**实习生周报**

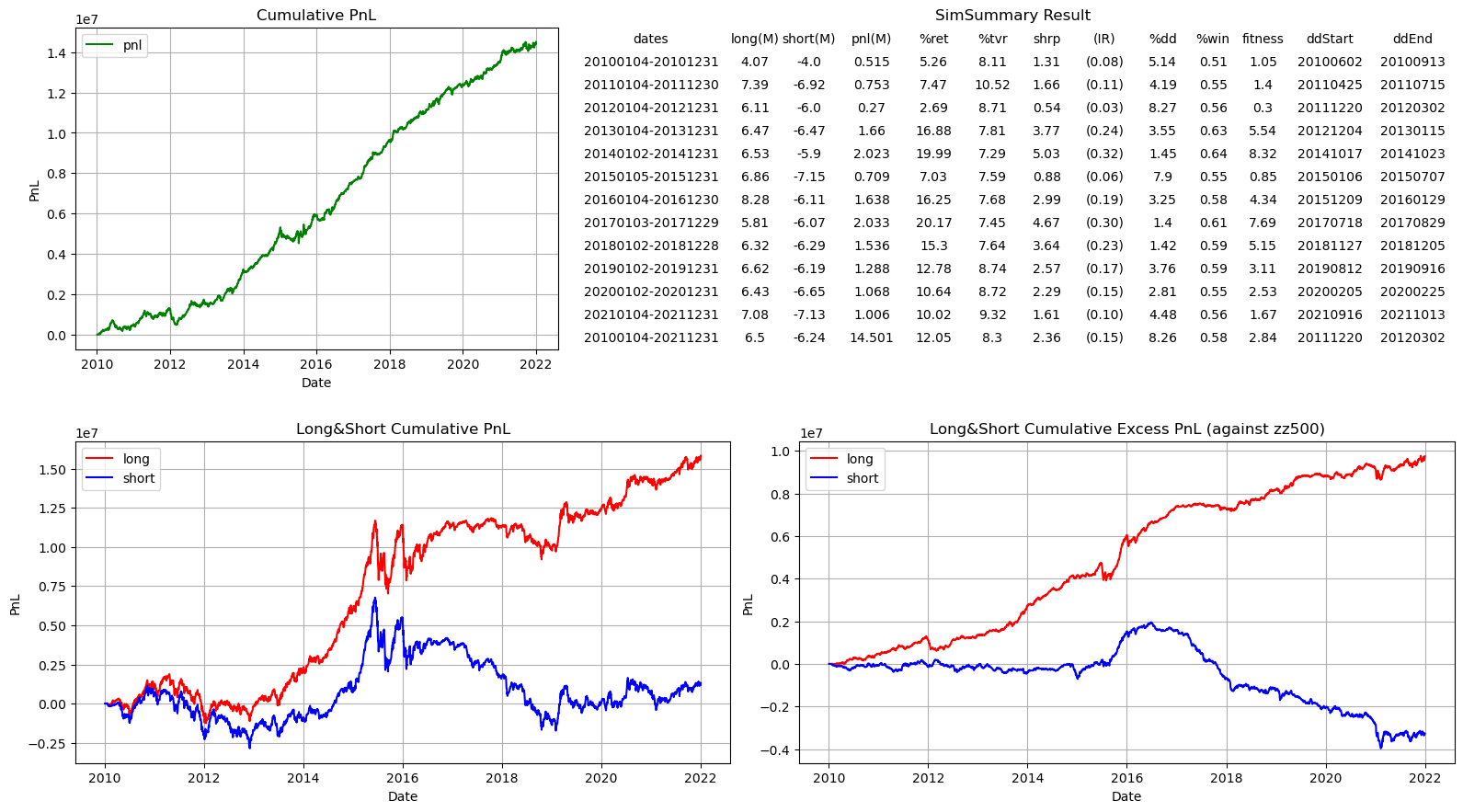
2021.01.10 实习生：崔晏菲

1. **近期小结**
   1. 任务目标是从写因子变成了写趋势因子，只写趋势
   2. 实现方法上从之前的堆砌数学公式，变成了先建立假设，再根据假设写相应的规则从而实现因子
   3. 尝试将开仓、平仓、止盈、止损加入因子逻辑中，逐渐尝试摆脱截面满仓中性化的框架，尝试从其中找到一些margin
   4. 目前刚刚在量价数据上熟悉了上述的写法，开始引入量价之外的数据（例如分析师、基本面、宏观经济指标等）与量价数据结合来构建因子
   5. 共写出6个趋势因子（5个量价趋势，1个量价+基本面趋势），并顺便写了2个非趋势因子
2. **因子展示**
   1. **AlphaLowVolatility**
      1. 因子思路

