

发布时间：2017-12-06

证券研究报告 / 金融工程研究报告

## Alphalens 使用教程

### 报告摘要:

Alphalens 是一款 Python 的工具包, 是 Quantopian 公司旗下三大开源包之一, 另外两个分别是 Zipline 和 Pyfolio. Alphalens 因其简单易用而又稳定科学的优势被广泛的量化分析师所青睐. 本文主要介绍 Alphalens 的回测框架, 提纲挈领地介绍了其中四个功能: 因子收益、因子 IC、因子换手以及事件研究。

在进行 Alphalens 的使用之前, 需要先对数据进行清洗。事前需要准备好因子数据、价格数据和行业分组数据。其中因子数据需要处理成 Dataframe 格式, 且其索引必须是二维日期和股票; 价格数据是列为时间, 行为股票的 Dataframe; 行业数据可有可无, 若有则可以设置成 Dict 格式或者 Series 格式。然后通过 Alphalens 的 get\_clean\_factor\_and\_forward\_returns 函数对原始数据进行清洗, 清洗之后返回股票、未来收益率和行业分类的因子数据 factor\_data, 然后调用 create\_full\_tear\_sheet 进行回测。

因子收益部分包括了因子分组超额收益分布直方图和琴型图、因子的累计收益曲线、超额收益曲线、因子加权收益、因子收益分布琴型图, 因子 spread 等结果。

因子 IC 部分包括了因子 IC 表、因子 IC 时间序列、因子 IC 分布图和 QQ 图、因子 IC 热力图等结果。

因子换手部分包括了因子分组平均换手率、因子换手率时间序列及因子排序自相关性等结果。

事件研究部分包括了因子分组平均超额收益随时间的关系及每组超额收益随时间的分布特征等。

Alphalens 只是在进行因子简单回测的第一步, 其优点在于可以批量的测试每个因子的性能, 极大的拓宽了投资者的思路而不用担心回测结果的正确性和稳定性。但是 Alphalens 也有其不足之处, 它仍然只是一个理想状态的回测, 真实的回测环境还要考虑手续费, 涨停买入和跌停卖出等限制。但瑕不掩瑜, 投资者可以修改代码以完善相应功能。

### 大、中、小盘净值曲线



	上周涨幅	今年累计涨幅
大盘指数	-2.97%	29.09%
中盘指数	-0.92%	10.22%
小盘指数	1.16%	-0.04%

### 配置建议

**择时观点:** 震荡看空沪深300, 震荡看平中证500, 震荡看平创业板指;

**风格配置:** 配置 30% 大盘, 30% 中盘, 40% 小盘

**行业推荐:** 银行、交通运输、纺织服装;

### 相关报告

《金融工程: HMM 指数择时研究之实战篇》

2016-09-26

《金融工程: HMM 指数择时研究之理论篇》

2017-01-17

《人工智能系列报告综述篇: 人工智能发展史及算法介绍》

2017-10-11

### 证券分析师: 高建

执业证书编号: S0550511020011

### 研究助理: 肖承志

执业证书编号: S0550116080014

021 2036 1264 xiaocz@nesc.cn

请务必阅读正文后的声明及说明

“慧博资讯”是中国领先的投资研究大数据分享平台

点击进入  <http://www.hibor.com.cn>

## 目 录

<b>1. 认识 Alphalens.....</b>	<b>4</b>
<b>2. Alphalens 使用教程 .....</b>	<b>4</b>
2.1. 第一步：数据预处理.....	4
2.2. 因子收益、IC 与换手率.....	7
2.2.1. 因子收益.....	9
2.2.2. 因子 IC 分析 .....	12
2.2.3. 因子换手率分析.....	13
2.3. 其他功能：事件研究.....	14
<b>3. 小结.....</b>	<b>16</b>

## 图表目录

图 1: 因子清洗和收益对齐.....	4
图 2: 数据清洗的第一个输入变量: 股票的因子值.....	5
图 3: 数据清洗的第二个输入变量: 股票的价格.....	5
图 4: 数据清洗的其他参数变量.....	6
图 5: 数据清洗的结果.....	7
图 6: 因子回测结果汇总.....	8
图 7: 因子收益: 分组平均超额收益 (收益中心化处理) .....	9
图 8: 因子收益: 分组平均绝对收益 (收益未中心化处理) .....	10
图 9: 因子收益: 分组超额收益分布图 (收益中心化处理) .....	10
图 10: 因子收益: 因子加权的 (多空) 累计收益净值曲线.....	10
图 11: 因子收益: 因子分组累计收益净值曲线.....	11
图 12: 因子收益: 因子分组超额收益净值曲线.....	11
图 13: 因子收益: 第十组-第一组平均收益分布情况.....	11
图 14: 因子 IC 分析: IC 时间序列图.....	12
图 15: 因子 IC 分析: IC 的分布图和 QQ 图 .....	13
图 16: 因子 IC 分析: IC 热力图.....	13
图 17: 因子换手率分析: 第一组 (绿) 和第十组 (蓝) 换手率 .....	14
图 18: 因子换手率分析: 因子排序的自相关图.....	14
图 19: 事件研究: 因子分组平均超额收益随时间的关系.....	15
图 20: 事件研究: 第一、二组超额收益分布随时间的关系 .....	15
图 21: 事件研究: 第三、四组超额收益分布随时间的关系 .....	15
图 22: 事件研究: 第五、六组超额收益分布随时间的关系 .....	16
图 23: 事件研究: 第七、八组超额收益分布随时间的关系 .....	16
图 24: 事件研究: 第九、十组超额收益分布随时间的关系 .....	16
表 1: 因子收益率分析: 不同周期下的收益率结果.....	9
表 2: 因子 IC 分析: 不同周期下的 IC 统计结果.....	12
表 3: 因子换手率分析: 不同周期下的分组换手率统计结果.....	13

## 1. 认识 Alphalens

我们先看看开发 Alphalens 的 Quantopian 公司是如何介绍它的：“Alphalens 是一款 Python 的工具包，用于分析 alpha 因子的表现。它创造了横截面分析的公平算法。Alpha 因子表达了一种在给定信息集与未来收益率的可预测关系，通过将这种关系应用于多个股票，我们可以希望产生一个 alpha 信号并交易它。开发一个好的 alpha 信号是有挑战性的，那么我们（Quantopian 公司）在哪些方面可以让事情变得更简单，并且您（Quant，量化分析师）又在哪些方面增加了最多的价值？我们认为，一套用于分析 alpha 因子的通用工具会产生深远的影响。”

这也是我们写这篇报告的目的，授人以鱼不如授人以渔。Quantopian 公司旗下有三大开源包，分别是 Alphalens，Zipline 以及 Pyfolio，Alphalens 用于因子开发，Zipline 用于回测策略，Pyfolio 用于策略分析。

运用 Alphalens，我们可以在做因子研究的过程中花更少的时间编写和运行回测框架。因此，我们可以更快地进行思维风暴，而不必怀疑算法最终的结果。用 Alphalens 构建严谨的工作流程将使策略的稳定性提高并且不易过拟合。

