

02 因子示例——从一日动量出发的研究

想法描述

因子计算

股票的一日涨跌幅cp_rtn（或者是一日动量mom_1）

思路来源

这个指标实在过于简单，正常来说我们切入因子的角度要比这个复杂得多。

逻辑猜测

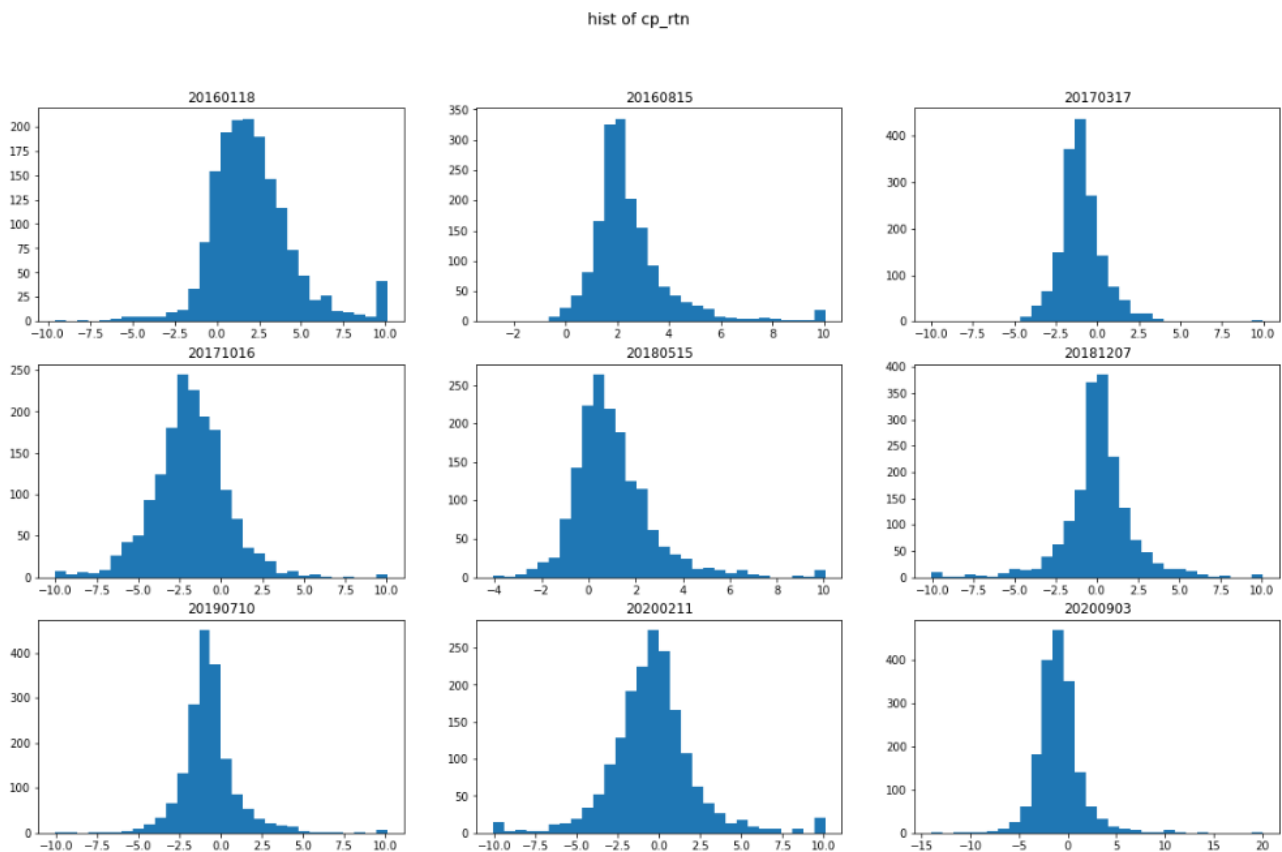
1. 散户有追涨杀跌的习惯，所以今天涨得多的票明天会接着涨，今天跌得多的明天还会跌。产生动量效应。
2. 由于股市的T+1规则，散户的追涨杀跌只体现在当日，隔夜情绪会在第二天对这一趋势进行修正，产生反转效应。
3. 可能只有大涨或者大跌的票产生趋势或者反转效应，因为只有大涨大跌的票才会引发关注和情绪。对于大部分涨跌幅很小的票（例如-2%-2%），可能很难说明天会出现什么动量或者反转。这两类票上面的现象可能会不一样。

因子描述

因子值分布

截面分布

- 随便挑一些天画当天的cprtn直方图，可以看到收益大致服从正态分布，符合预期（但是偏度和峰度可能跟正态分布有较大区别！）
- 另外，由于涨跌停有点像两个“吸收壁”，因此分布多少会存在肥尾。下图的20160118、20200211，涨停处的异常非常显眼。