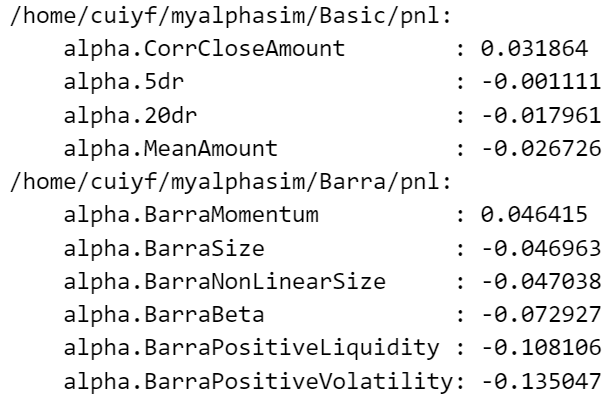
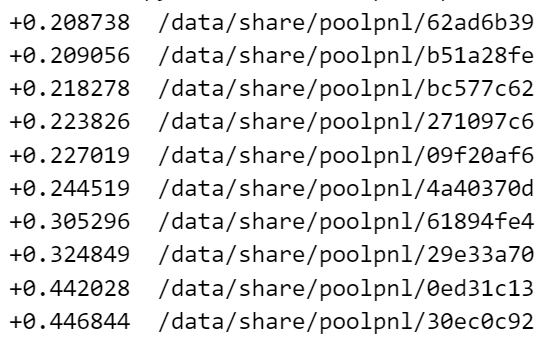
对于股价来说，在价格的波动中，趋势和反转信号一定是同时存在的，只不过有时反转强盖过了趋势信号而已。因此试图将趋势信号分解出来。假设股价每天的波动率有高有低，在过去的一段时间内，低波动率的交易日的收益率理论上趋势信息会更多；而高波动率的交易日的收益率理论上会更凸显反转趋势。

因此我取每只股票过去160个交易日的数据，以同一天全市场的波动率的平均值外加一个标准差作为阈值，将日内价格振幅低于阈值的交易日数据挑选出来，计算这些天相对指数的超额收益率的平均值，再加上我自己特制的开仓平仓和止盈止损规则，就得到了动量信号。

* + 1. 因子表现



* + 1. 因子相关性

和风格因子的相关性 和因子库的相关性

* + 1. 因子讨论

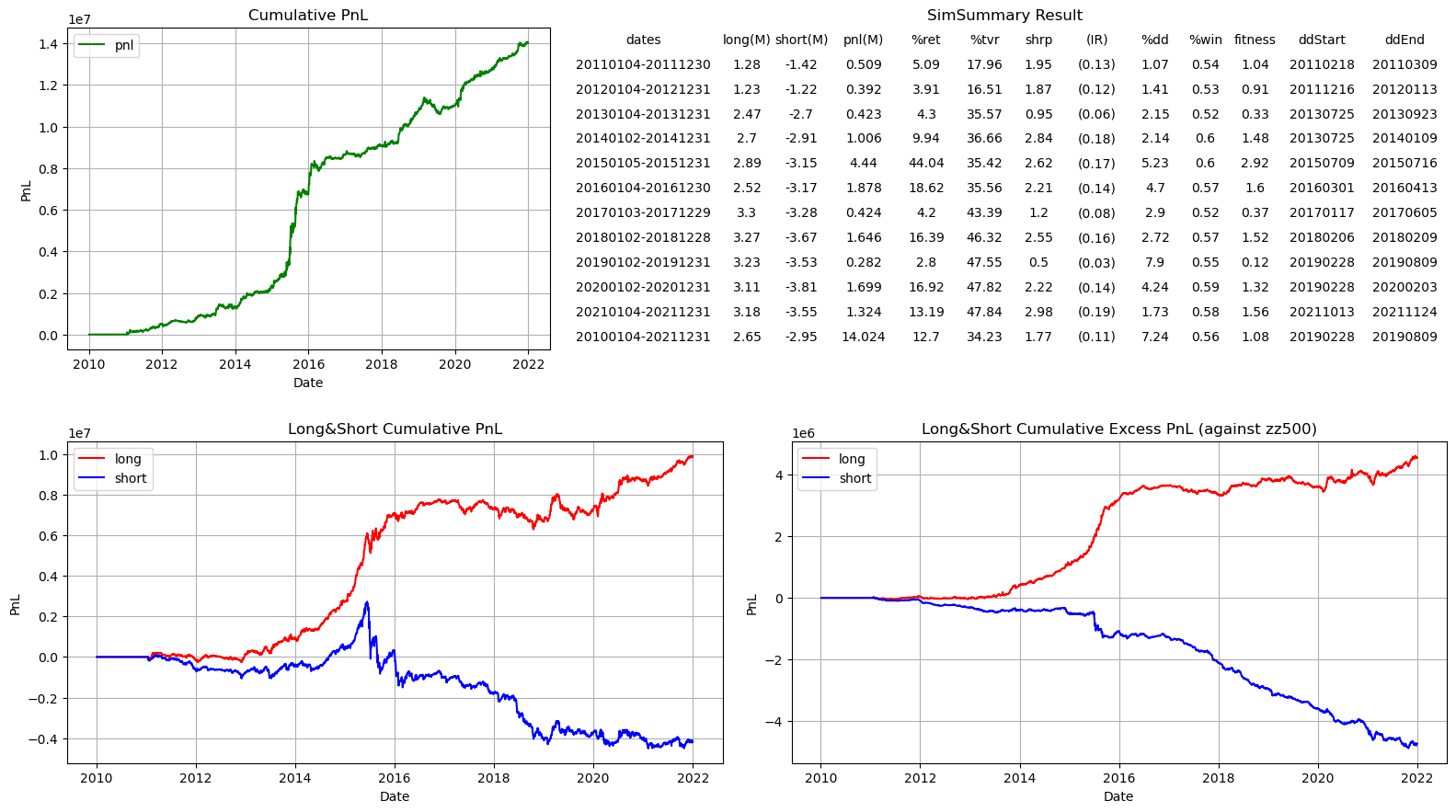
该因子几乎没有空头收益，空头头寸唯一的作用就是对冲beta。并且多头收益稳定跑赢中证500。我之后写的因子都有这样的特征，故不再赘述。