## 2. Alphalens 使用教程

### 2.1. 第一步：数据预处理

首先，我们要做的就是数据的预处理。可以说，只要做好这一步几乎就完成了 95% 的工作。而这一步其实也非常简单，只需要用一个函数实现即可。下面我们将全面解读这个函数。

我们先看一下这个函数 `get_clean_factor_and_forward_returns`，它的描述是“将因子数据、价格数据以及行业分类按照索引对齐地格式化到一个数据表中，这个数据表的索引是包含日期和资产的多重索引”，我们理解就是获取清洗后的因子及其未来收益（可以包含行业，也可以不包含行业），并将它们的收益对齐。

图 1：因子清洗和收益对齐

```
def get_clean_factor_and_forward_returns(factor,
                                         prices,
                                         groupby=None,
                                         by_group=False,
                                         quantiles=5,
                                         bins=None,
                                         periods=(1, 5, 10),
                                         filter_zscore=20,
                                         groupby_labels=None):
    """
    Formats the factor data, pricing data, and group mappings
    into a DataFrame that contains aligned MultiIndex
    indices of date and asset.
```

数据来源：东北证券，Alphalens

通过图 1 我们可以了解到这个函数有 2 个输入变量：股票的因子值（factor），股票的价格（prices），以及 7 个参数变量：股票的行业分组（groupby），是否按行业分组（by\_group），分组个数（quantiles），直方图个数（bins），因子换手周期

(periods), 异常值阈值设定 (filter\_zscore), 行业分组标签 (groupby\_labels)。

第一个输入变量是股票的因子值, 它是一个序列 (Series) 并且具有多重索引 (MultiIndex), 该多重索引的两个索引分别是日期 (date) 和股票代码 (asset), 而且日期的索引层级 (level 0) 优先级高于股票代码的索引层级 (level 1)。正如图 2 所示那样, 这个 Series 的第三列 (前两列是索引) 才是股票的因子值, 如苹果公司 (AAPL) 的因子值是 0.5。实现多重索引的方法是 df.set\_index(['date', 'asset'])。

图 2: 数据清洗的第一个输入变量: 股票的因子值

```
Parameters
-----
factor : pd.Series - MultiIndex
A MultiIndex Series indexed by date (level 0) and asset (level 1),
containing the values for a single alpha factor.
::
-----
      date | asset |
-----
      | AAPL | 0.5 |
      | BA  | -1.1 |
2014-01-01 | CMG  | 1.7 |
      | DAL  | -0.1 |
      | LULU | 2.7 |
-----
```

数据来源: 东北证券, Alphalens

图 3: 数据清洗的第二个输入变量: 股票的价格

```
prices : pd.DataFrame
A wide form Pandas DataFrame indexed by date with assets
in the columns. It is important to pass the
correct pricing data in depending on what time of period your
signal was generated so to avoid lookahead bias, or
delayed calculations. Pricing data must span the factor
analysis time period plus an additional buffer window
that is greater than the maximum number of expected periods
in the forward returns calculations.
'Prices' must contain at least an entry for each date/asset
combination in 'factor'. This entry must be the asset price
at the time the asset factor value is computed and it will be
considered the buy price for that asset at that date.
'Prices' must also contain entries for dates following each
date/asset combination in 'factor', as many more dates as the
maximum value in 'periods'. The asset price after 'period'
dates will be considered the sell price for that asset when
computing 'period' forward returns.
::
-----
      | AAPL | BA | CMG | DAL | LULU |
-----
Date  |      |   |     |     |      |
-----
2014-01-01 | 605.12 | 24.58 | 11.72 | 54.43 | 37.14 |
2014-01-02 | 604.35 | 22.23 | 12.21 | 52.78 | 33.63 |
2014-01-03 | 607.94 | 21.68 | 14.36 | 53.94 | 29.37 |
-----
```

数据来源: 东北证券, Alphalens

第二个输入变量是股票的价格数据, 它是一个二维数据表 (DataFrame), 行是时间, 列是股票代码。图 3 描述说了两件事。第一是输入的价格数据必须是正确的, 必须是按照信号发出进行回测的, 否则会产生前视偏差 (lookahead bias) 或者使用到“未来函数”, 可以加一个缓冲窗口递延交易来解决。例如, 通常按照收盘价的

回测其实就包含了这样的前视偏差，所以递延到第二天开盘价回测。第二是不能有一列或者一行全为空值，就每行每列至少得要一个值，这个值是因子值计算出来时的价格并且也用这个价格买入。当然有买入也要有卖出，同理在这个买入价格给出之后，也要给出一个卖出的价格，根据这两个价格从而计算因子的未来收益率。

下面我们介绍其余七个参数变量，如图 4。

图 4: 数据清洗的其他参数变量

```
groupby : pd.Series - MultiIndex or dict
    Either A MultiIndex Series indexed by date and asset,
    containing the period wise group codes for each asset, or
    a dict of asset to group mappings. If a dict is passed,
    it is assumed that group mappings are unchanged for the
    entire time period of the passed factor data.
by_group : bool
    If True, compute statistics separately for each group.
quantiles : int or sequence[float]
    Number of equal-sized quantile buckets to use in factor bucketing.
    Alternately sequence of quantiles, allowing non-equal-sized buckets
    e.g. [0, .10, .5, .90, 1.] or [.05, .5, .95]
    Only one of 'quantiles' or 'bins' can be not-None
bins : int or sequence[float]
    Number of equal-width (valuewise) bins to use in factor bucketing.
    Alternately sequence of bin edges allowing for non-uniform bin width
    e.g. [-4, -2, -0.5, 0, 10]
    Chooses the buckets to be evenly spaced according to the values
    themselves. Useful when the factor contains discrete values.
    Only one of 'quantiles' or 'bins' can be not-None
periods : sequence[int]
    periods to compute forward returns on.
filter_zscore : int or float
    Sets forward returns greater than X standard deviations
    from the the mean to nan.
    Caution: this outlier filtering incorporates lookahead bias.
groupby_labels : dict
    A dictionary keyed by group code with values corresponding
    to the display name for each group.
```

数据来源：东北证券，Alphalens

第一个参数变量是行业分类 `groupby`，它是可以是一个多重索引的序列，也可以是一个字典（dict）。如果它是一个序列的话，那么其多重索引必须要包含之前的 `factor` 的多重索引；如果它是一个字典的话，那么其股票代码的个数也必须包含 `factor` 的股票代码个数。这都是为了保证每个股票都有行业分类。该值默认是空值（None），也就是默认不按照行业分组计算收益。

接着看第七个参数 `groupby_labels`，它也是一个字典格式，它的意思是把所有的行业贴上标签。比如说，（煤炭）中信对应数字 1，（钢铁）中信对应数字 2，以此类推。之所以要设计这个替代可能是因为显示方便（Python2.x 版本对中文编码显示容易出错，Python3.x 版本则没有这个问题），又或是中文字符运算较慢。