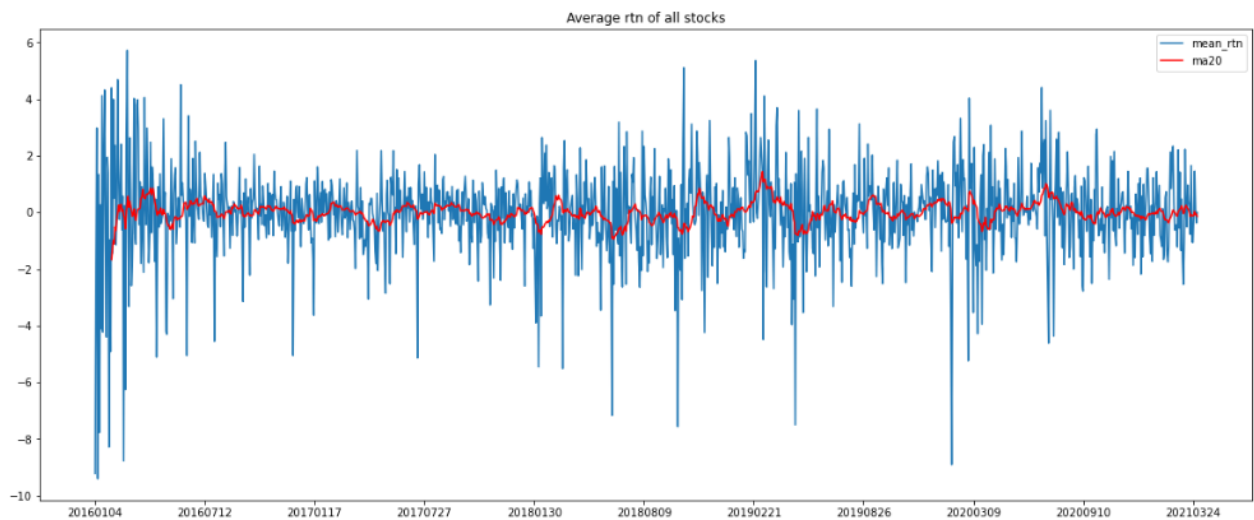
时序分布

- （图一）因子的均值和分布模式在时序上没有明显的变化（这是显然的，因子均值基本就是指数收益，长期看总是在0左右波动）
- （图二）`cp_rtn`的std在时序上也基本是平稳的（或者你认为它在19年以来有缓慢的上升趋势？）
- （图三）教大家画个小提琴图，可以看看时序上的分布情况。（这个小提琴图可以看出来`cp_rtn`的分布和正态分布的峰度存在很大差距）

```

: fig = plt.figure(figsize=(20, 8))
  plt.plot(fac_df.mean(), label='mean_rtn')
  plt.plot(fac_df.mean().rolling(20).mean(), label='ma20', color='red')
  plt.legend()
  plt.xticks(fac_df.columns[::int(len(date_list) / 10)])
  plt.title('Average rtn of all stocks')
  plt.show()

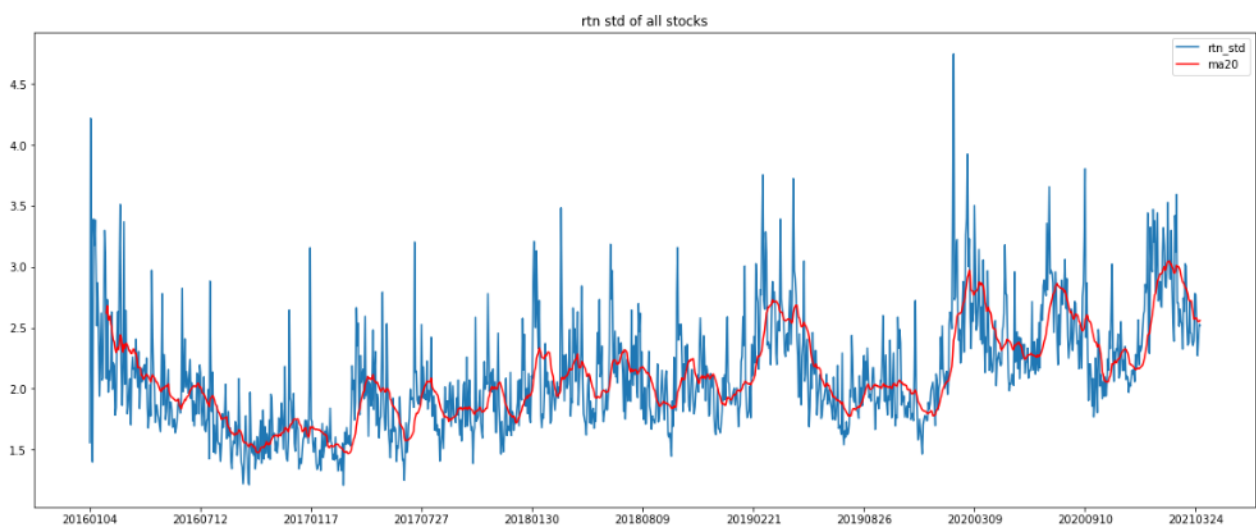
```



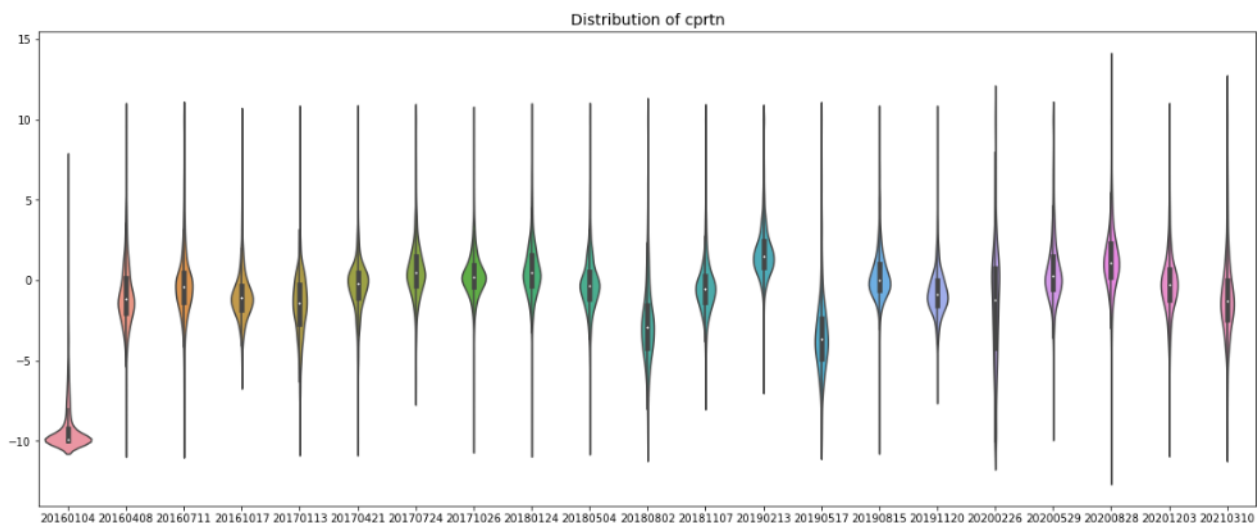
```

fig = plt.figure(figsize=(20, 8))
plt.plot(fac_df.std(), label='rtn_std')
plt.plot(fac_df.std().rolling(20).mean(), label='ma20', color='red')
plt.legend()
plt.xticks(fac_df.columns[::int(len(date_list) / 10)])
plt.title('rtn std of all stocks')
plt.show()

