

知

写文章

登录



【多因子模型】 Barra模型讲解 (2)



感到不爽 · 5 天前

接着第一部分继续，第一部分地址：[【多因子模型】 Barra模型讲解 \(1\) - 知乎专栏](#)

第一部分主要说的是多因子模型的框架，这一部分主要谈谈多因子模型具体实施的步骤

一、因子有效性的验证方法

1. IC（信息系数）方法

IC的最初始的定义可以写成： $IC = \text{corr}(f, R)$ ，表示的是因子本期预计收益率与下一期真实收益率之

间的相关系数，但是本期的收益率还要通过线性回归来进行估计，所以通常情况下，为了简化我们的估计步骤，我们定义 $IC = \text{corr}(X, R)$ ， X 表示的是当期因子值， R 表示下一期的对应的收益率。一般情况下， IC 的计算是日频的，利用市场上三千多只股票的因子暴露与第二天该股票的收益率进行计算。

但是这里需要说明的是，由于因子值和因子收益率这两个序列并不服从一定的分布函数（如正态分布、T分布等），所以使用平时计算相关性的公式所以这里采取了非参数的方法，所以这里计算因子的 IC 值，采取的spearman秩和相关性，具体的包在Python的scipy.stats包中的spearmanr函数。一般来说， IC 值大于0.03，就认为这个因子的有效性比较高了。

2. Barra方法

Barra提供了一系列评价因子是否有效的方法：

- 1.单因子回归方程系数T检验的绝对值均值，通常该值大于2认为是比较理想的结果
- 2.单因子回归方程系数T检验绝对值均值大于2的占比，用于解释因子显著性的分布特征
- 3.利用回归出的因子收益率，计算因子的年化收益率
- 4.计算因子的年化收益率波动率
- 5.计算因子收益率/因子波动率
- 6.计算因子收益率与基准收益率（如沪深300）的相关性，相关性越低越好
- 7.因子的自稳定系数

其中 v_n^t 表示股票 n 在第 t 天的WLS回归所采用的加权重，即第一部分说的市值或者是残差。通常情况下，因子自稳定系数大于0.9算是比较不错的，小于0.8则认为因子并不具有一定的稳定性。

3. 因子排序法

将因子排序，分成五档或者十档，看看每一档的走势。不过这种算法有些问题，因为每一档因子在其他因子上的暴露是不确定的。比如说，将市值因子分成十档，看每一档上的股票的平均走

势，这种方法虽然控制了市值因子的暴露，但是对于其他因子，如行业，却没有控制暴露，所以这种方法要慎用。

4. 控制因子暴露的因子排序法

原理同3，只不过增加了约束条件，控制了其他因子在评测因子各个档上的暴露。但是，个人并不觉得这个方法用起来方便，因为通常情况下，多因子模型要进行全市场选股的，全市场3000多只股票，解一个三千多个变量的优化方程，其鲁棒性还是存在问题的。

二、因子取舍与因子合成

通过第一部分说的三种方法，就可以得到挑选出来一些因子了，接下来就是对于因子的筛选工作了。

在计量经济学中，多元线性回归要考虑到多重共线性问题；所以，对于相关性较高的因子，可以采用两种方法进行处理：

- (1) 根据因子的有效性进行排序，挑选最有效的因子进行保留，其余的就直接删掉了
- (2) 对因子集合进行合成，尽可能多的保留更多有效因子信息

因子合成的主要方法：

- (1) 等权法（相关性高的因子等权重进行合成，合成一个复合的因子）
- (2) 历史收益率加权法

利用因子一段时期的历史收益率作为权重进行加权，合成一个复合因子

- (3) 历史信息比例（IR）法

利用一段时间因子的信息比例（即因子收益率均值/因子收益率波动率）作为权重进行加权（研报中比较推荐这种方法）

- (4) PCA（主成分分析）

PCA是一种线性降维的方法，但是这种方法在金融中慎用，因为PCA会破坏因子之间的关系。

三、通过多因子模型进行股票选择

基本上讲到这里，多因子框架基本的逻辑就很清楚了，再回顾一下：

- step1: 因子选择，去极值（三种方法：Zscore标准化、分位数标准化、MAD标准化）
- step2: 中性化（行业中性化、市值中性化）
- step3: 因子有效性检验（IC法、BARRA法）
- step4: 相关性强的因子合成新的复合因子（等权法、历史收益率加权法、历史信息比例加权法、PCA）
- step5: 因子加权进行股票选择

对于如何进行因子加权，最经典的是参考文献6提出的IR方法。

1. IR方法

令：

其中， v 表示待估计的因子权重值，是个 $1 \times K$ 维的向量， K 是因子数量， Σ_{ic} 表示日频IC的协方差矩阵， \bar{IC} 表示IC的均值。解下列的优化问题，求 v 的值：