然后是第二个参数 `by_group`，它是布尔型（bool）。如果取值是 True 的话，也就是计算的结果都将按照行业分组之后，再行业内部进行计算。如果取值是 False 的话，则是所有成分股进行计算。分组计算可以理解成是传统的行业中性的做法。

第三个参数 `quantile` 可以和第四个参数 `bins` 放在一起讲，他们都是整型（Int）或者是数值型的列表。我们在进行因子研究的其中重要一步就是对因子进行分组，可以简单的按照数量等分成若干组，也可以认为设定分组的依据。例如，我想知道



中心值和极端值的情况，我可以将 quantile 分成四组，设置成[0,0.05,0.5,0.95,1]。Quantile 是从数据的个数角度进行分类的，bins 是从数据取值的范围进行划分的，类似于直方图（histogram），就不多赘述了。要注意的是，quantile 和 bins 只能设置其中一个，另一个必须是 None。

第五个参数是调仓周期 periods，它是一个不可变长度的列表——元组（tuple）格式的数值，如(1,5,10)，而不可以是数值或列表 list。如果要计算下一期的收益率，则 periods=(1,)，而不可以是 periods=1 或者 periods=[1]。

第六个参数是异常值阈值倍数的设置，filter\_zscore 的取值是一个数值，代表几倍标准差，当未来的收益率超过这个倍数将变成 nan 格式。**需要特别注意的是，这个用法将会导致前视偏差**，因为我们事前将涨的特别多的或者跌的特别多的股票收益率去掉了。但是如果要避免这个问题，只要把 filter\_zscore 的值设置的特别大即可，默认是 20 倍标准差，应该够了。

以上就是 get\_clean\_factor\_and\_forward\_returns 的全部输入变量和参数变量。通过这个函数得到的结果如图 5。它是一个合并的数据表，多重索引。得到 1,5,10 列中对应的数字就是该股票的收益率<sup>1</sup>，factor 列中对应的该股票的因子值，group 列为对应的行业分组，factor\_quantile 为横截面上因子所在哪个分组。得到的结果我们称之为 factor\_data。

图 5: 数据清洗的结果

```
Returns
-----
merged_data : pd.DataFrame - MultiIndex
A MultiIndex Series indexed by date (level 0) and asset (level 1),
containing the values for a single alpha factor, forward returns for
each period, the factor quantile/bin that factor value belongs to, and
(optionally) the group the asset belongs to.
::
-----
|          | 1 | 5 | 10 | factor | group | factor_quantile |
-----
date | asset |
-----
| AAPL | 0.09 | -0.01 | -0.079 | 0.5 | G1 | 3 |
-----
| BA | 0.02 | 0.06 | 0.020 | -1.1 | G2 | 5 |
-----
2014-01-01 | CMG | 0.03 | 0.09 | 0.036 | 1.7 | G2 | 1 |
-----
| DAL | -0.02 | -0.06 | -0.029 | -0.1 | G3 | 5 |
-----
| LULU | -0.03 | 0.05 | -0.009 | 2.7 | G1 | 2 |
-----
::
```

数据来源：东北证券，Alphalens

## 2.2. 因子收益、IC 与换手率

正如我们上述所说，如果以下命令运行成功，则代表已经处理好 95% 的工作了。剩下的工作就是计算因子收益，分析因子 RankIC，以及换手率等结果了。

<sup>1</sup> 下文会提到关于收益率不同标准计算的结果，如绝对收益，超额收益，行业超额收益等。

```
factor_data = get_clean_factor_and_forward_returns(factor,
                                                    prices,
                                                    groupby=None,
                                                    by_group=False,
                                                    quantiles=10,
                                                    bins=None,
                                                    periods=(1, 5, 10),
                                                    filter_zscore=20,
                                                    groupby_labels=None)
```

而实际上，Alphalens 有一个特别强大的功能叫 tears 模块，它会生成一张很大的表图，里面是一张张被称之为撕页（tear sheet）的图片，记录所有与回测相关的结果。这部分就是剩下的 5% 的内容。

```
create_full_tear_sheet(factor_data,
                       long_short=True,
                       group_adjust=False,
                       by_group=False)
```

其中，long\_short 参数含义是是否计算多空组合的收益，group\_adjust 参数含义是是否按照行业调整或者行业中性后的收益，by\_group 参数含义与前面相同，即是否按照行业分组展示。

图 6: 因子回测结果汇总

```
@plotting.customize
def create_full_tear_sheet(factor_data,
                           long_short=True,
                           group_adjust=False,
                           by_group=False):
    """
    Creates a full tear sheet for analysis and evaluating single
    return predicting (alpha) factor.

    Parameters
    -----
    factor_data : pd.DataFrame - MultiIndex
        A MultiIndex DataFrame indexed by date (level 0) and asset (level 1),
        containing the values for a single alpha factor, forward returns for
        each period, the factor quantile/bin that factor value belongs to, and
        (optionally) the group the asset belongs to.
        - See full explanation in utils.get_clean_factor_and_forward_returns
    group_adjust : bool
        Demean forward returns by group before computing IC.
    by_group : bool
        If True, perform calculations, and display graphs separately for
        each group.
    long_short : bool
        Should this computation happen on a long short portfolio? if so, then
        factor values will be demeaned across the factor universe when factor
        weighting the portfolio. Additionally, mean quantile returns will be
        demeaned across the factor universe.
    """

    plotting.plot_quantile_statistics_table(factor_data)
    create_returns_tear_sheet(factor_data,
                              long_short,
                              by_group,
                              set_context=False)
    create_information_tear_sheet(factor_data,
                                  group_adjust,
                                  by_group,
                                  set_context=False)
    create_turnover_tear_sheet(factor_data, set_context=False)
```

数据来源：东北证券，Alphalens

create\_full\_tear\_sheet 只是汇总各个模块的结果，通过图 6 我们便可一目了然：



包括 4 个部分：第一个是绘制因子分组统计结果的表格，第二部分是绘制因子分析的“撕页”，第三部分是绘制因子 IC 分析的“撕页”，最后部分是绘制因子换手率的“撕页”。因此我们可以一一将它拆解。

### 2.2.1. 因子收益

下面我们进入第一部分：`create_returns_tear_sheet` 这一功能。它同样有三个参数，这三个参数跟 `create_full_tear_sheet` 的参数是一样的。它能够很简便的帮助我们做出因子回测的结果。这里我们以市值（size）因子为例展示该因子的按月回测结果。

首先，我们看到的是表 1 当中不同周期下的因子整体收益情况。其中的 alpha 和 beta 是通过回归得到的，回归的自变量是全市场的平均收益，因变量是市值因子的因子加权（多空）组合收益。之后再将 alpha 年化。之后还有不同周期下的第一组和第十组的平均收益，以及他们收益差（spread）的平均值。