```

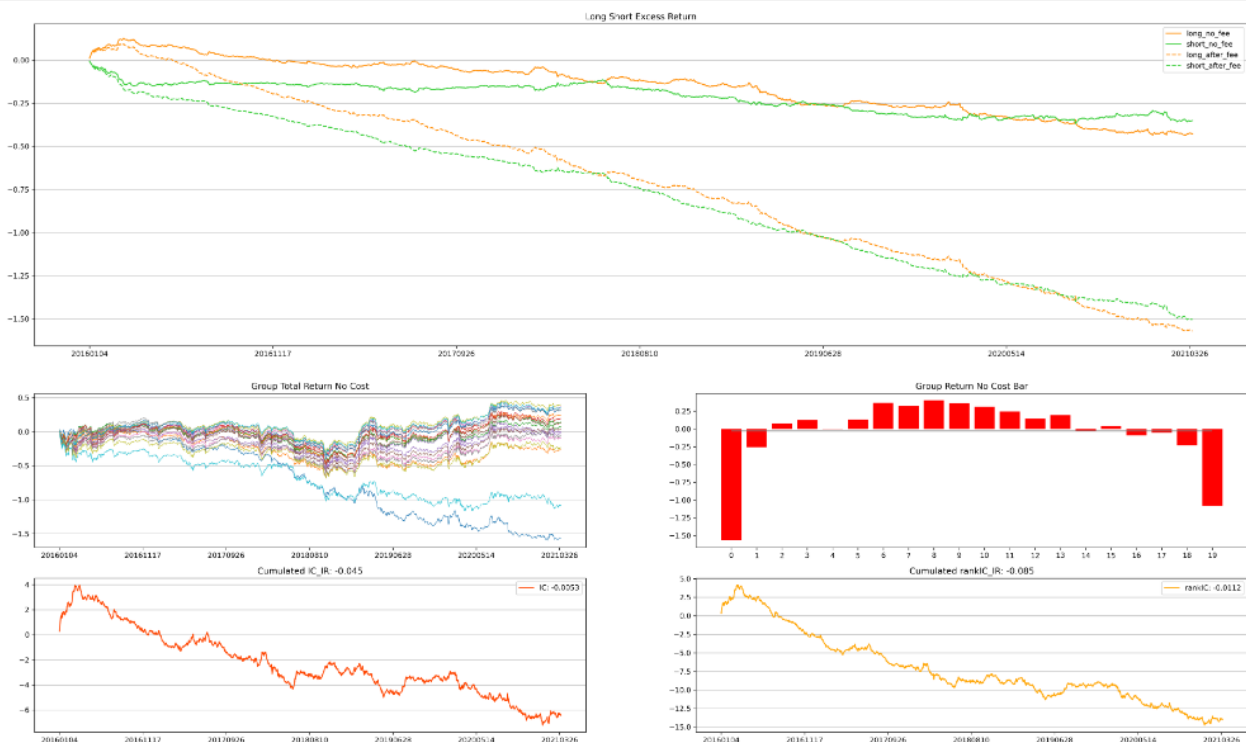


```
fig = plt.figure(figsize=(20, 8))
sns.violinplot(data=fac_df.loc[:, ::int(len(date_list) / 20)])
plt.title('Distribution of cprtn', fontsize=14)
plt.show()
```



因子回测

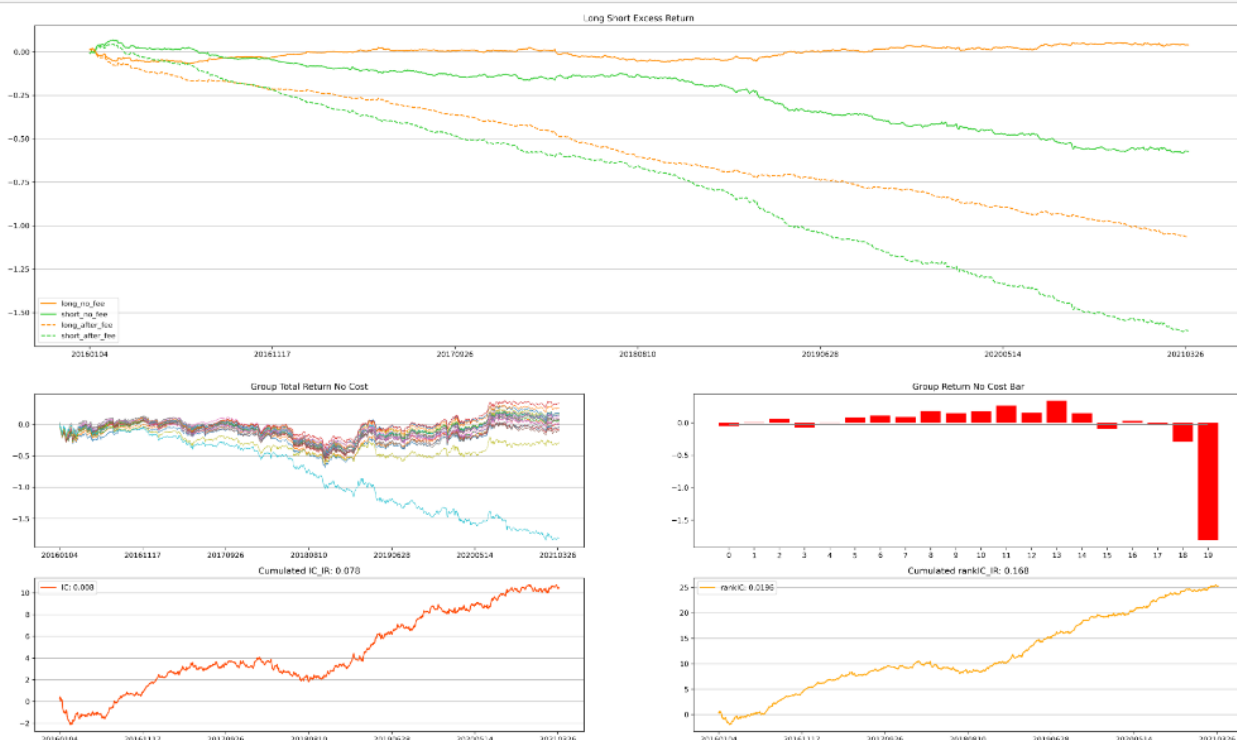
```
: res1, res2 = alter_backtest(factor_df=normal_process(fac_df), benchmark='1800', start_date='20160101', end_date='20210331',
                             rtn_df=RET_df.copy(), head='factor', ifplot=True)
res1
```