* 1. **AlphaIntradayMomentumD0**
     1. 因子思路

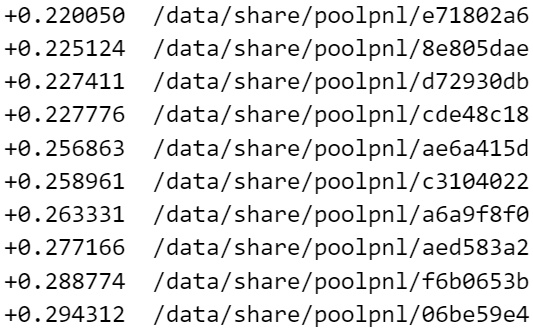
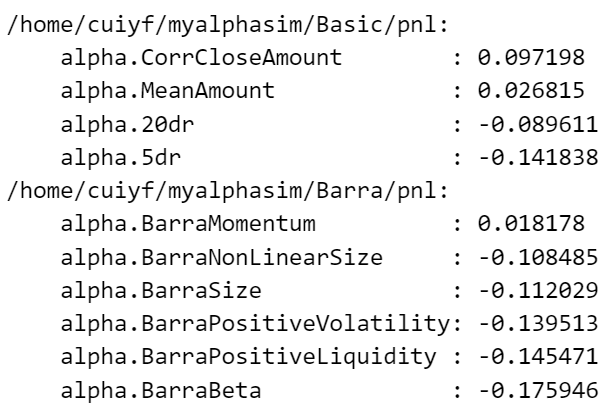
股票每天早上刚开盘是存在动量的。开盘的走势很大程度上决定了收盘的走势。因此我写了一个delay 0的因子，计算每天刚开盘半小时的个股收益率，并计算这半小时内的个股价格振幅，并计算这个振幅相对于全市场平均的振幅的比率，作为调整系数。调整系数和个股收益率相乘，再计算这个指标相对于过去一季度的zscore，就得到了因子值。