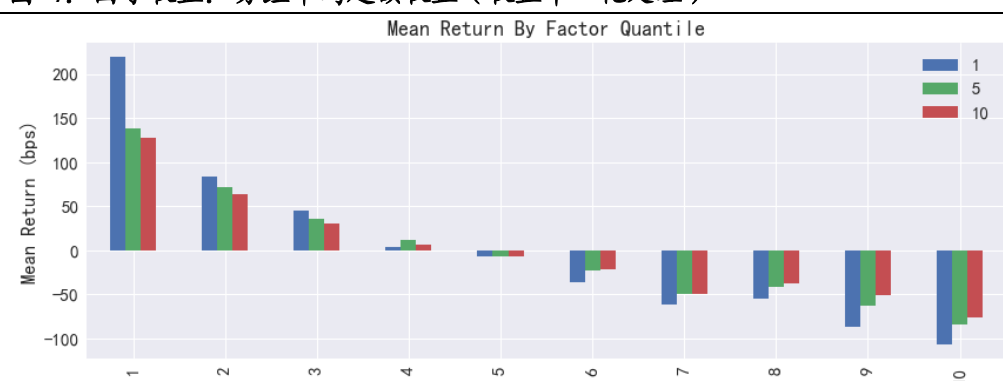
接下来是不同周期下的因子分组收益的表现。我们把市值因子分成十组，图 7 展示的是每组在未来 1、5、10 期的平均超额收益率，即在截面上对所有的股票收益率做 demean 处理，也称为中心化（Zero-centered 或者 Mean-subtraction）。因此，从图 7 我们可以看出市值因子的平均超额收益具有非常好的单调性，即市值越小，平均超额收益率越高；市值越大，平均超额收益率越低。

表 1: 因子收益率分析：不同周期下的收益率结果

Period	1	5	10
Ann. alpha	-0.073	-0.054	-0.072
beta	-0.139	-0.1	-0.017
Mean Period Wise Return Top Quantile (bps)	-103.044	-79.877	-73.296
Mean Period Wise Return Bottom Quantile (bps)	215.919	134.467	124.51
Mean Period Wise Spread (bps)	-295.088	-215.957	-196.528

数据来源：东北证券，Alphalens

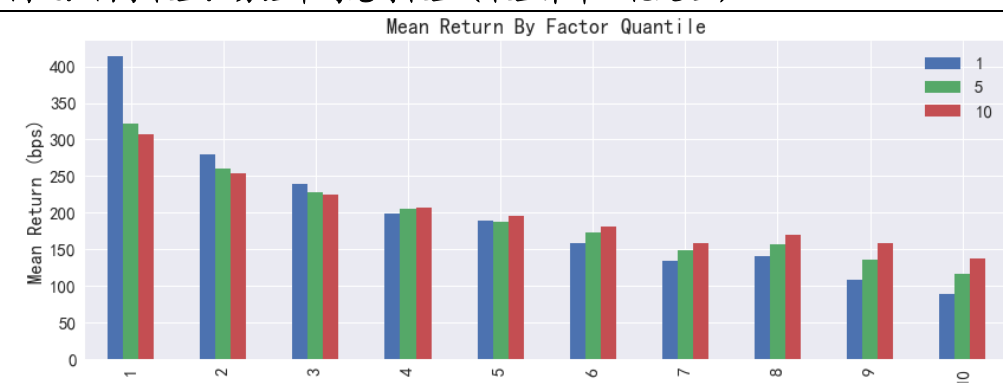
图 7: 因子收益：分组平均超额收益（收益中心化处理）



数据来源：东北证券，Alphalens

图 7 和图 8 的区别是，图 8 使用的是绝对收益率而不是超额收益率。为了图形的美观和避免重复，下面展示的图形均为未来一期收益率的结果。

图 8: 因子收益: 分组平均绝对收益 (收益未中心化处理)



数据来源: 东北证券, Alphalens

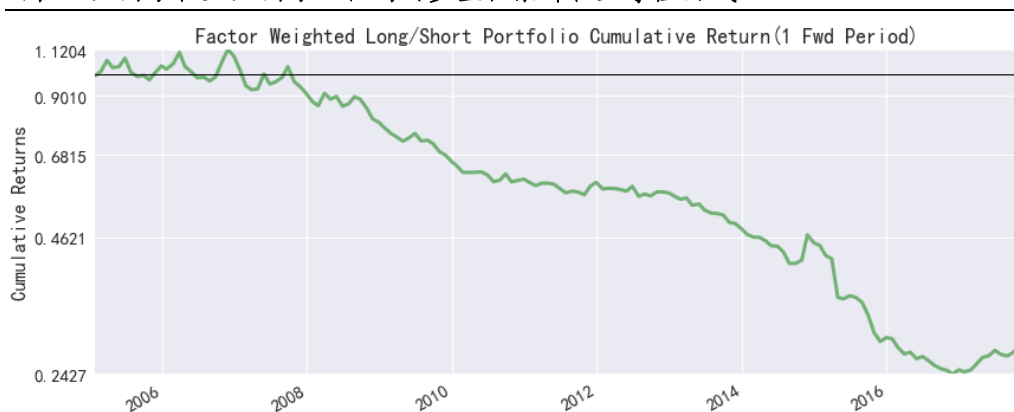
图 9 展示的是小提琴图 (violin plot)。小提琴图由两部分构成, 外层是一个左右对称由核密度估计 (Kernel Density Estimation) 拟合的分布函数。默认情况下, Alphalens 作图只显示[1%,99%]分位数样本内的取值; 内层是一个盒形图 (box plot), 记录了三个四分位值与极值。小提琴图的好处是可以把不同组的超额收益样本分布情况放在一起进行比较, 感兴趣的读者可以进一步查阅相关资料。

图 9: 因子收益: 分组超额收益分布图 (收益中心化处理)