- 显然cprtn本身不存在和次日ret的线性预测能力
- 但是！因子分组体现出显著的倒U形，也就是说今天跌得多或者涨得多的票明天都是不好的，但是如果涨跌幅比较小，那么会有超出平均水平的收益

根据上述分析结果，修改一下测试的因子，把cprtn变成abs(cprtn)，重新进行回测

```
res1, res2 = alter_backtest(factor_df=normal_process(abs(fac_df)), benchmark='1800', start_date='20160101', end_date='20210331',
                           rtn_df=RET_df.copy(), head='factor', ifplot=True)
res1
```



并没有出现想要的多头组，但是已经出现了明显的空头组，且因子分组朝着有序的方向移动了。显然这个操作对于提升预测能力是有帮助的。现在我们需要思考一下，abs(cprtn)到底在表达什么

猜想验证

提出猜想

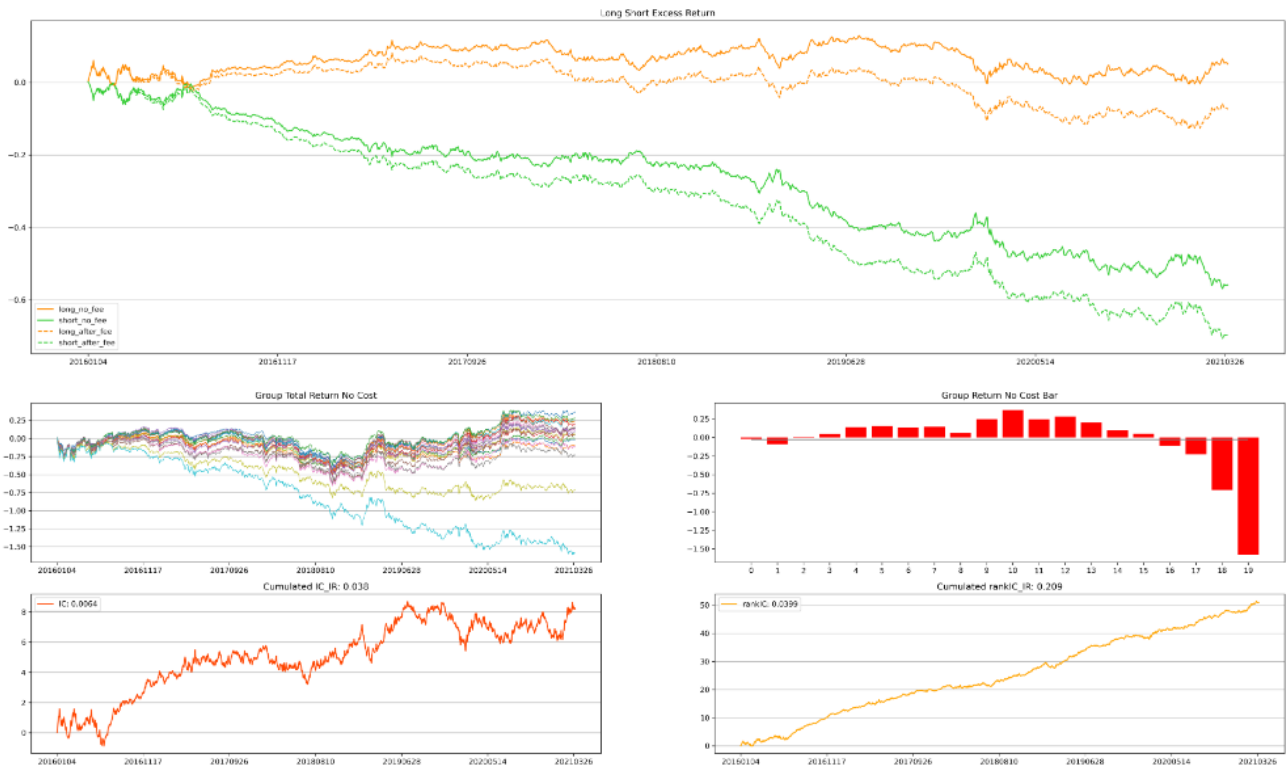
cprtn的回测图中，倒U形的分组收益已经否定了逻辑猜测中的1和2，而是在说涨或者跌，幅度大都不好，幅度小比较好。这其实不是在衡量rtn，而是在衡量波动的大小。由此我们可以提出一个猜想：波动小的票存在正向alpha。

验证猜想

什么指标可以衡量波动呢？

- 最容易想到的就是过去N天的RangeRate

```
1 fac_df = roll(rr, 10, 'mean')
2 fac_df = fac_df * fac_mask
```