解得 V 之后，用 V 乘以各个因子暴露，加权求和，得到的结果就是每只股票最终的得分。将股票得分进行排序，就可以选择前30、50只股票进行投资了。

2. WLS收益率加权法

多因子模型（一）介绍了利用WLS估计因子的日收益率，将估计的因子收益率乘上因子暴露值，再加上自身的特异收益率 u ，就可以算出来每只股票的预计收益率了。这时候可以选择预计收益率前30、50的股票进行投资。

这样说有问题吗？当然是有的，因为此时，你要进行了投资了，你只能知道历史的因子日收益率，知道今天的因子暴露，可是你并不知道，今天各个因子的收益率是多少，那么怎么办？估计呗，四种方法估计因子今天的收益率。

- 历史平均法

这个不用说了，简单平均即可。

- EWMA（指数加权平均法）

本质思想是衰减的平均值法，即越靠近预测日的值所占的权重越大。

- 时间序列方法（AR、ARMA）

模型的具体推倒参阅蔡瑞胸《金融时间序列分析》。

- 其他牛逼的方法（随机森林、深度学习等）

比如说X可以是前10天的收益率，Y是下一天的收益率，这样就可以去用机器学习的方法做。

3. 各种牛逼的方法开始发挥用处了

第三种方法只是提供个人的一种思路，这个时候，是不是就可以尝试RNN及各种RNN改进版这些东西了？

其中X表示每只股票的因子暴露，Y则是每只股票的收益率，这个收益率可以是第二天的，也可以是未来几天平均的；市场上所有的股票的一段时间的因子数据合成一个大的train_X矩阵，train_Y就是对应的收益率，数据量足够大，可以用深度学习试试，不过效果不一定好。当然，也有人这么做的，上了实盘。

除了RNN，其他的XGBOOST，Extratree、Randomforest是不是也都可以一试？

四、barra框架下的多因子风险模型

这一部分，着重讲解Barra框架下的股票投资模型。其实到了第三部分，多因子框架的大体规则和流程都说的差不多了，这部分主要讲讲barra的多因子模型是怎么估计。

回到第一部分多因子最核心的公式上来： $R=Xf+u$

barra风险估计分成两个部分：第一是因子收益率的协方差矩阵，第二是每只股票的特异收益率

1. 因子收益率的协方差矩阵估计

首先，利用日频因子的历史收益率，通过EWMA（加权移动平均）计算日频率的因子收益率协方差矩阵F

其中，指数衰减权重 $\lambda = 0.5^{1/h}$ ，表示赋予越靠近当前日期越高的权重，h表示样本长度。

之后呢，虽然使用了EWMA方法，但是barra认为，这个时候协方差矩阵会存在风险的持续性高估或者低估问题，要进行调整，所以引入了因子截面偏差统计量 B_t^F 和因子波动率乘子 λ_F

定义：

K是因子个数， f_{kt} 表示第k个因子第t的收益率， σ_{kt} 表示第k个因子从开始到第t天的收益率的波动率。通常情况下， B_t^F 应当等于1，如果 B_t^F 大于1，则证明因子预测的风险过小；（国泰君安研报里面写的是过大，个人觉得是错的，barra的use里面写的就是过小，因为只有 σ_{kt} 小了， B_t^F 才应该大）。之后，定义：

这里的 λ_F 计算，也是EWMA方法， W_t 表示的是EWMA的衰减权重。

最后，就得到了我们调整之后的因子协方差矩阵：

$$\tilde{F} = F * \lambda_F^2$$

然后，我个人再说一句我的观点，国泰君安的研报后面说用Newey-West方法将日频数据调整到月频，我认为这又错了，Newey-West是调整异方差性和自相关性的方法（接下来会细说），与日频、月频数据无关，而且barra的USE4中并没有这一部分的叙述。

2. 股票特异性风险估计

股票的特异性风险估计基本算是barra里面最复杂的一个部分了。我在这仅仅就是说一说我对于barra每一部分的个人理解。

- step1: 同因子收益率协方差矩阵一样，进行EWMA；

h 是EWMA的样本长度， j 表示第 j 只股票。

- step2: Newey-West调整

定义：

C_n^{NW} 为Newey-West调整方法，后面大括号里的，就是EWMA之后的特异矩阵。

问题来了，这个Newey-West调整到底如何应用？

我对于这个问题也是百思不得其解，先贴出Newey-West方法的权威解释，来自蔡瑞胸的《金融时间序列分析》

可以看到，这里的Newey-West是针对残差的异方差性或者自相关性，调整了待估参数的协方差矩阵，显然，这与barra的Newey-West调整方法不符的。

最后，我在参考文献7中找到了一种Newey-West调整的方法，我也不知道正确与否，但是至少能够解释的通Step2的过程：

参考文献7认为

其中：

Z_t 是一列的时间序列，长度为T期，且 $E(Z_t) = u$ ，m是Newey-West调整的滞后期的长度，Bartlett权重函数 $w(j, m) = 1 - [j/(m + 1)]$ ，当 $j \leq m$ 的时候， $w(j, m)$ 的取值为0。m的取值可以按照蔡瑞胸的说法。