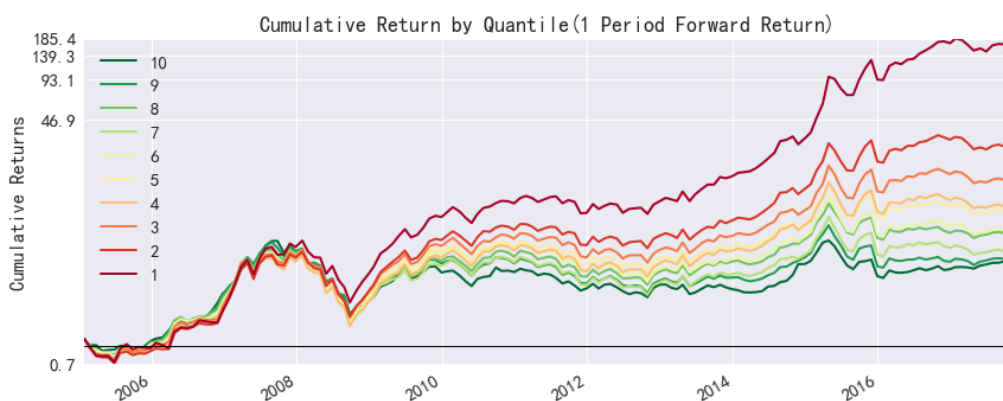
数据来源: 东北证券, Alphalens

图 10: 因子收益: 因子加权的 (多空) 累计收益净值曲线



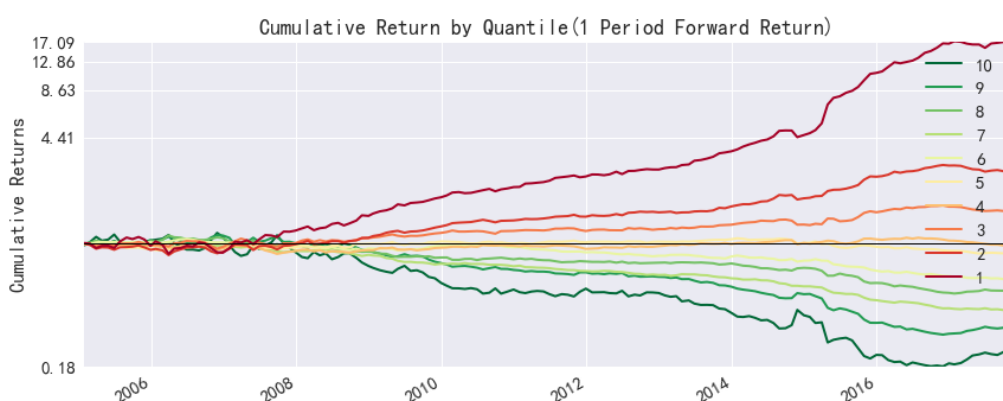
数据来源: 东北证券, Alphalens

图 11: 因子收益: 因子分组累计收益净值曲线



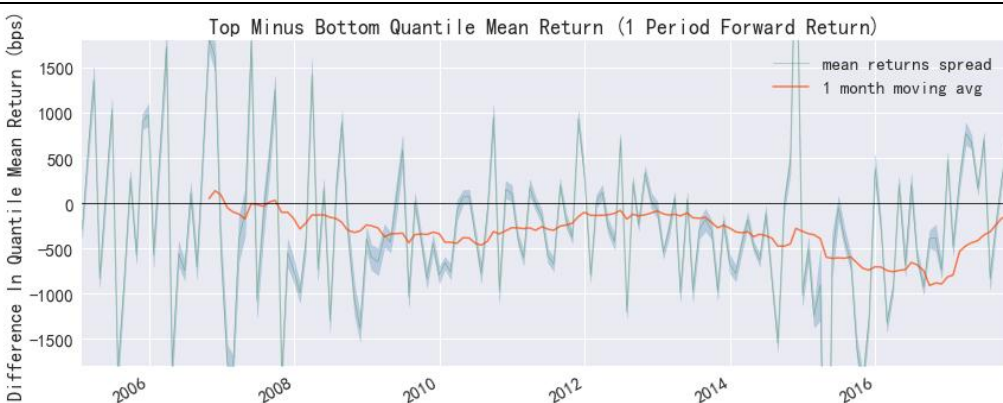
数据来源: 东北证券, Alphalens

图 12: 因子收益: 因子分组超额收益净值曲线



数据来源: 东北证券, Alphalens

图 13: 因子收益: 第十组-第一组平均收益分布情况



数据来源: 东北证券, Alphalens

图 10 展示的是将所有因子按照因子权重进行加权后的收益。可以看到市值因子在 14 年底和 17 年以来都经历了一段较大的回撤。图 11 与图 8 对应, 我们可以看到市值最小的一组表现特别抢眼<sup>2</sup>。图 12 与图 7 对应, 从中可以看到市值因子

<sup>2</sup> 由于这里只是为了展示作图结果, 选取的是全 A 样本股, 也未作其他限制和处理。

的单调性非常好。图 13 展示了第十组的平均收益减去第一组的平均收益的分布情况，其中“一个月移动平均”是针对日度回测而言的，因此如果用月度回测，表示的是过去 20 个月的移动平均。

以上就是 create\_returns\_tear\_sheet 这一函数的全部内容，涵盖了与因子收益相关的各类图形。

### 2.2.2. 因子 IC 分析

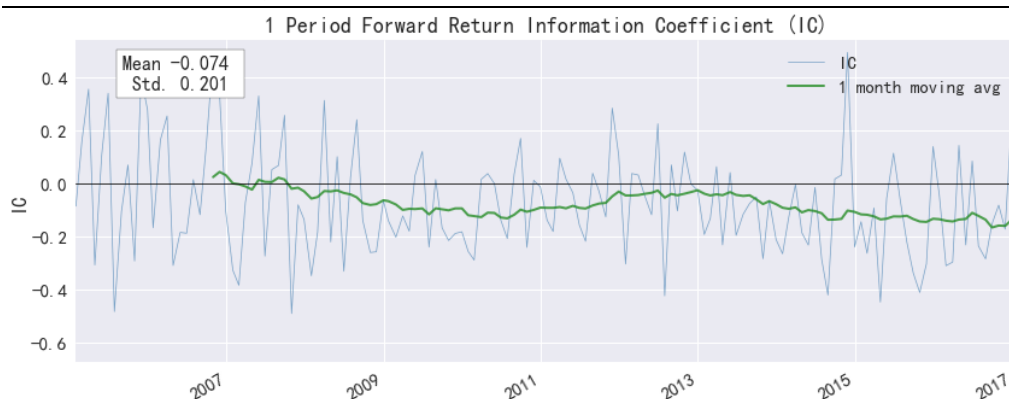
下面我们进入第二部分：create\_information\_tear\_sheet 这一功能。它同样有三个参数，这三个参数跟 create\_full\_tear\_sheet 的参数是一样的。这里我们继续以市值因子为例展示该因子的按月回测结果。

表 2: 因子 IC 分析：不同周期下的 IC 统计结果

Period	1	5	10
IC Mean	-0.074	-0.129	-0.168
IC Std.	0.201	0.212	0.208
t-stat(IC)	-4.424	-7.359	-9.722
p-value(IC)	0.000	0.000	0.000
IC Skew	0.486	0.404	0.244
IC Kurtosis	-0.014	-0.320	-0.539
Ann. IR	-1.304	-2.169	-2.866