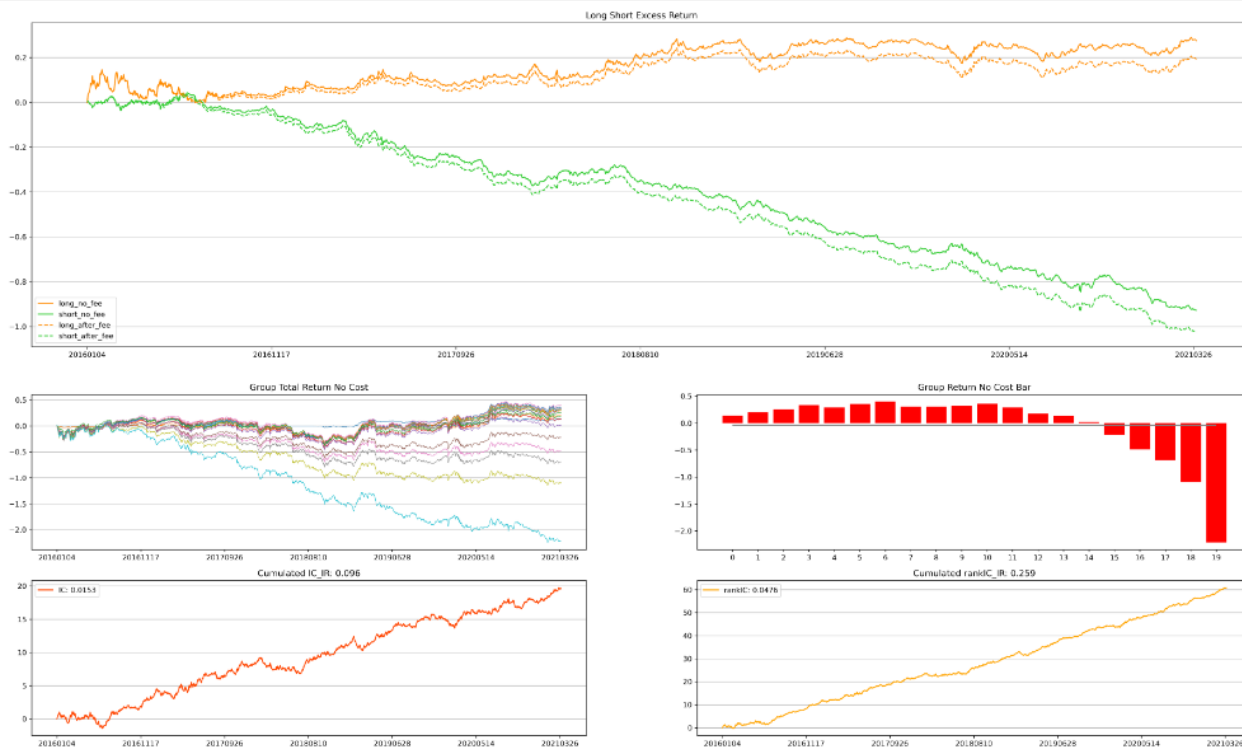


多头组收益好像还是不明显

- 或者可以考虑使用过去一段时间rtn的std

```
1 fac_df = roll(cprtn_test, 20, 'std')
2 fac_df = fac_df * fac_mask
```

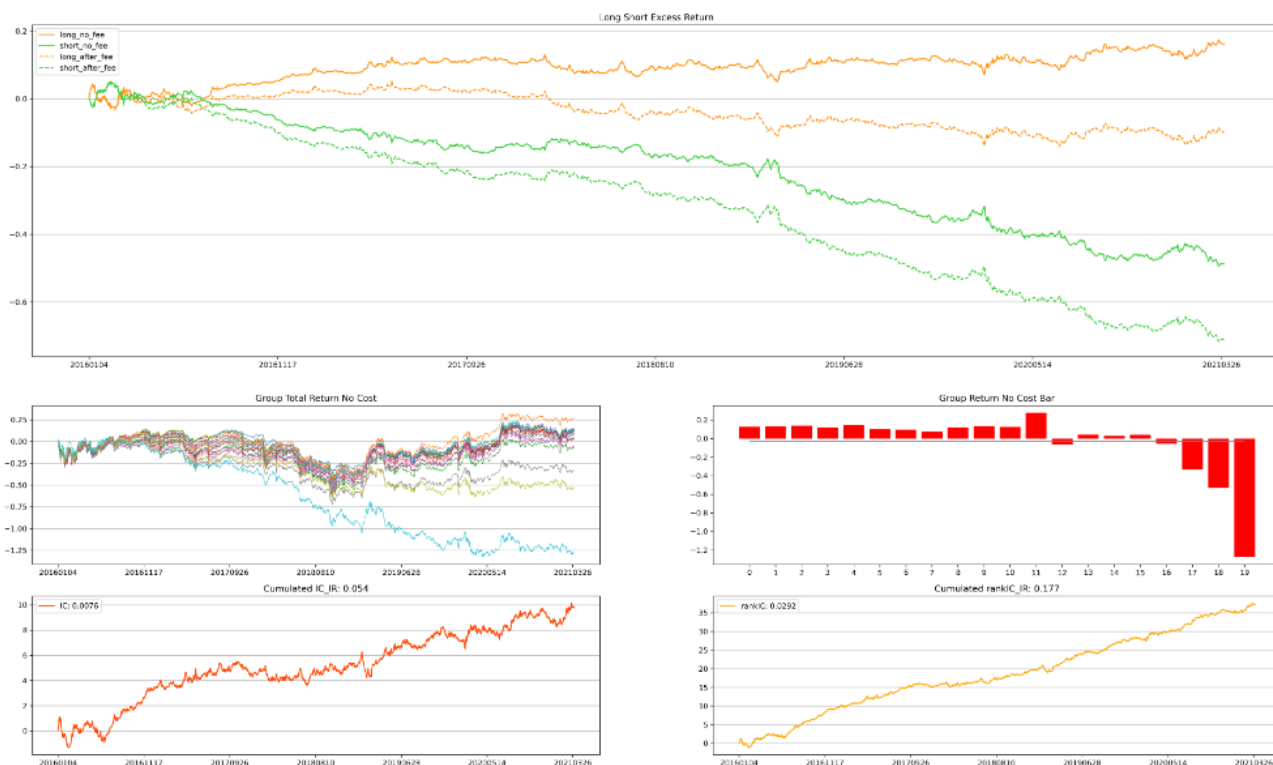
```
res1, res2 = alter_backtest(factor_df=normal_process(fac_df), benchmark='1800', start_date='20160101', end_date='20210331',
                           rtn_df=RET_df.copy(), head='factor', ifplot=True)
res1
```



