

主动投资的基本原理 . 第25讲

 渔阳 2018-07-12 20:00

字数 4,126 阅读需 11分钟

“ 欢迎来到量化小学 ”

▲ 加入“[量化小学](#)”校友圈儿提问交流

投资全球更要投资自己

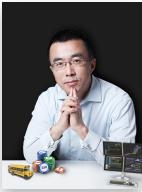
 我的订阅


0


收藏



来自特辑



量化小学
解放你的投资动手能力



最近更新

【学业总结】量化学习的脉络梳理，以及



详细内容请在wifi环境下观看视频

<本期课程4426字，视频19分57秒，请合理安排学习时间>

继续学习提高的路径

2019-04-12更新

进阶研究：集成学习和深度学习. 第31讲

2019-03-28更新

内容阅读

大家好，欢迎来到量化小学。今天我们开始新的一章：**因子研究**。

我们先简单介绍本章主要内容，然后谈谈主动投资最基本的逻辑框架。量化的主动投资是一个比较困难的事情，一般而言属于一个专业人士的范畴。

普通投资者，包括金融人士，个人投资理财时，我还是建议大家多考虑资产配置，多使用被动指数化投资的方式，也可以考虑使用智能投顾等比较新的方法。这样省时省力，效果理想。如果想做主动投资，那最好还是去买一些比较靠谱的基金，让专业人士帮你打理。

这章节内容围绕主动投资，即如何寻找能够跑赢指数的股票。这章涉及内容较多，主要是因子的研究方法。从架构上来看，我们先讲主动投资的基础原理，即主动投资的逻辑框架和发力点。然后，从易到难讲述因子的研究方法。从最简单的分组法到多元线性回归法，再到现在关注度较高的机器学习法。

机器学习法涉及多个话题，包括逻辑回归、分类、集中学习等。量化小学是一个介绍性的课程，重在介绍这些方法的宏观逻辑，具体的技术内容，会给大家提供参考资料，有兴趣的同学们可以在课后深入学习。

我们仅仅知道研究方法是远远不够的，我们还需要研究工具与样例程序。我们有quantOS，quant<OS>中有一个名为JAOS的组建，它主要用于选股和因子研究。我们前几期谈到过JAQS，它还可以用作期货的CTA与回归策略。本章，我们会多次使用JAQS。以上就是本章的主要内容。



主动投资的基本原理

主动投资的基础框架与发力点

这里再次提及度量积极投资效率的关键指标，即信息比率。它的基本公式是用投资组合超过指数的超额收益除以投资组合相对于指数的积极风险，这是风险收益比的概念。比例越高，投资效率就越好。这就说明投资者指数偏离不多，可以取得超额的收益。

那么怎样达到一个比较高的IR？这就涉及到投资组合的基本原理，英文就是Information Ratio。它看上去简单，就是把IR分成两部分，IC乘以 $\sqrt{\text{Breath}}$ 。IC,你可以将其理解为研究的精度，即主动投资。我们要对未来做一些预测。那么，预测和实际结果间的相关性，就是IC。

Breath又是什么？它是研究的一种广度。通俗来说是独立投资次数，它的基本原理和赌博原理不谋而合的，无非就是两招，一是提高单次胜率，二是多玩几次。如果单次胜率高，玩的次数又足够多，输钱的概率就会大大变小，投资的效率就会随之提升。对应到主动投资上就是两点，一是找到与实际结果相关性高的预测方法，二是预测次数尽可能多。次数多了，正确的概率就会大大提高。

“广度”提升：横截面与时间轴

横截面

1.在给定的时间点，扩大选股空间。因为有更多的股票，相当于有更多做预测的机会。从中可以选出好的股票做多，差的股票做空，所以扩大选股空间是根据主动投资的基本原理。这对提高IR有好处。



2.注意“扩张有度”，因为因子研究本质来说是在股票之间寻找共性，这样因子效果才会更好。举一个稍微极端的例子，如果把中国的股票和美国的股票放在一起研究，乍一看选股空间变大了，但实际上由于中国的股票和美国的股票共性不多，实际效果反而会变差。

这就有艺术的成分了，就是如何决定在差不多一致和足够大的空间中做出平衡。有时候因子研究，把同行业的股票放到一起研究，效果会更好。或者把市值大的和市值小的股票分别放在一起研究，这样效果也会变好。因此，横截面就是在保持共性的前提下，尽可能的扩大选股空间。

时间轴

1.从理论的角度来说，提高决策的频率往往有助于提升IR。

2.从实践的角度，因为是有交易成本的，决策频率不能太高，否则你的钱将都会用于上交手续费。最后，你不但不能赚钱，还会稳定的亏钱。实际上，你必须得在决策的频率和交易的成本之间做出平衡。这点大家多实践就会有经验。