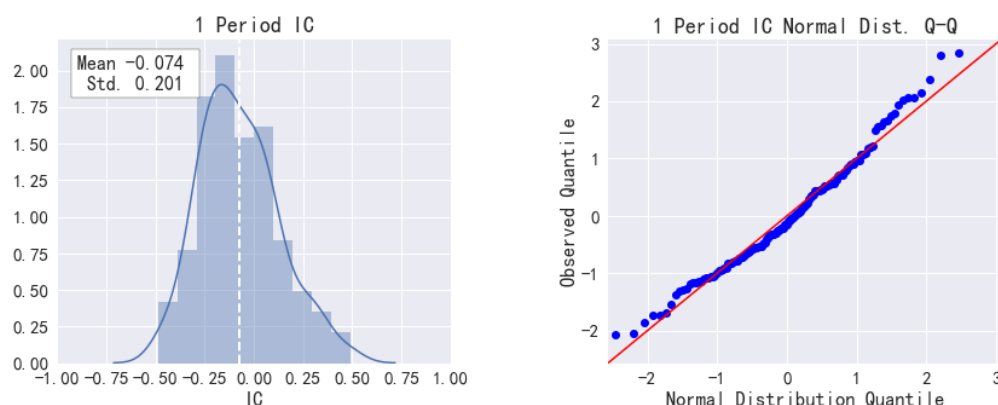
数据来源：东北证券，Alphalens

图 14: 因子 IC 分析：IC 时间序列图



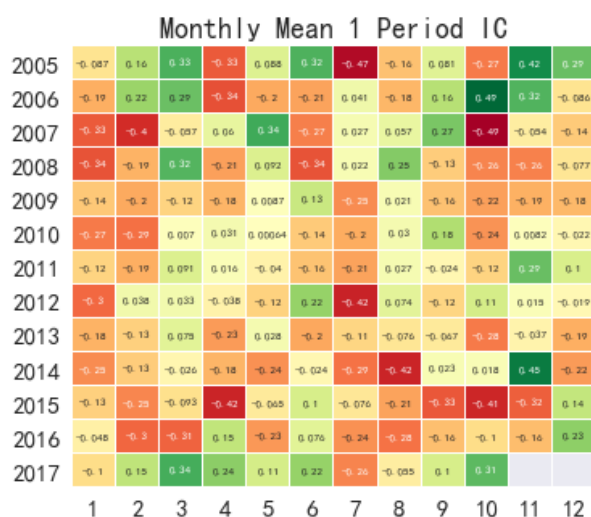
数据来源：东北证券，Alphalens

图 15: 因子 IC 分析: IC 的分布图和 QQ 图



数据来源: 东北证券, Alphalens

图 16: 因子 IC 分析: IC 热力图



数据来源: 东北证券, Alphalens

### 2.2.3. 因子换手率分析

下面我们进入第三部分: create\_turnover\_tear\_sheet 这一功能。这里我们继续以市值因子为例展示该因子的按月回测结果。这里的换手率指的是单边的成分股的调整个数占比。

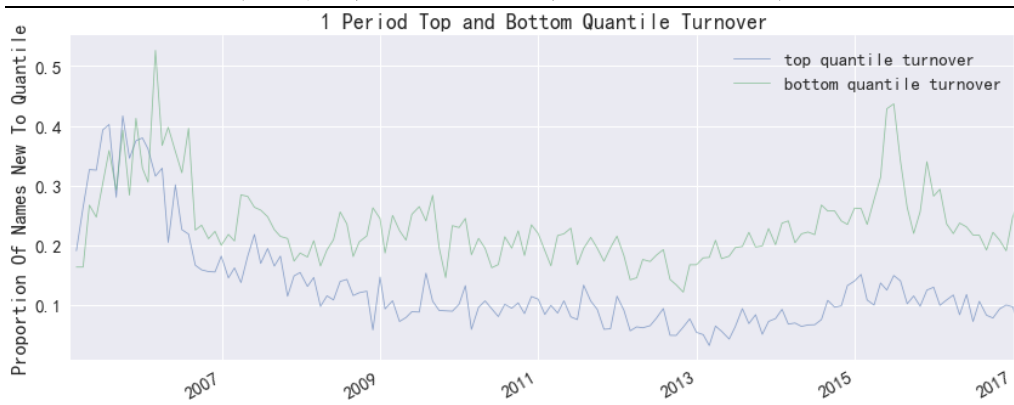
表 3: 因子换手率分析: 不同周期下的分组换手率统计结果

Period	1	5	10
Quantile 1 Mean Turnover	0.234	0.351	0.490
Quantile 2 Mean Turnover	0.365	0.566	0.698
Quantile 3 Mean Turnover	0.411	0.629	0.753
Quantile 4 Mean Turnover	0.427	0.659	0.773
Quantile 5 Mean Turnover	0.425	0.664	0.774
Quantile 6 Mean Turnover	0.405	0.644	0.761
Quantile 7 Mean Turnover	0.364	0.592	0.718
Quantile 8 Mean Turnover	0.301	0.506	0.647

Quantile 9 Mean Turnover	0.224	0.373	0.517
Quantile 10 Mean Turnover	0.130	0.185	0.279

数据来源：东北证券，Alphalens

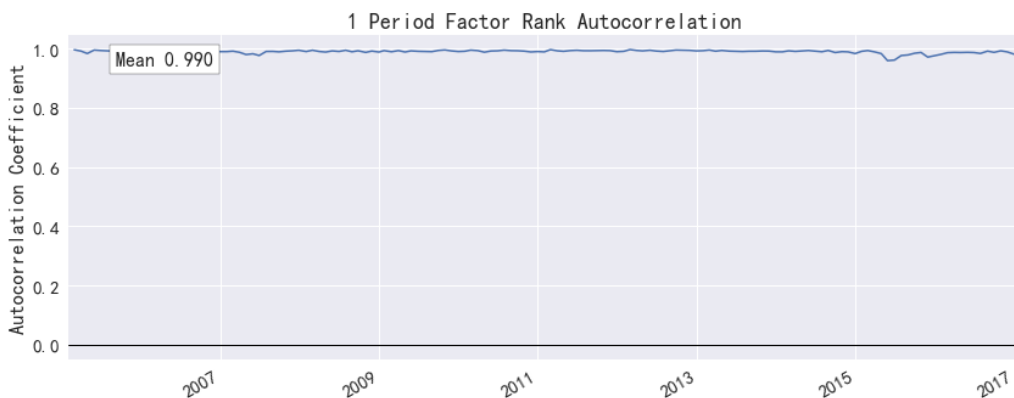
图 17: 因子换手率分析: 第一组 (绿) 和第十组 (蓝) 换手率



数据来源：东北证券，Alphalens

图 18 展示的是因子当期因子排序与下期因子排序的相关性序列图。如果说图 17 只刻画了第一组和第十组的换手率，那么图 18 刻画的是整个横截面的换手率，截面换手率越低，因子的自相关性就越高。

图 18: 因子换手率分析: 因子排序的自相关图



数据来源：东北证券，Alphalens

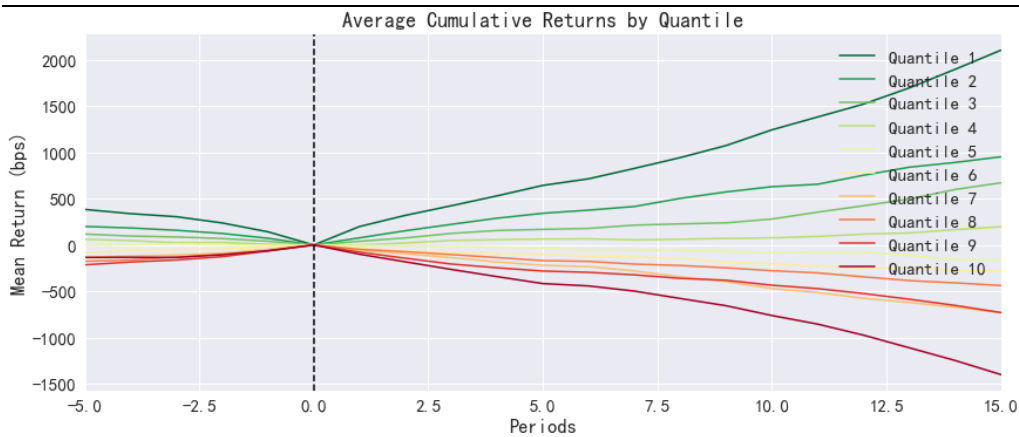
### 2.3. 其他功能: 事件研究

除了上述介绍的三个函数外，还有第四个函数: `create_event_returns_tear_sheet`，顾名思义，是用于事件收益研究。这里我们继续以市值因子为例展示该事件收益的研究结果。