也就是我认为开盘的超额价格波动是能体现当天的日内趋势的，因为这可能代表了大量散户入场。

* + 1. 因子表现



* + 1. 因子相关性



和风格因子的相关性 和因子库的相关性

* + 1. 因子讨论

这个因子多头的收益没有跑输过中证500，并且近期开始重新跑赢。

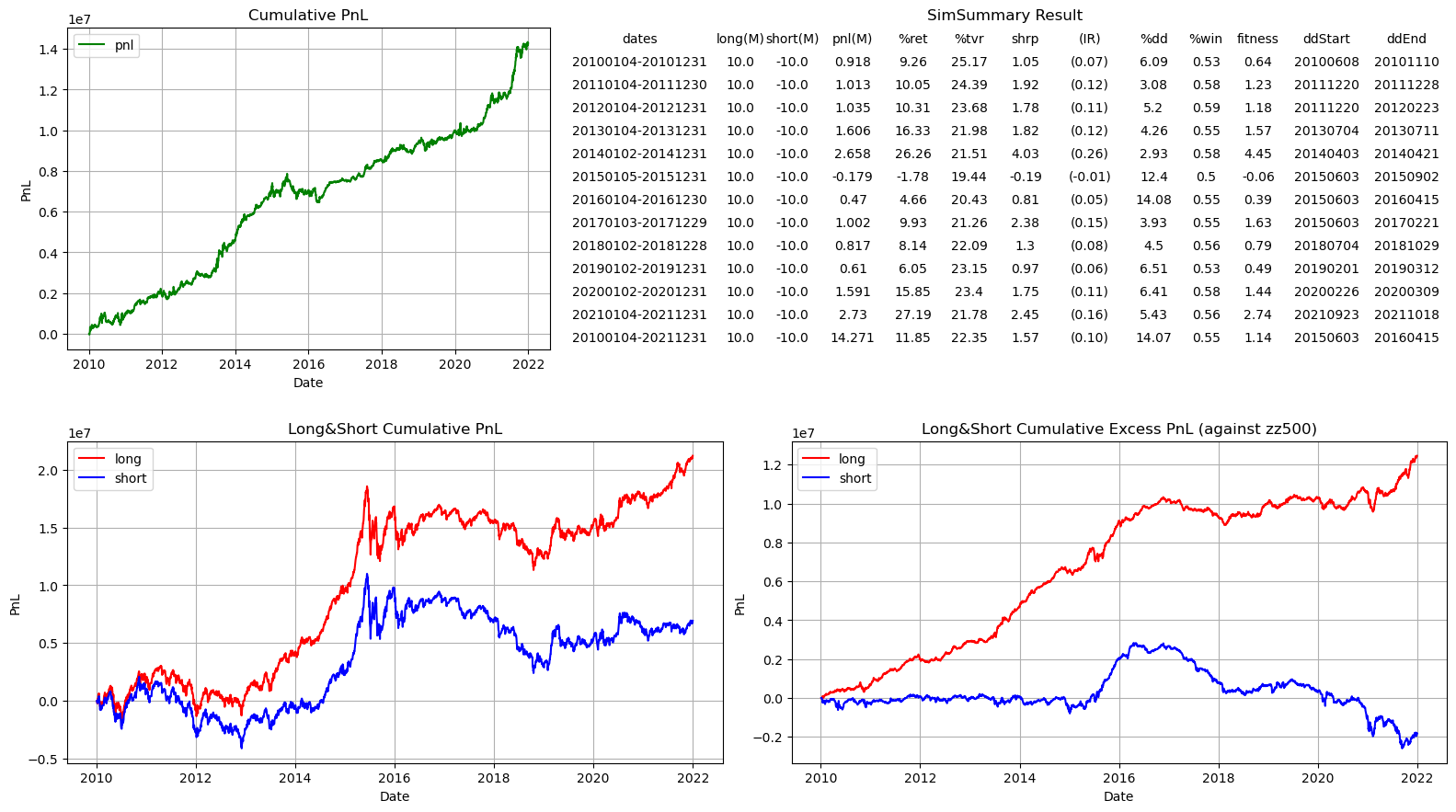
并且这个因子的回撤很小，尤其是2015年股灾期间很好地选出了不会狂跌的股票用于做多，导致股灾期间大赚了一把。

* 1. **AlphaDragonComove**
     1. 因子思路

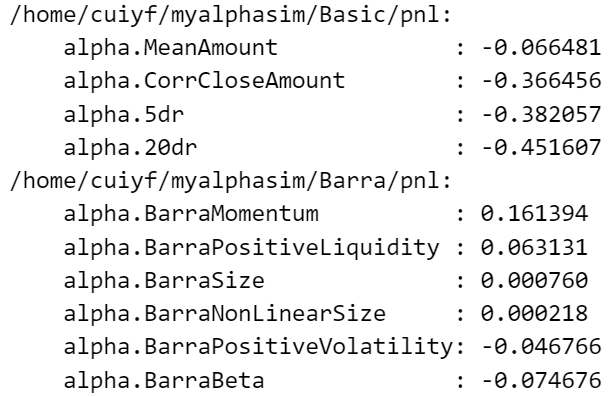
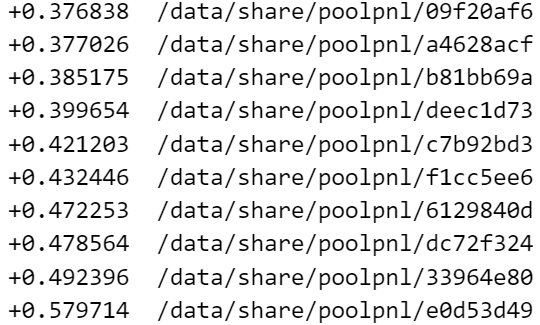
在A股中，经常会一个行业一起涨，或者一个概念一起涨。这往往是一些行业龙头带动整个板块一起涨，例如最近的“华为概念股”。因此这个因子试图找出行业龙头对整个行业的引领作用。

我将股票按照industry分组，计算最近21天每只股票的累积成交量，并计算组内zscore，组内zscore大于1.9的我就认为是龙头股票，并且我把最近21天有过涨跌停的股票也看做龙头股。计算每组的龙头股的最近21天的收益率的均值，将其应用到组内每只股票上，作为龙头股对个股的引领作用。接着对于每只股票，再用此指标除以个股近期的波动率，等于是惩罚波动率较大的，因为波动率大可能意味着趋势不明显。这样就得到了因子值。

* + 1. 因子表现



* + 1. 因子相关性

和风格因子的相关性 和因子库的相关性

* + 1. 因子讨论

根据同花顺上的新闻以及很多散户的投资逻辑，龙头股对行业、板块的引领作用确实是存在的。但往往不止局限于行业，例如产业链上下游、创始人人物关系等都可以作为龙头股引领的途径。因此这个因子还有巨大的改进空间。