A股市场，通常日级别到月级别是一个比较好的频率。国外市场会更高一些，原因很简单，因为国外市场的交易成本比较低。大家更易朝频率高的交易发展。

深入思考

1.首先是从策略研究难度来看：择时〉行业滚动〉选股

一般而言，择时是最难的。择时就是判断市场涨跌。这是大家平时关心的问题。但实际上它的难度最大。因为你做择时，你的判断的次数少。比如我们做一次择时，通常我们需要几天甚至一个月。假设我做月度择时，那一年中我只能做出12次判断。



根据刚才我们提到的投资基本原理，你才做12次独立判断，那么你的IR就不会特别高。次数太少了，广度不够。那选股为什么相对容易？因为你的选股空间大。你可能同时有八百到一千只股票，如果你还是每个月选一次的话，你独立判断的次数就比择时多。行业轮动是居于两者之间的，这也策略开发难度所在。

2.量化股票策略精力分配（通常）：选股〉行业滚动〉择时

策略研究时，我们通常在选股策略上多分配时间。其次，我们考虑行业轮动的模型，最后，考虑择时的模型。根据这一投资基本原理，择时的模型，IR很难做得很高，也就是它的主动投资效率较低，或许根本就不值得花这个时间。

在国外，顶级股票量化基金往往是市场中性，完全不做择时。刚才我们提到的几点原因，一是择时难做，接近有效市场；二是择时付出的精力和所得到的收益不成正比；三如果以后大家有经验了就会察觉，从风险占比的角度来看，风险额度分配到择时上也是不划算的。

“精度”提升：预测能力强的因子

根据主动投资的基本原理，广度之外我们还需一定的预测精度，即我们需要知道预测能力强的因子。在这里IC是英文Information coefficient的缩写，它是度量预测因子X和被预测变量r之间的相关性的数学量。从数学关系式，我们可以发现r跟X的关联性。

一般而言，r通常是某种收益率。它可以直接是股票的本身收益率或者是减掉指数影响后的超额收益率，也可以是残差收益率。残差收益率是去掉其他因子影响后的结果，比如行的影响，实际影响等。所以这预测的是未来的某个收益率，



那么我们如何预测它呢？这就是所谓的因子，因子是各种各样的。

因子例1:市盈率

举了例子，比如我们用股票的市盈率来做预测的因子，那市盈率越低就说明这股票越便宜。

因子例2:行业中性分组下的市盈率

那它有没有超额收益呢？我们可以放到X试一下，我们可以做一个行业中性的分组，在这种情况下研究它的市盈率。从通常逻辑讲，不同的行业，它的经济基本面不一样。比如银行股的市盈率低一些，科技股的市盈率高一些。根据我们刚才讲过的原理，你把相似性高的东西放在一起研究，效果会更好。这样的话，我们可以先做一个行业中性，再来看它的市盈率。

因子例3:市值中性分组下的市盈率

类似的，可能大股票的市盈率会相对低一点，小股票的市盈率呢会相对高一点，所以我可以做一个市值中心的分组。从这个角度研究市盈率，所以这是三个不同的因子。

我们课后的样例是分别研究因子在A股市场不同选股空间中的有效性，大家可以自行阅读，尝试把这些因子替换成其它的因子。

这就是因子研究的基本框架。总结一下，我们的目标是要寻找IC高而且稳定的预测因子。

这一点，我特别要强调一个稳定，就是IC，它在不同的历史时期也可能不一样，比如某个因子可能有一段时间管用，有一段时间又不管用，那它稳定性就不够好。



反之，穿越牛熊，穿越各种情况的因子，IC就比较稳定。所以我们寻找的目标第一是IC高，第二是比较稳定，那具体用什么样数学指标我们后面会有所涉及。

第二就是方法。从最简单的分组法道回归法，再到机器学习法等，我们在这一章中会逐步涉及。

第三是因子研究，你必须要有工具，传统上大家都是各自的去写一套工具，但现在有了排放比较好的工具，也有了开元的技术环境，所以quant<OS>的JAOS组件，它的功能就很强大。这里涉及到选股研究两个比较主要的模块，一个叫Single digger即信号研究；另外一个直接模拟历史选股和交易的回测工具。大家可以登录JAOS网站去看具体的相关文档。

因子研究的一般注意事项

回测时间要足够长

1.你需要覆盖不同的经济周期以及牛市、熊市等各种情况。如果时间不够长的话，你可能找到一些规律。比如说有一类股票的周期股，像煤炭钢铁，包括化工，在经济景气的时候，这些周期股的盈利通常都很好，但在经济不景气的时候，这些周期股的盈利就比较差。

那么，如果你想研究这些周期股的表现，你需要足够多的历史数据，最好覆盖一个完整的经济周期。所以经济周期都是以年为单位，你不能用太少的数据去测量。

2.降低“巧合”被错误认定为“因子”的概率

如果你用过少的数据进行回测，你以为找到正确的规律，但这只是一种巧合而已。有时，



我们关注财经媒体，经常会看到一些类似的说法：“这个市场已经五连跌了，那么历史上五连跌之后会……”其实，这些所谓的报道经常都是带有偏见的。

因为数据不够多，所以你总能找出一些巧合。比如说5月份市场倾向于上涨，但是你想A股从开始到现在总计不到30年，所以一共不到30个5月份，你很容易找到这样的巧合，它究竟是真正的规律还是只是一种历史数据的巧合，这就很难辨别，但是你覆盖的周期如果足够长的话，就能够避免这样的问题。

