

单因子测试（中）——分层测试法

原创 量化小白H 量化小白上分记 2018-12-23

收录于话题

#多因子模型

12个 >

本文总结单因子测试的分层测试法。与回归法相比，分层测试法相对繁琐，但能展示更多细节。

分层测试法的思路是在统一的规则下，利用单因子构建投资组合进行回测，然后对投资组合的表现进行全面评价，通过投资组合的表现说明因子的有效性。

组合构建方法

每月末，把当月因子进行预处理之后，将股票按因子值排序后（本篇默认降序），等分为10组（或5组），买入第一组的股票，持有到下月最后一个交易日，重复上述过程。

这里需要说明的几点是

1. 为什么一定选每月末选股，并没有什么非常好的支撑逻辑，只是目前大家都是这样做的。
2. 选取第一组的前提是因子值越大的股票越好，但有的因子可能是越小越好(比如PB，市值因子)，这类因子，应该买入最后一组的股票，但为了编程统一，一般对这类因子取倒数(比如PB，一般用 $BP = 1/PB$)。
3. 回测时，也可以同时对最后一组做空（虽然实际中A股市场并不能这样操作），这样做的目的在于保持市场中性，消除市场下跌的对于因子测试结果的影响。实际操作中，一般会选用相应的衍生品（如股指期货）对冲风险。
4. 一般会把每一组当成一个投资组合，即每期构造5个投资组合，去看五个投资组合各自的表现，**好的因子应该使各组合收益曲线明显分离，并且收益各组之间收益关系是单调的。**

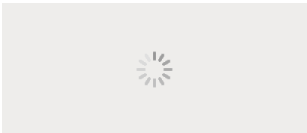
5. 细节：因子数据用到财报数据的，需要注意到实际交易日期（tradeDate）和财报日期（rptDate）的区别，rptDate数据必须是给定时间下能取到的，否则会有未来信息，最好输入能取到的最近的财报日期，比如现在是7月14日，那报告日期就应该用当年的3月31日，如果说只想用年报和半年报的数据，就应该用去年的12月31日，这样子得到的数据才是正确的。

评价指标

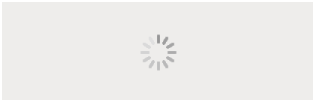
1. 组合收益率/净值

组合收益率/净值是分层测试法里评价因子有效性最重要的指标，在计算日度收益率的基础上，计算组合月度，年度的收益率，并与基准收益率相比较，可以明显看出因子的效果。

组合收益率要考虑组合的加权方式，一般采用等权或者市值加权，一些特殊的加权方式，因子测试时不会用到。不论哪种方法，都有准确计算和近似计算两种，准确计算来说，以市值加权为例，假设第t期买入n只股票，价格分别 $P_{n,t}$ ，股票股本 $M_{n,t}$ ，则t期所有股票的总价值为

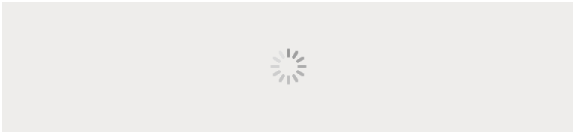


则从第t期到t+1期的组合收益率为



当然这里计算对数收益率也是可以的。

近似计算方法，每次首先计算出每只股票的收益率，再算他们的市值加权平均作为组合的收益率，用公式表示如下：



第一种方法计算时非常麻烦，因为每一期都需要对股票价格进行市值加权。如果是计算月度收益率还好，只是每月算两个加权的总价值，但如果要算逐日的收益率，这种方法就效率很低了，而且要非常精确还要考虑股票的分红配股的影响，基本不太可能算的准确，只是测试因子的话，没必要做这么准，所以一般采用的是第二种方法，而且第二种方法python可以用`np.weight`或者`np.mean`快速完成，想对简单。