从事件研究的结果来看，市值因子的买入后的超额收益是非常显著的，而且随时间的增加超额收益是不断增加的，单调性也非常好。

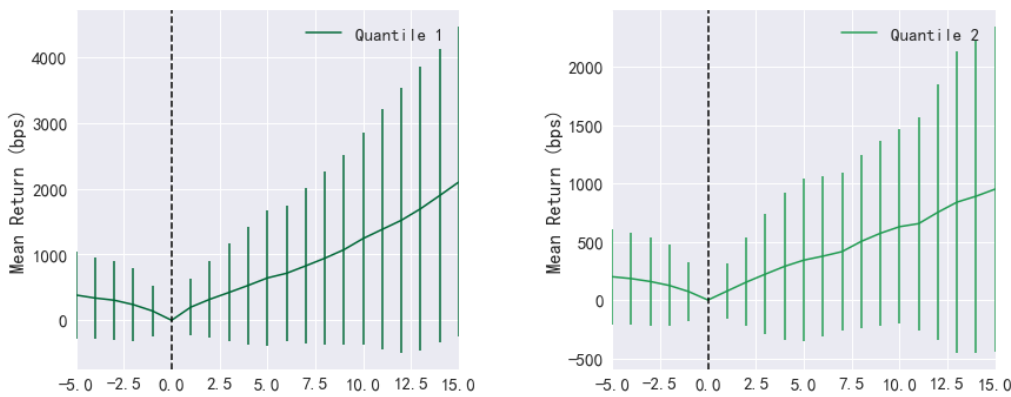


图 19: 事件研究: 因子分组平均超额收益随时间的关系



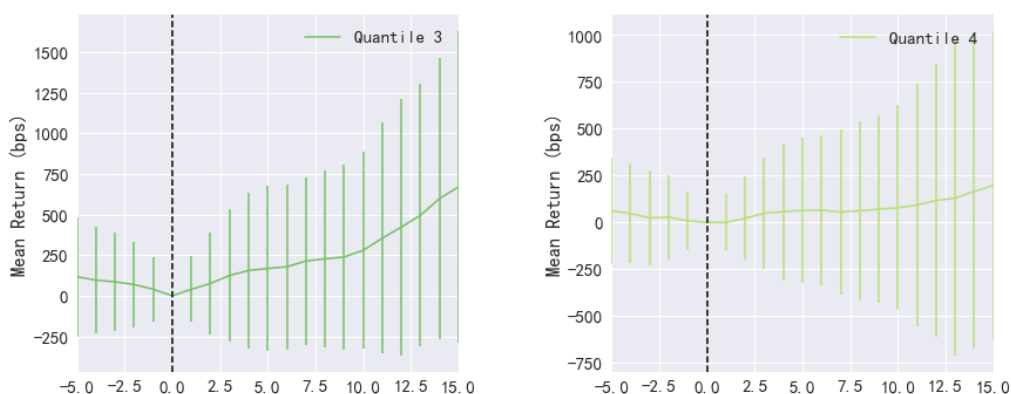
数据来源: 东北证券, Alphalens

图 20: 事件研究: 第一、二组超额收益分布随时间的关系



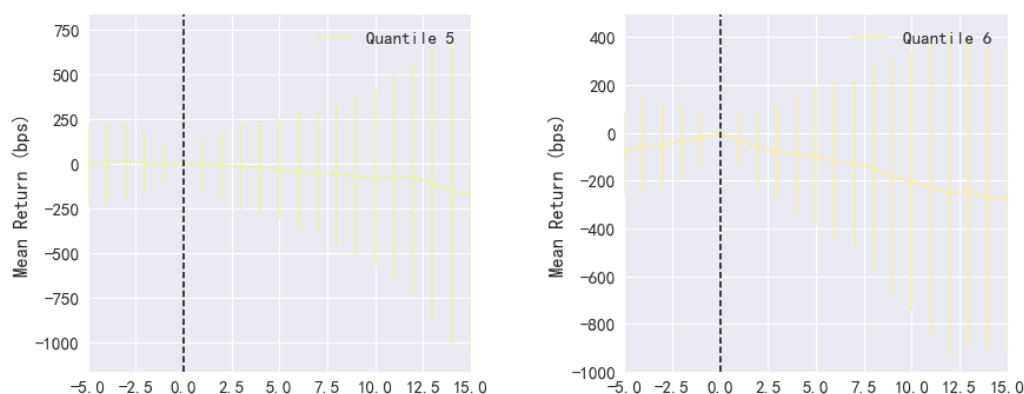
数据来源: 东北证券, Alphalens

图 21: 事件研究: 第三、四组超额收益分布随时间的关系



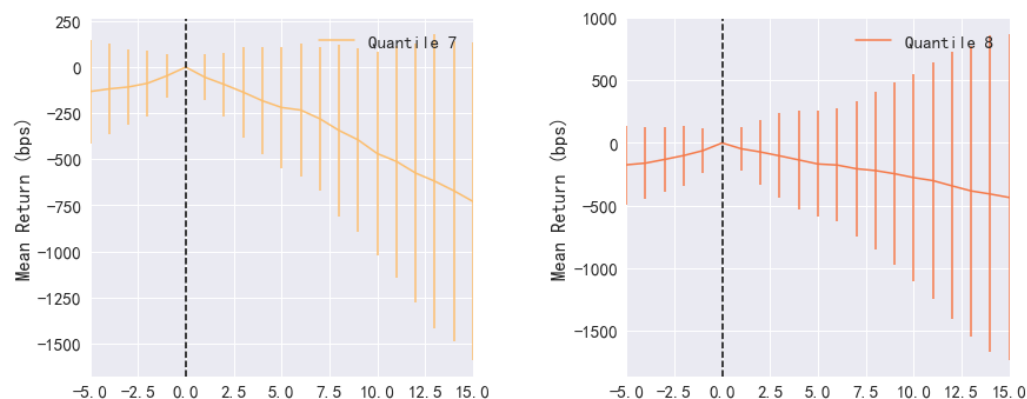
数据来源: 东北证券, Alphalens

图 22: 事件研究: 第五、六组超额收益分布随时间的关系



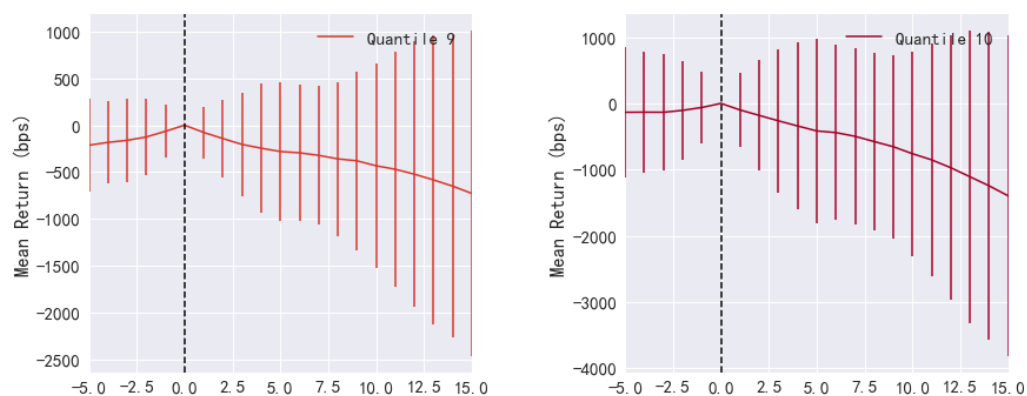
数据来源: 东北证券, Alphasens

图 23: 事件研究: 第七、八组超额收益分布随时间的关系



数据来源: 东北证券, Alphasens

图 24: 事件研究: 第九、十组超额收益分布随时间的关系



数据来源: 东北证券, Alphasens

### 3. 小结

本文主要是希望通过介绍国外开源的工具包 alphasens 为投资者在因子回测上

提供便利。正如本文在一开始所说，只要处理好因子值、价格与行业的关系之后，便已经做好了 95% 的工作（实际上，例如剔除次新股、ST、PT 与涨停跌停等股票，包括沪深 300 选股、中证 500 选股，这些工作都包含在这 95% 的范围），剩下的部分 alphalens 均能帮我们高效的处理，这也是其优势所在。Alphalens 的回测功能虽然强大，但是离实际回测还是有一定的差距。例如，因子的调仓成本未考虑，未剔除涨停买进和跌停卖出。但瑕不掩瑜，投资者可以自行根据需要再原有的模板上进行改进，以适用于真实环境下的回测。

**分析师简介:**

肖承志: 金融工程研究助理, 同济大学应用数学硕士, 2016年加入东北证券研究所。

**重要声明**

本报告由东北证券股份有限公司(以下称“本公司”)制作并仅向本公司客户发布, 本公司不会因任何机构或个人接收到本报告而视其为本公司的当然客户。

本公司具有中国证监会核准的证券投资咨询业务资格。

本报告中的信息均来源于公开资料, 本公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅反映本公司于发布本报告当日的判断, 不保证所包含的内容和意见不发生变化。

本报告仅供参考, 并不构成对所述证券买卖的出价或征价。在任何情况下, 本报告中的信息或所表述的意见均不构成对任何人的证券买卖建议。本公司及其雇员不承诺投资者一定获利, 不与投资者分享投资收益, 在任何情况下, 我公司及其雇员对任何人使用本报告及其内容所引发的任何直接或间接损失概不负责。

本公司或其关联机构可能会持有本报告中所涉及的公司所发行的证券头寸并进行交易, 并在法律许可的情况下不进行披露; 可能为这些公司提供或争取提供投资银行业务、财务顾问等相关服务。

本报告版权归本公司所有。未经本公司书面许可, 任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的, 须在本公司允许的范围内使用, 并注明本报告的发布人和发布日期, 提示使用本报告的风险。

本报告及相关服务属于中风险(R3)等级金融产品及服务, 包括但不限于A股股票、B股股票、股票型或混合型公募基金、AA级别信用债或ABS、创新层挂牌公司股票、股票期权备兑开仓业务、股票期权保护性认沽开仓业务、银行非保本型理财产品及相关服务。