在这里我也简单地列了几个概念，一个是coincidence（巧合），另一个是correlation（相关），最后一个是causation（逻辑因果）。他们显然是由弱到强的关系。我们不感兴趣的是巧合，因为巧合，它发生完全是随机的。比如说刚才我们提到的说每周三股市倾向于上涨，或者每年5月份股市上涨概率大，像这样的规律其实巧合概率较大。

一般而言，我们关心的是相关。这就是我们先去观测一件事情，比如说PE值，市盈率偏低，然后他它能跟股票未来的收益相关，这是我们一般研究的范围。

逻辑因果，一些研究大神会采用因果方式阐述问题。实际上，金融市场通常逻辑因果关系不强。所以从研究的角度来看，我们找到相关性即可。另外，IC值通常不会特别高，我们之后也会讲到，IC超过0.2是非常少的。

类似的东西放在一起研究

这涉及到因子选股，有些因子是在全A选股效果好，就是刚才我们说的选股空间大有助于提高IR。但是也有些全A的一个子集，例如中证500指数或某个行业，效果也比较好。这是因为中证500差不多是相对而言较接近的一些股票放在一起选。某一个行业较接近的股票放在一起，效果也会更好。



有些原始因子选股效果好，有些是行业中性或者市值中性处理后效果比较好。大家可以想象到的排列组合是很多的，因此你需要一个比较高效的工具来帮助你做这些事情。

因子处理

1.预处理

在计算基本因子的时候需要做一些预处理：取对数、去查分、或者做排序等，目的是更接近正态分布或者一致分布。然后，需要去极值。

2.标准化

我们利用排序，均值和方差把它转化为z-score

3.去中性

去除行业市值其他因素的影响

在这里，我们先给大家介绍了基本框架，你知道要注意的具体事项，我们列举出来了常见的一些方法和名词。如果你有不懂的地方，比如标准化，你可以查阅相关文档。

今天，我们讲了因子研究的大致框架和一般注意事项。下次开始，我们就要细看来一些具体方法。最后，我向大家提供一些学习资料，首先还是金融终端教学频道里面的样例程序，另外就是我反复提到的《Active Portfolio Management》，这本书非常经典，大家可以自行阅读。

书中两章与今天的内容相关性较大，第一是主动投资的基本原理，第二是做预测的基本原理。书中还会提到一些数学公式了。如果你理解了今天我们的课程框架，你再去阅读这两章就会比较容易理解。特别是希望走专业道路的同学。



今天的课程就到此结束。

-END-

加入“量化小学”的见识圈，关注动态

感谢您订阅本特辑，扫描下方二维码或[点击圈子链接](#)，即可加入专属见识圈子提问交流





量化小学



渔生

小学而大不遗，量化师生联谊会

感谢大家订阅《量化小学》，这里是学校见识社群，你可以随时提问、随时互动，我们一起投资，一起分享！



风险提示及免责条款

市场有风险，投资需谨慎。本文不构成个人投资建议，也未考虑到个别用户特殊的投资目标、财务状况或需要。用户应考虑本文中的任何意见、观点或结论是否符合其特定状况。据此投资，责任自负。

写评论

请发表您的评论



发布评论

华尔街见闻

- 关于我们
- 广告投放
- 版权与商务合作
- 联系方式
- 意见反馈

声明

未经许可，任何人不得复制、转载、或以其他方式使用本网站的内容。
评论前请阅读网站[“跟帖评论自律管理承诺书”](#)

法律信息

- 版权声明
- 用户协议
- 付费内容订阅协议
- 隐私政策

违法和不良信息

举报电话: 021-60675200 (周一到周五9:30-11:30, 13:00-18:30)
举报邮箱: contact@wallstreetcn.com
网站举报: [点击这里](#)



华尔街见闻APP



华尔街见闻公众号



微博@华尔街见闻



中央网信办
违法和不良信息举报中心

上海市互联网
违法和不良信息举报信息

[违法和不良信息举报受理和处置管理办法](#)

[清朗·财经违规内容专项整治公告](#)



举报中心

友情链接

[腾讯财经](#) | [财经网](#) | [澎湃新闻](#) | [界面新闻](#) | [全景财经](#) | [陆家嘴金融网](#) | [富途牛牛](#) | [网易财经](#) | [凤凰网财经](#) | [虎嗅](#)

© 2010 - 2022 上海阿牛信息科技有限公司 版权所有 沪ICP备13019121号



沪公网安备 31010102002334 号 增值电信业务经营许可证沪B2-20180399