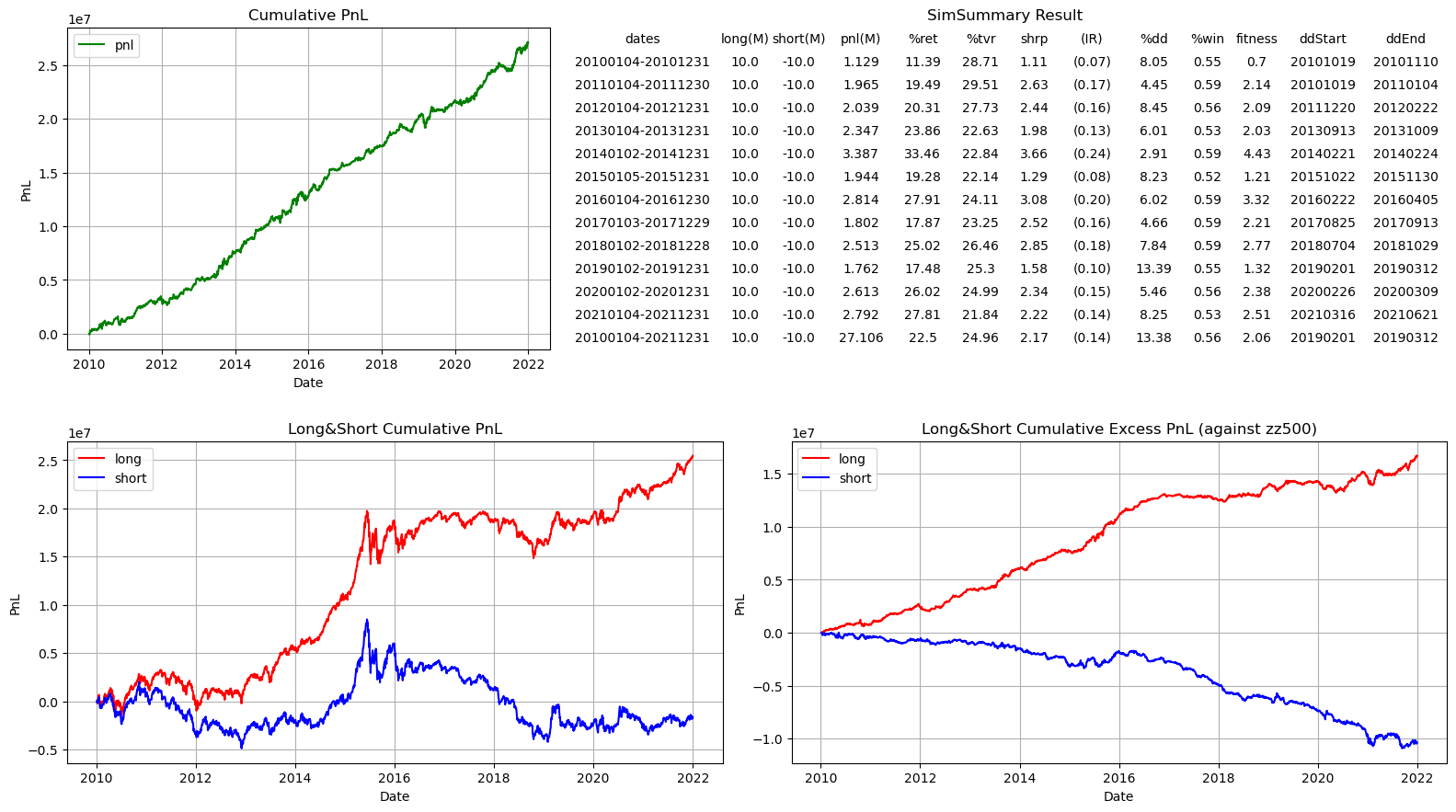
* 1. **AlphaPcaDragonComove**
     1. 因子思路

这个因子就是想对上一个龙头引领做出改进。试图只使用量价数据，能否挖到比行业更深乘次的股票之间的关联性？

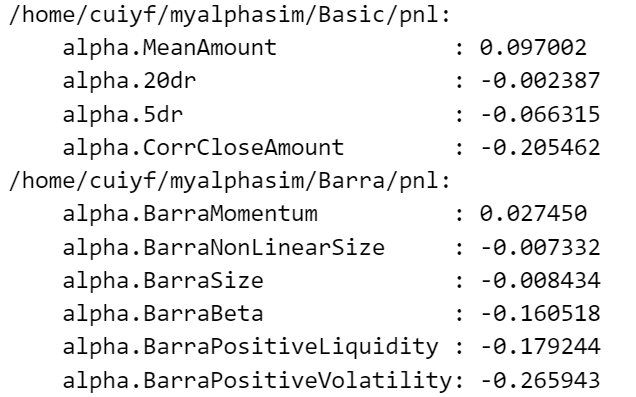
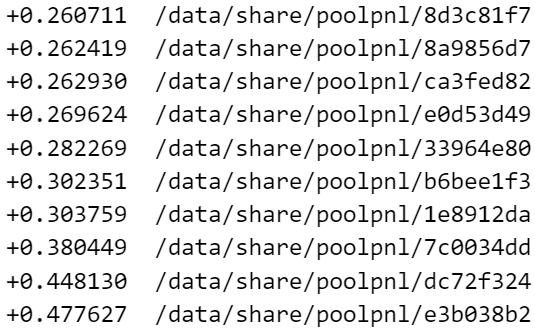
首先定义一个股票需要哪些特征，我认为行业、中期收益率、波动率和换手率是一只股票的基本特征，其中波动率可分为日内振幅、白天夜间的波动差异以及收益率的中期标准差，分别反映来了不同维度的波动信息。再辅以收益率的偏度和峰度这两个统计量作为辅助特征。就构成了我从量价指标对一只股票的刻画。那么如何根据这些指标来抓住股票之间的关联从而分组呢？