- step3: 贝叶斯压缩

barra给这部分命名叫做贝叶斯压缩，这部分认为，单只股票的特异性风险不仅取决于自身，还取决于与它市值类似的股票特异性风险的均值。

定义最终的矩阵公式：

其中， v_n 是加权系数， $\hat{\sigma}_n$ 表示step2计算出的股票n的协方差,然后，定义

w_n 表示的是市值加权系数；这里引入市值因子，假设将市值因子分成了十档，每一档上有若干只股票。所以这里的 S_n 也有十个，分别是每一档市值上的股票风险值 σ 的加权平均，加权系数取决于市值。

这里， q 是常数。

$\Delta_\sigma(s_n)$ 就可以理解成每一档市值上股票的方差。

- step4:引入特异截面偏差统计量 B_t^F 和特异波动率乘子 λ_F 之后的值

这部分内容与因子收益率协方差矩阵的风险调整类似，首先定义特异截面偏差统计量：

这里的 σ_{nt} 是step3中计算出的第n只股票第t天的波动率， u_{nt} 是第n只股票第t天的特异收益率， w_{nt} 是股票n第t天采用WLS回归时的加权系数。

再定义特异波动率乘子：

需要说明一点的是，这里的 w_t 与之前因子波动率乘子的 w_t 相同，都是定义一个EWMA的衰减值。

最后得到调整之后的股票特异收益率：

经过四步之后，barra的特异风险矩阵的计算也算是结束了。

写在最后的一点感悟

最后，再提出一个问题，barra的风险模型可以用来做什么？

个人认为，barra的风险模型理论上可以用来选择股票的，但是全市场三千多只股票，如果单纯的去解一个均值-方差模型这样的凸优化问题，显然会带来很大的偏差。所以barra模型，我认为可以在得到了想要的30、50只股票之后，计算应当以什么样的比例投资于这三五十只股票。



多因子模型（一）基本算作是对券商研报的总结汇总，多因子模型（二）算是我个人对于模型的一些理解了。欢迎大家拍砖。

参考文献：

1. 国泰君安，数量化专题之五十七：基于组合权重优化的风格中性多因子选股策略
2. 爱建证券，多因子系列之一：多因子模型梳理探索
3. 华泰证券，多因子系列之一：华泰多因子模型体系初探
4. 东方证券，选股因子数据的异常值处理和正态转换——《金工磨刀石系列之二》
5. Barra, USE4

- 6. Qian, Quantitative Equity Portfolio Management, modern techniques and applications
- 7. Deepa Dhume Dattay, Wenxin Du, Nonparametric HAC Estimation for Time Series Data with Missing Observations
- 8. 蔡瑞胸，《金融时间序列分析（第三版）》
- 9. 禹惊廉，基于特征修正协方差矩阵的中国A股市场多因子模型构建
- 10. 禹惊廉，股票收益率协方差矩阵的特征修正方法
- 11. Jose Menchero ,Jun Wang ,D.J. Orr, Eigen-Adjusted Covariance Matrices,Improving Risk Forecasts for Optimized Portfolios

BARRA模型

☆ 收藏  分享  举报

 45



3 条评论

写下你的评论



胡翔
噢噢，板凳吗？写的不错，通俗易懂，支持一下
4 天前



张强Jenkin
先mark一下！
4 天前



杜家昌

个人认为barra模型主要是为了去预测市场的风险或者给定的组合风险，其实对于选股意义不大，选股更注重的是收益，那么预期收益模型可能更对路，比如FF的五因子

3 天前

文章被以下专栏收录



清华大学量化投资协会成果集萃

清华量化协会关于量化交易模型、策略的学术探讨

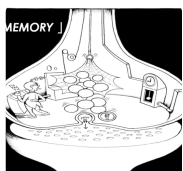
[进入专栏](#)

推荐阅读

AC米兰 16-17 赛季小结

这几天米兰夏季转会工作开展得太过迅速，以至于我感觉新赛季都快要来了。穆萨基奥已经宣布加盟，凯西也已经到了米兰体检，传闻中的罗德里格斯和莫拉塔看起来也是很有希望。米兰球迷已经情不... [查看全文](#) >

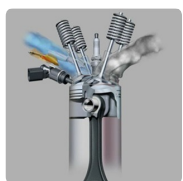
DHLLL · 6 天前 · 编辑精选 · 发表于 只是聊足球而已



神经科学的记忆| 记忆的脚步

文/ insoulter 本文已被知乎日报收录：都听过不少关于「记忆」的惊人研究，这里是几条实打实... [查看全文](#) >

insoulter · 1 个月前 · 编辑精选 · 发表于 行为与认知神经科学



可变气门正时系统简介（四）

最近的事情有点忙，所以第四篇拖了有点久了，有了上三篇的基础，我们可以来看一下中间锁止的... [查看全文](#) >

Crown · 1 个月前 · 编辑精选



靠谱配方 | 萌啾啾小鸡棉花糖

为什么要做小鸡形状的棉花糖？因为可爱呀！~~~~~



~~~~~... 宜有至义 >

焙忘录 · 13 天前 · 编辑精选