个人感觉，编程实现上来说，即使是第二种方法，想快速计算投资组合的逐日收益率也不是想象中那么简单的事情，循环效率非常低。

2. IC

IC定义为当期各股票因子值与下一期各股票收益率的相关系数。

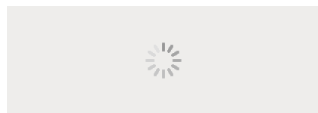
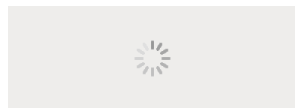
显然，IC绝对值越大，表明因子有效性越高。这里的相关系数有两种计算方法，一种是常用的**皮尔逊相关系数**，即线性相关系数。另一种是**斯皮尔曼相关系数**，也叫作秩相关系数（rankIC），这种相关系数反映的是两个变量之间单调的相关关系，对于非线性关系的变量，效果更好。

通过观察回测期间IC的分布情况，可以看出因子整体表现情况。

3. IR,ICIR

收益率和IC都是从收益角度对于因子的衡量，并没有考虑到因子的稳定性(风险性)，在收益率和IC基础上，可以计算IR,ICIR。

IR又称为信息比率。衡量某一投资组合优于一个特定指数的风险调整超额报酬。



之后将只做多第一组的组合称为多头组合，同时做多第一组和做空最后一组的组合称为多空组合。

回测区间：2010年1月1日-2018年6月30日

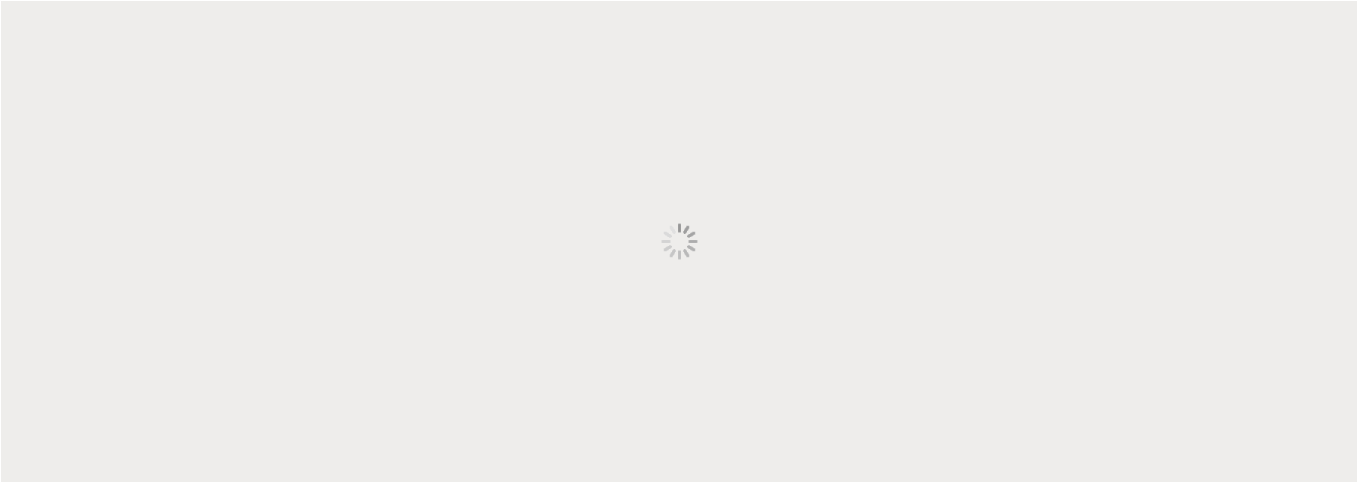
基准指数：沪深300指数

分组数：10

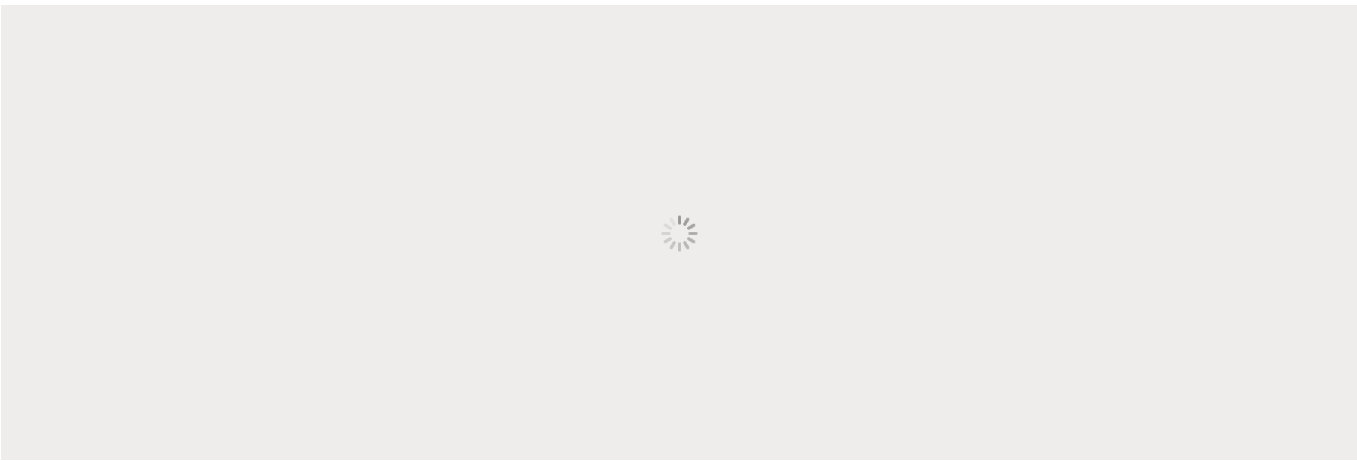
蓝色线为组合净值，**黑色线**为基准指数净值（HS300），**红色**为最大回撤率。

PB因子 (windcode : **pb_lf**)