我想到了一种基于机器学习的方法——PCA。PCA会将这些特征融合分解，并且将数据投影到方差最大的那个维度从而进行降维。注意到，方差最大就意味着数据的可分性是最强的，因此我可以只保留方差最大的这一个维度，这样数据不仅有了可分性，还很方便通过数据的数值大小来对股票进行分组。这样得到的分组是原有特征的线性组合，可解释性非常强。我根据这样的分组再计算龙头股的引领作用，预期就能得到好的结果。

* + 1. 因子表现



* + 1. 因子相关性

和风格因子的相关性 和因子库的相关性

* + 1. 因子讨论

可以看到仅仅是用了PCA降维再做聚类，因子的表现就好了很多，pnl几乎成了一条直线，每年都很稳定。这说明PCA确实成功地把一些信息加入到了行业信息中做补充。

那为什么不多加几个维度呢？因为量价数据噪声太多，我们只有第一个维度能确保信噪比最强，之后的维度噪声都会盖过信号值。我试过单独只保留第二个维度进行聚类，发现这样的pnl在2017年之后就变成平的了。而从第三个维度开始，pnl曲线就像噪声曲线一样了。因此这也解释了为什么我之前不做降维使用GMM等聚类方法聚出的类效果都不尽如人意——因为数据噪声太多了。

