所以上式中左侧 sum 得到的就是属于行业 l_p 的所有股票的市值权重之和，也就是行业 l_p 的市值权重，即 S_{l_p} 。

👍 赞



Niko 回复 石川 (作者)

3 个月前

昨晚睡觉之前突然想出来了哈哈哈 还是谢谢您的解答！

👍 赞



不跑偏的阿屎

3 个月前

您好，想问那个关于行业因子暴露和国家因子暴露线性相关的问题，那么是因子暴露线性相关的话，为什么要对因子收益率进行限制呢？希望可以得到回答，谢谢！

👍 赞 ↩ 回复 🗑 踩 🚩 举报



石川 (作者) 回复 不跑偏的阿屎

3 个月前

求解的是 $r = Xf + u$ 里面的因子收益率 f 。因为 X 里面的 $1 + P$ (P 是行业数) 存在共线性，所以 f 求解不为唯一，因此对 f 加一个 constraint，这就是为什么对收益率 f 加限制



知乎



首发于
川流不息

证明，可以参看。

👍 赞



不跑偏的阿屎 回复 石川 (作者)

3 个月前

好的，谢谢您的回复！

👍 赞