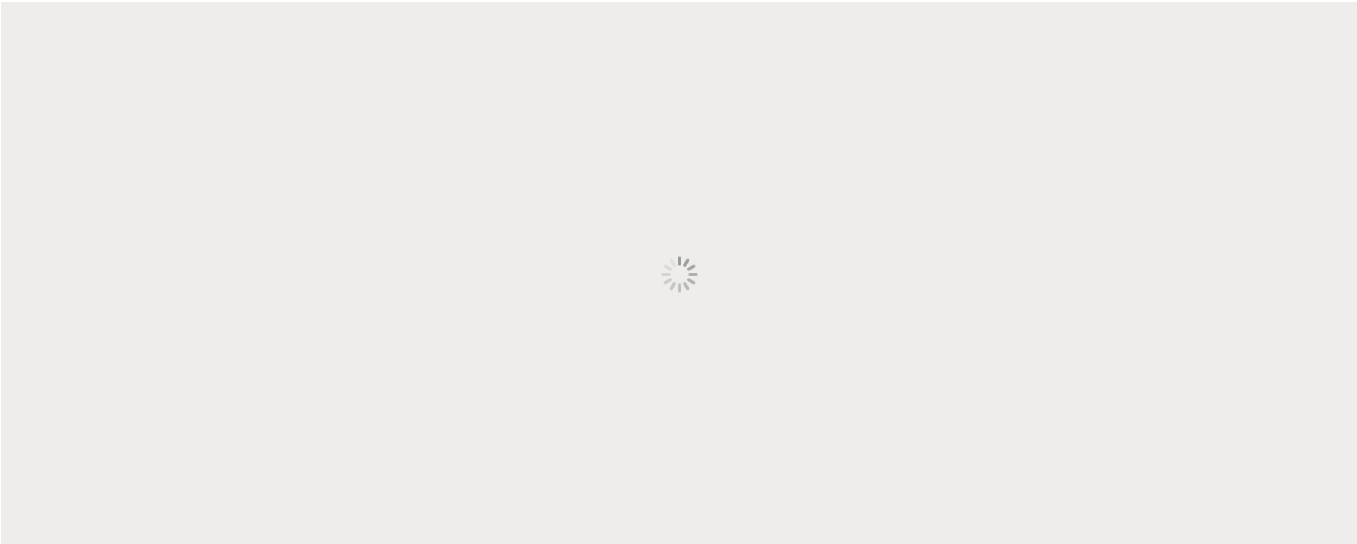
- 多头组合净值曲线（因子中性化）



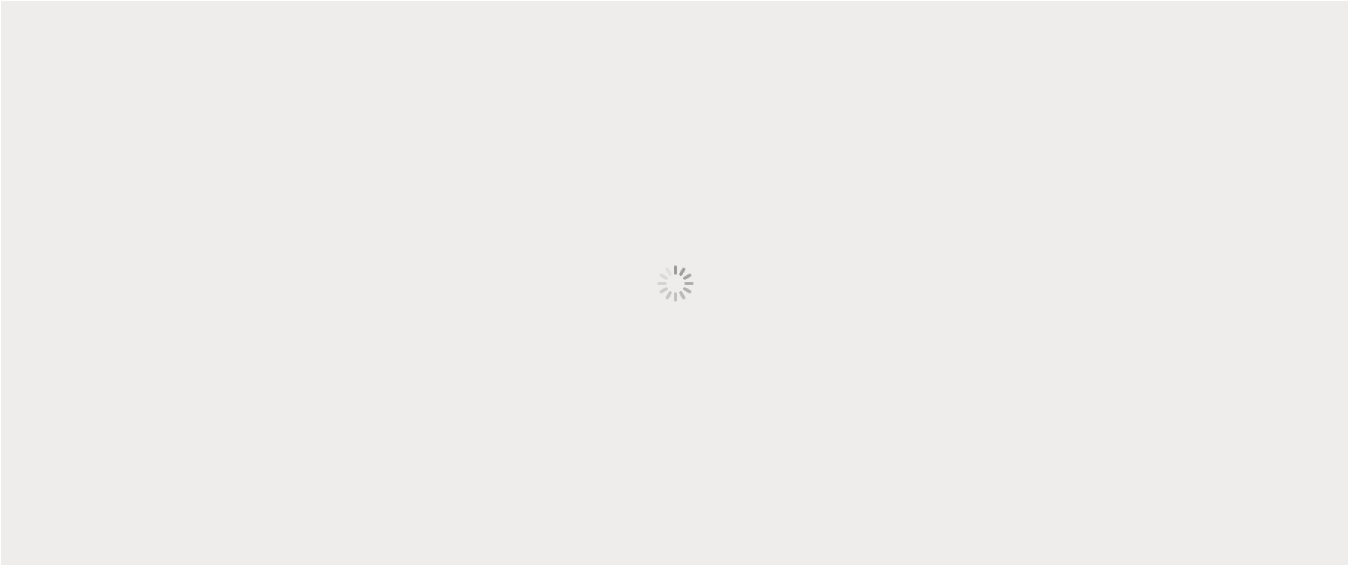
- 多空组合净值曲线（因子中性化）



- 多头组合净值曲线（因子未中性化）



- 多空组合净值曲线（因子未中性化）



市值因子（因子中性化）

- 多头组合净值曲线



- 多空组合净值曲线



- 市值因子不同组对比



可以看出，不论是PB因子还是市值因子，相对于沪深300指数，均有明显超额收益，但今年以来净值曲线存在回撤。

此外，同等条件下，多空组合的回撤明显小于多头组合，净值曲线更平滑。

未做中性化的因子效果明显比做中性化的因子效果差。



代码

```
1  if_neutral_industry = False
2  if_neutral_mktcap = False
3  monthdata['BP'] = 1/monthdata.pb_1f
4  # 标准化，中性化
5  monthdata = norm(monthdata,if_neutral_industry,if_neutral_mktcap)
6
7  if_ascending = False # 是否升序排列
8  bp_rank = monthdata[['BP']].rank(ascending = if_ascending)
9  monthdata.loc[bp_rank['BP']<=len(bp_rank)/group_num,'bp_rank'] = 'long'
10 monthdata.loc[bp_rank['BP']>len(bp_rank)*(group_num - 1)/group_num,'bp_rank'] = 'short'
11 # 多空组的股票代码
12 s_l = monthdata.loc[monthdata.mktband=='long',['logmktcap']]
13 s_s = monthdata.loc[monthdata.mktband=='short',['logmktcap']]
```

1. 20180306-方正证券-方正证券“远山”量化选股系列（一）：规矩，方正单因子测试之评价体系

收录于话题 #多因子模型·12个 >

< 上一篇

单因子测试（上）——因子中性化

下一篇 >

单因子测试（下）——回归测试法

喜欢此内容的人还喜欢

因未发奖金，一名程序员删代码泄愤，平台崩溃近三个小时

猿大侠



暴涨40%！北京楼市，正处于启动前夜？

米宅北京



马光远：最终压垮房地产泡沫的是低生育陷阱

光远看经济