rtn的rolling_std已经产生了一个因子的雏形：有显著区分的多空头，有较显著有序的分组收益

- 或者可以考虑过去一段时间股价的波动（std of prc），注意prc要归一化

```
1 fac_df = roll(cp, 10, 'std') / roll(cp, 10, 'mean')
2 fac_df = fac_df * fac_mask
```



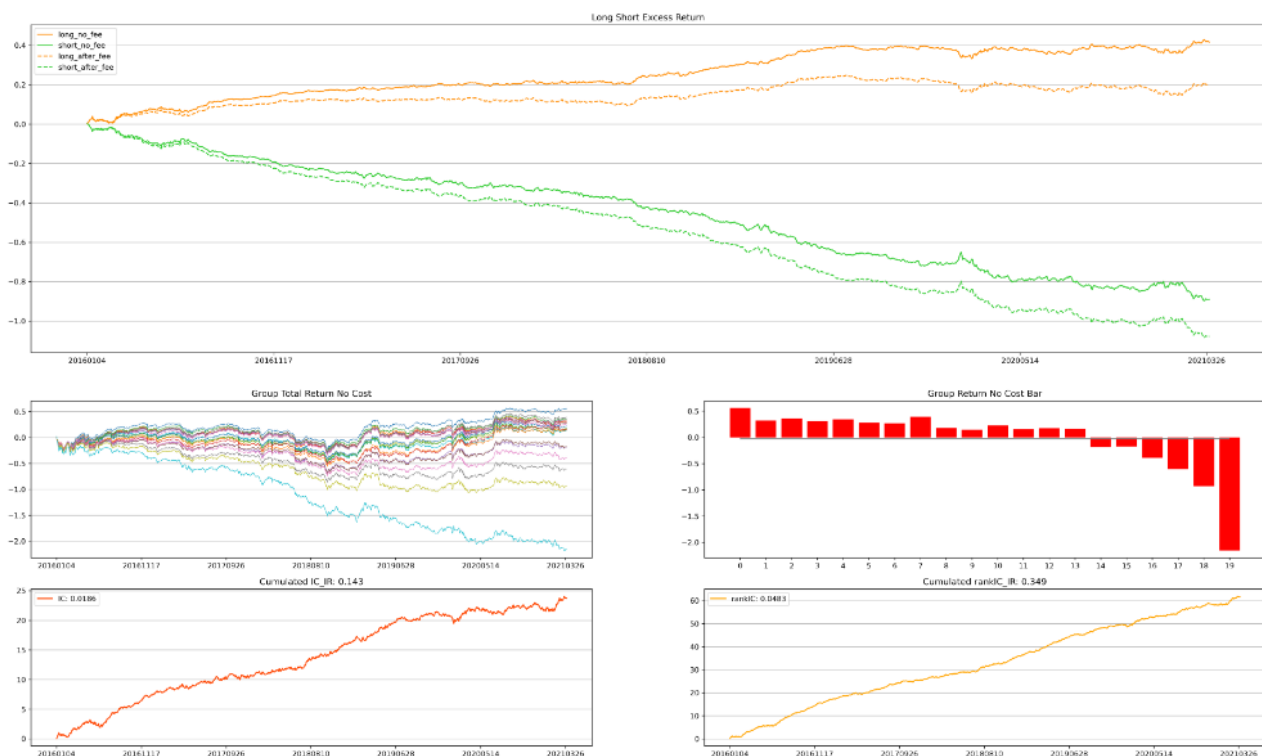
因子含义其实和rtn_std差不多，效果好像也并不比rtn_std明显

以上几个因子测试可以告诉我们的信息是：日间低波动是一个正向的alpha，日间波动大小对未来收益有预测能力。接下来我们根据这个被验证的猜想写因子并进行优化

逻辑外推

1. 我们对“波动”的定义是什么？价格产生了移动就是波动吗？假如某天指数大跌5%，而某只股票仅跌了0.1%，这只股票今天的“波动”是大还是小呢？“波动小”是收益贴近0还是走势随波逐流？也许可以考虑用指数（或者当天cprtn的均值代替0）