若本公司客户(以下称“该客户”)向第三方发送本报告, 则由该客户独自为此发送行为负责。提醒通过此途径获得本报告的投资者注意, 本公司不对通过此种途径获得本报告所引起的任何损失承担任何责任。

**分析师声明**

作者具有中国证券业协会授予的证券投资咨询执业资格, 并在中国证券业协会注册登记为证券分析师。本报告遵循合规、客观、专业、审慎的制作原则, 所采用数据、资料的来源合法合规, 文字阐述反映了作者的真实观点, 报告结论未受任何第三方的授意或影响, 特此声明。

**投资评级说明**

股票 投资 评级 说明	买入	未来 6 个月内, 股价涨幅超越市场基准 15%以上。
	增持	未来 6 个月内, 股价涨幅超越市场基准 5%至 15%之间。
	中性	未来 6 个月内, 股价涨幅介于市场基准-5%至 5%之间。
	减持	在未来 6 个月内, 股价涨幅落后市场基准 5%至 15%之间。
	卖出	未来 6 个月内, 股价涨幅落后市场基准 15%以上。
行业 投资 评级 说明	优于大势	未来 6 个月内, 行业指数的收益超越市场平均收益。
	同步大势	未来 6 个月内, 行业指数的收益与市场平均收益持平。
	落后大势	未来 6 个月内, 行业指数的收益落后于市场平均收益。

**东北证券股份有限公司**

网址: <http://www.nesc.cn> 电话: 400-600-0686

地址	邮编
中国吉林省长春市生态大街 6666 号	130119
中国北京市西城区锦什坊街 28 号恒奥中心 D 座	100033
中国上海市浦东新区杨高南路 729 号	200127
中国深圳市南山区大冲商务中心 1 栋 2 号楼 24D	518000

**机构销售联系方式**

姓名	办公电话	手机	邮箱
<b>华东地区机构销售</b>			
袁颖 (总监)	021-20361100	13621693507	yuanying@nesc.cn
王一	021-20361267	13761867866	wangyi@nesc.cn
王博	021-20361111	13761500624	wangbo@nesc.cn
李寅	021-20361229	15221688595	liyin@nesc.cn
杨涛	021-20361106	18601722659	yangtao@nesc.cn
刘睿	021-20361112	13120728563	liurui@nesc.cn
阮敏	021-20361121	13564972909	ruanmin@nesc.cn
李喆莹	021-20361101	13641900351	lizy@nesc.cn
李流奇	021-20361102	13120758587	lilq@nesc.cn
齐健	021-20361258	18221628116	qijian@nesc.cn
<b>华北地区机构销售</b>			
李航 (总监)	010-58034553	18515018255	lihang@nesc.cn
殷璐璐	010-58034557	18501954588	yinlulu@nesc.cn
温中朝	010-58034555	13701194494	wenzc@nesc.cn
夏一然	010-58034561	15811140346	xiayiran@nesc.cn
曾彦戈	010-58034563	18501944669	zengyg@nesc.cn
金婷	010-58034559	13439651862	jinting@nesc.cn
<b>华南地区机构销售</b>			
邱晓星 (总监)	0755-33975865	18664579712	qiuxx@nesc.cn
刘璇	0755-33975865	18938029743	liu-xuan@nesc.cn
王之明	0755-33975865	15999555916	wangzm@nesc.cn
林钰乔	0755-33975865	13662669201	linyq@nesc.cn
周逸群	0755-33975865	18682251183	zhouyq@nesc.cn