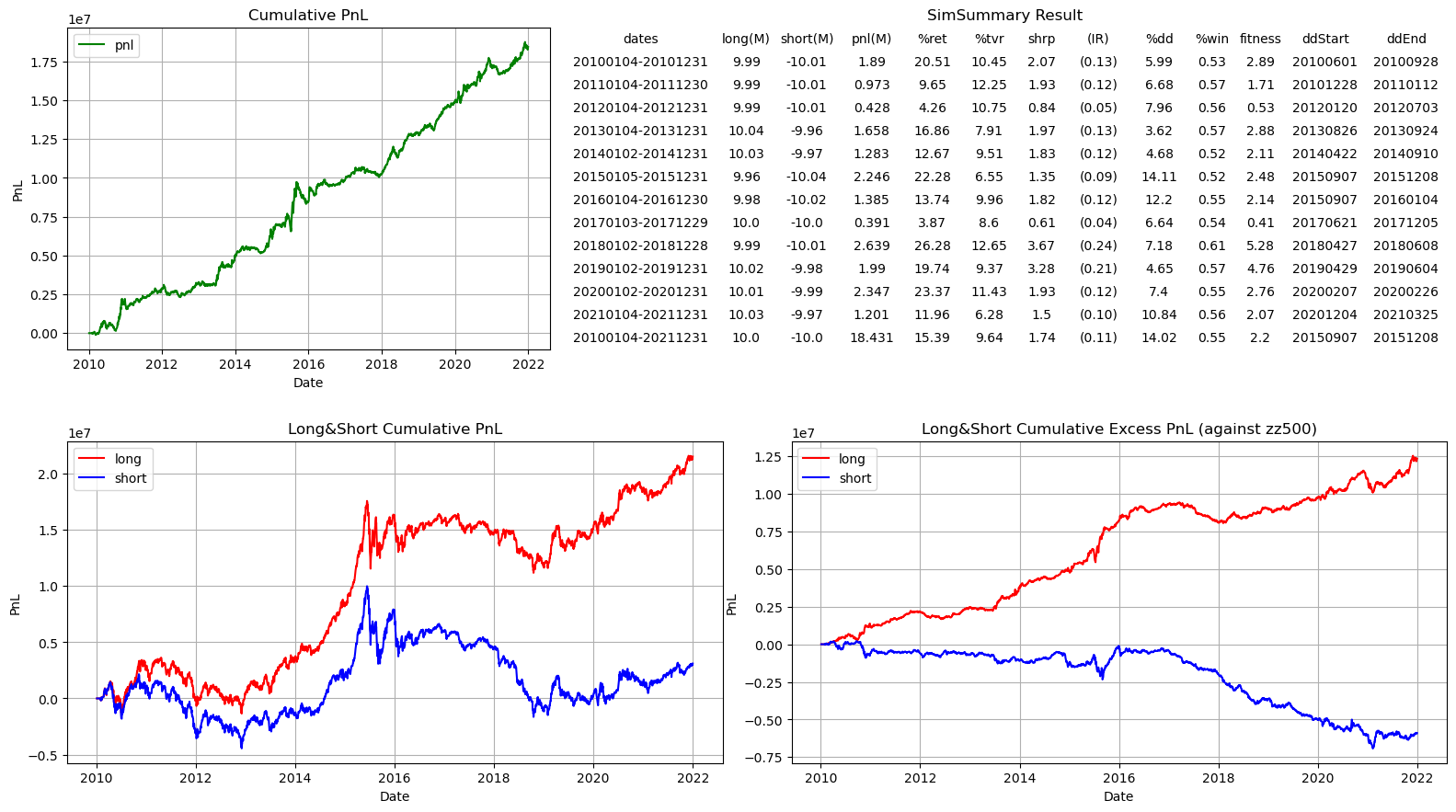
* 1. **AlphaFallRebund**
     1. 因子思路

这个因子想尝试一下时点动量。在一些特殊时刻，整个市场会发生较大的变化，从而导致这些时刻的信号值是有长期深远的动量效应的。例如A股全市场大跌反弹的动量。

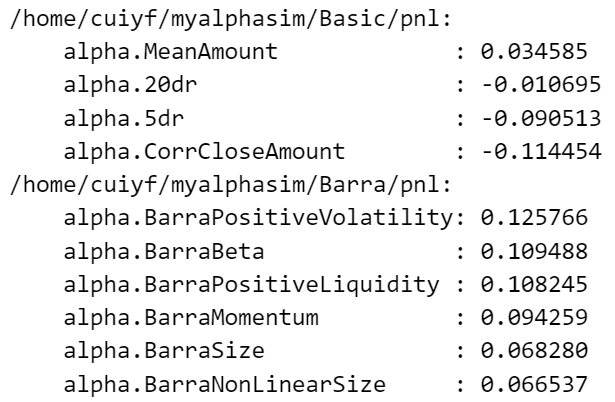
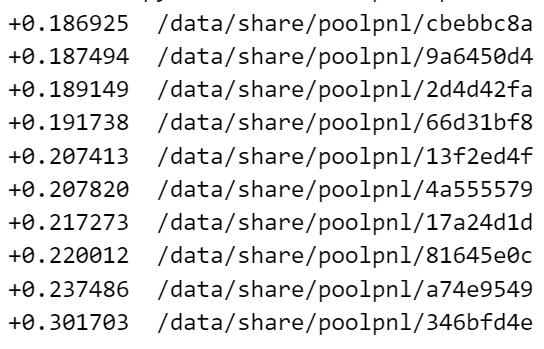
观察中证500的历史曲线，发现其日间波动率大约在1%~2%左右。因此我定义中证500单日涨幅超过1%就称其为大涨。若当日发生了大涨，那么回看过去一段时间的指数，如果在大涨日之前的过去一段时间指数的累积跌幅超过了5%，并且在这段下跌期内，单日涨幅均未超过大涨的阈值，则认为这是一段大跌行情。如果大涨和大跌同时存在，则认为该日是一个“大跌反弹日”，行情会在这一刻发生重要变化。当然了，这个判断大涨的阈值会随着市场整体波动率的大小进行微调。

在大跌反弹日当天，我计算每只股票的当日收益率相对于过去21天的zscore，作为信号值，并将该信号值从大到小分成10组，每组内所有股票的因子值是组内信号值的均值。这样就得到了我这个因子。因子值会一直保持不变，直到下一个大跌反弹日来临。