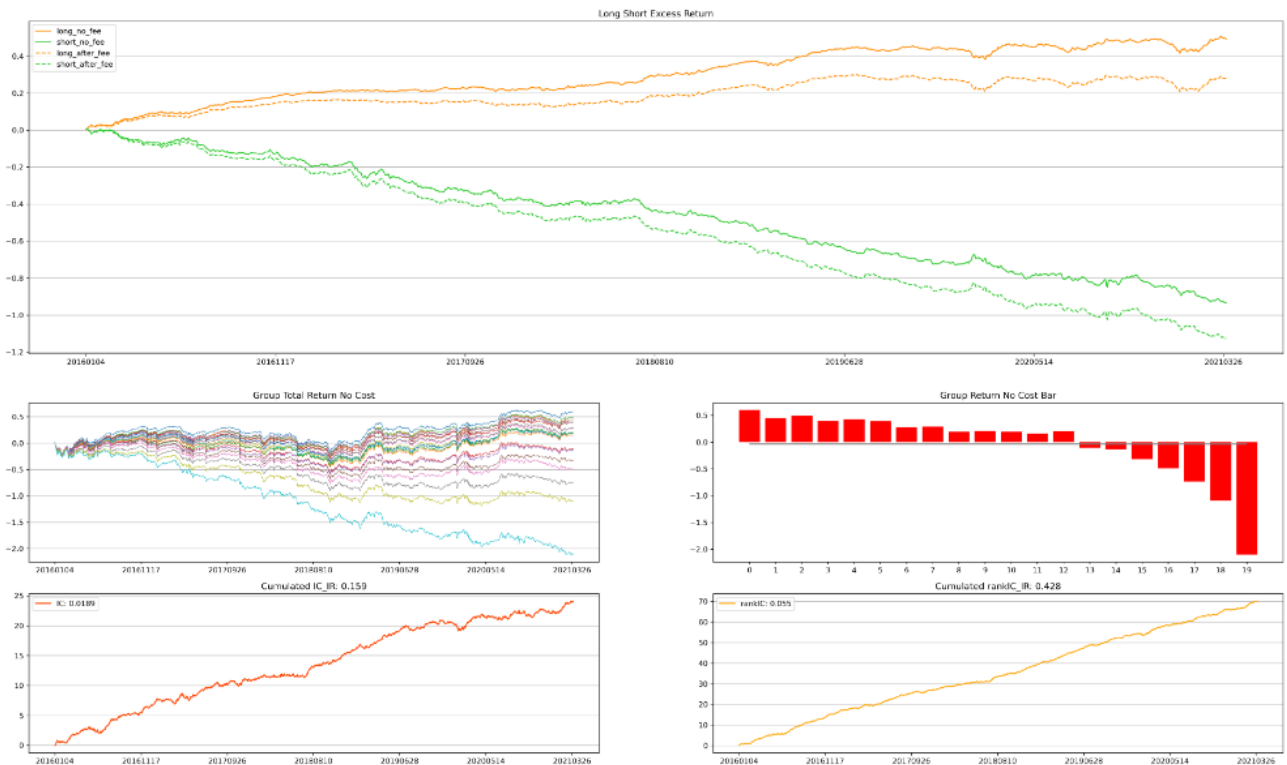
```
1 fac_df = abs(cprtn_test - cprtn_test.mean())
2 fac_df = roll(fac_df, 10, 'mean')
3 fac_df = fac_df * fac_mask
```

	AlphaRet	AlphaRetNC	AlphaSharpe	AlphaSharpeNC	AlphaDD	TurnOver
2016	0.119	0.162	2.487	3.369	0.029	0.134
2017	0.000	0.043	0.001	1.339	0.019	0.136
2018	0.049	0.091	1.097	2.043	0.021	0.133
2019	0.046	0.088	1.168	2.241	0.024	0.133
2020	-0.059	-0.019	-0.966	-0.308	0.050	0.128
2021	0.043	0.052	2.212	2.687	0.021	0.123
sum	0.056	0.098	1.000	1.895	0.065	0.131

这个逻辑修正可以说对因子的预测能力产生了根本性的变化！大幅提升了因子的预测能力，一定是一个正确的优化方向。

2. 在一开始对cprtn的直方图统计中，我们看到涨跌停处会出现异常分布。这些作为个别情况出现的涨跌停可能波动的计算造成很大影响。也许可以考虑计算时去除这些情况



	AlphaRet	AlphaRetNC	AlphaSharpe	AlphaSharpeNC	AlphaDD	TurnOver
2016	0.158	0.200	4.321	5.459	0.014	0.132
2017	-0.021	0.021	-0.676	0.685	0.024	0.133
2018	0.098	0.140	2.255	3.236	0.019	0.134
2019	0.034	0.076	0.913	2.008	0.026	0.130
2020	-0.027	0.013	-0.443	0.212	0.053	0.125
2021	0.033	0.043	1.574	2.029	0.032	0.127
sum	0.064	0.106	1.324	2.271	0.079	0.130

在计算cpnrtn时去除涨跌停，对该因子的表现有较显著的提升。之后的研究中可以默认使用去除涨跌停的票池。

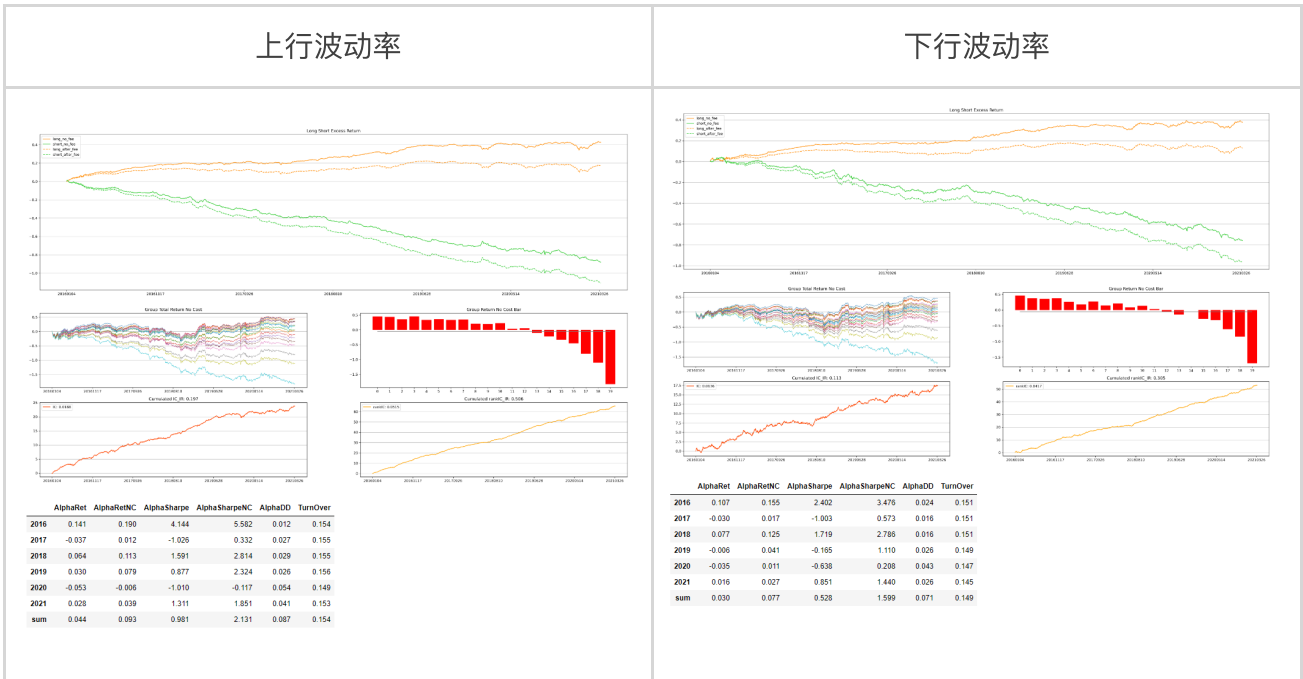
到这里我们已经做出一个弱有效的因子了，可以做一个参数搜索，找到最佳的回看期（参数搜索这一步也可以早一点做，但是参数不能设置太密）