* + 1. 因子表现



* + 1. 因子相关性

和风格因子的相关性 和因子库的相关性

* + 1. 因子讨论

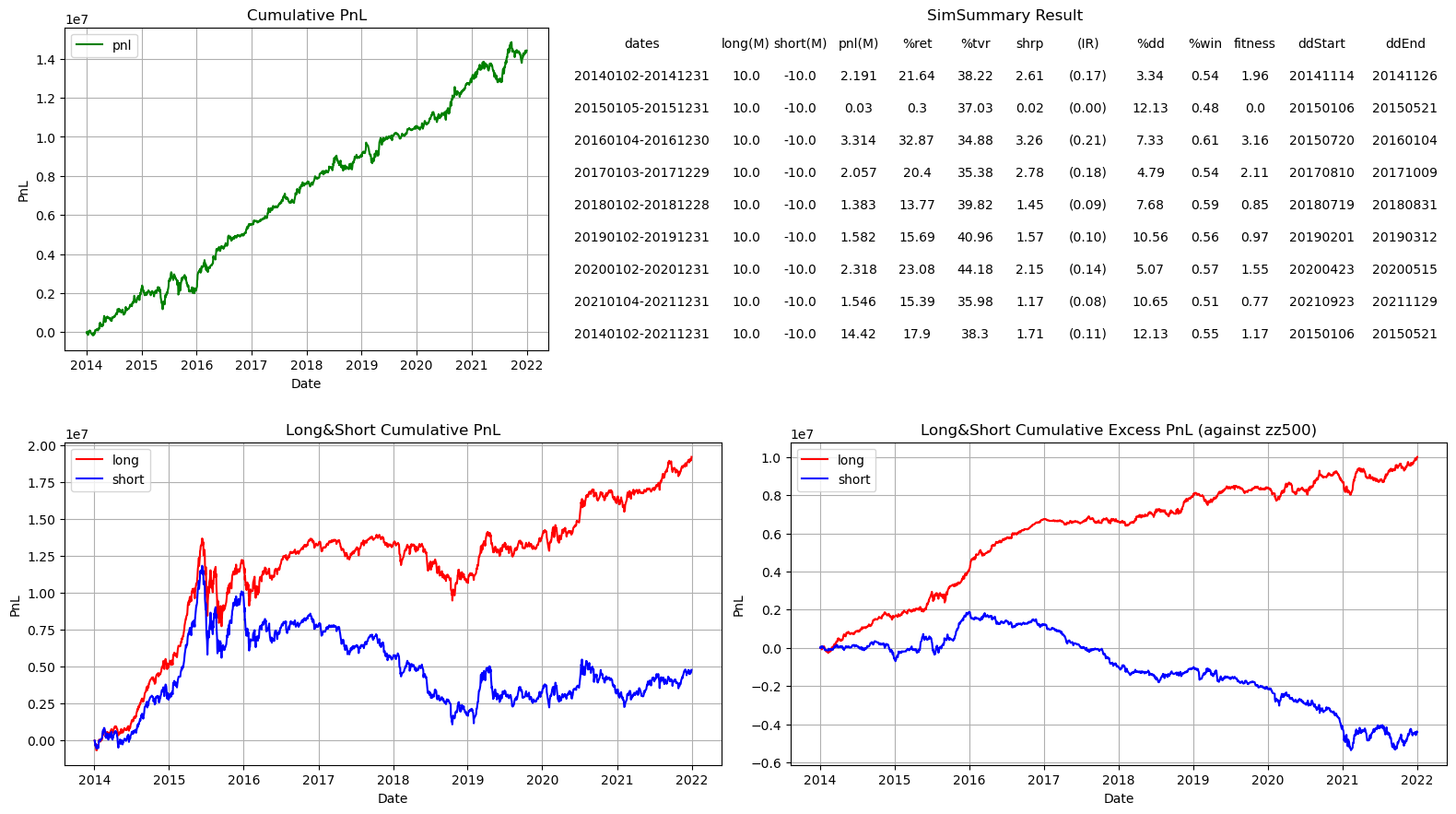
可以看到，这个因子的多头收益非常高，跑赢了中证500良多。并且该因子是一个罕见的正波动率、正流动性的因子。

* 1. **AlphaFundamentalPcaDragon**
     1. 因子思路

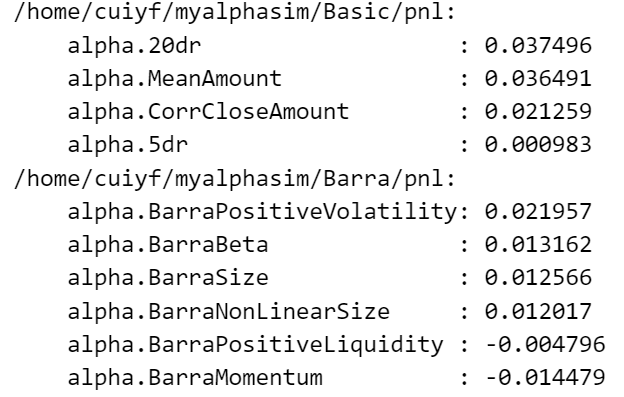
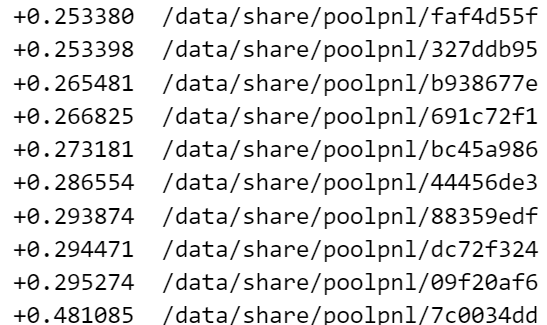
在量价趋势上历练了足够多之后，开始做基本面和量价数据的结合。这个因子仍然是对龙头引领因子的扩展延伸。只不过我这次使用的是行业、市盈率、资产周转率、毛利率、应收账款占总收入的比率作为描述公司的特征。之所以选取这五个，是因为市盈率代表公司当前股价的“便宜”程度；资产周转率反映出公司的营运能力；毛利率和应收账款代表了公司在上下游的议价能力。

对这些特征使用PCA降至一维后再分组计算龙头股的引领效果，就得到了我这个因子。

* + 1. 因子表现



* + 1. 因子相关性

和风格因子的相关性 和因子库的相关性

* + 1. 因子讨论

该因子和前几个龙头股因子相关性很低，最高只有0.005，这意味着这样分组确实挖掘到了新的龙头股引领网络。并且该因子多头每年都跑赢了中证500，十分难能可贵。