	1	3	5	10	20	40	60
AlphaRet	-0.116	-0.011	0.037	0.064	0.079	0.074	0.065
AlphaRetNC	0.091	0.094	0.109	0.106	0.103	0.088	0.075
AlphaSharpe	-3.531	-0.434	0.778	1.324	1.462	1.221	1.068
AlphaSharpeNC	2.642	2.368	2.555	2.271	1.971	1.496	1.261
AlphaDD	0.072	0.077	0.078	0.079	0.094	0.102	0.106
TurnOver	0.651	0.329	0.225	0.130	0.075	0.043	0.030

	1	3	5	10	20	40	60
AlphaRet	-0.004	0.051	0.068	0.081	0.083	0.077	0.073
AlphaRetNC	0.108	0.116	0.115	0.110	0.100	0.087	0.080
AlphaSharpe	-0.295	0.930	1.245	1.412	1.377	1.220	1.131
AlphaSharpeNC	2.410	2.349	2.232	1.991	1.708	1.408	1.266
AlphaDD	0.076	0.077	0.078	0.087	0.097	0.104	0.108
TurnOver	0.353	0.204	0.147	0.091	0.054	0.031	0.023

左表为rolling_mean，右表为ewm_mean。综合考虑可以选取rolling5或者ewm3。

3. 同为波动，向上波动和向下波动的效果一致吗？如果把波动分拆成上行波动和下行波动，二者表现相似还是有所区别呢？



二者没有显著差别，且单个表现不如整体的波动率因子表现。这个优化尝试是无效的。

因子分析

观察因子的回测结果，可以发现在2019年下半年以后，因子绩效显著变差，多头收益几乎完全消失



	AlphaRet	AlphaRetNC	AlphaSharpe	AlphaSharpeNC	AlphaDD	TurnOver
2019	-0.032	0.003	-2.181	0.211	0.021	0.216
2020	-0.062	0.007	-1.142	0.126	0.048	0.217
2021	0.031	0.048	1.611	2.490	0.030	0.225
sum	0.002	0.072	-0.571	0.942	0.078	0.219

一般情况下是头组收益降低甚至为负，但这个因子却没有出现这样的情况。说明并不是头组的衰减导致因子多头的失效。

因子优化

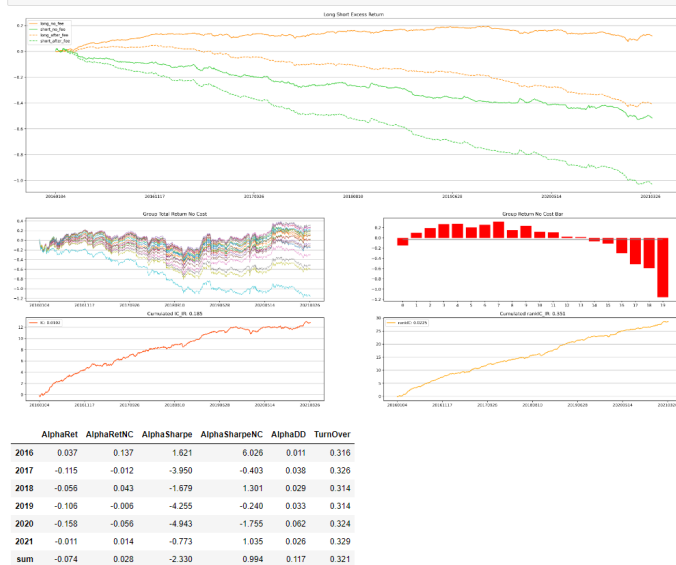
该因子有一个显著的问题，就是低波动本质上是一个市场风格，而不是纯alpha收益

```
calc_exposure(fac_df)
```

```
residual_volatility: 0.496
momentum: 0.185
growth: 0.112
earnings_yield: -0.164
book_to_price: -0.286
liquidity: 0.252
range_1: 0.510
tovr_1: 0.491
mom_5: 0.110
range_5: 0.632
tovr_5: 0.517
mom_10: 0.150
range_10: 0.564
tovr_10: 0.461
mom_20: 0.161
range_20: 0.492
tovr_20: 0.405
```

```
fac_df = neutralize(['residual_volatility', 'range_5', 'size'], normal_process(fac_df))
```

```
res1, res2 = alter_backtest(factor_df=normal_process(fac_df), benchmark='1800', start_date='20160101', end_date='20210331',
                             rtn_df=RET_df.copy(), head='factor', ifplot=True)
```



对因子进行中性化后，因子alpha几乎消失。由此可见大部分收益都是低波这一市场风格带来的。

Demo总结

本文只是帮助展示写因子过程的一个demo，并没有提供一个真正的alpha因子。事实上，本文的切入点“低波动”本身就是一个风格因子。但是在把它当做一个alpha因子的研究过程中，每一步的分析、思考和优化，对于写alpha因子是通用的。希望大家能举一反三，不受流程框架束缚，灵活思考，开拓思路，写出具有独创性的好因子。