yumingming@xyzq.com.cn
S0190514100003

gongmin@xyzq.com.cn
S0190119020038

2019 10 17

Wind A
AdaBoost

1.
2.
3.
4.

- 1.
2. 3. 4.
5.

Wind A

- 51 AdaBoost
Wind A 1 -1 51
500

AdaBoost 2014 10 27 2019 8 30
41.31% 1.41
24.67% 0.98
7.66% 0.26

- AdaBoost
1. 2. 3.
4.

- 2015 6 1 2019 8 30 15.90%
0.72
13.16% 0.70 Wind A -11.18%

1 ....- 4 -
1.1 ....- 4 -
1.2 ....- 6 -
1.3 ....- 7 -
1.3.1 ....- 7 -
1.3.2 ....- 8 -
1.4 ....- 9 -
1.4.1 ....- 9 -
1.4.2 ....- 11 -
2 AdaBoost ....- 12 -
2.1 AdaBoost ....- 12 -
2.2 AdaBoost ....- 13 -
2.3 AdaBoost ....- 15 -
3 ....- 17 -
3.1 ....- 17 -
3.2 Wind A ....- 17 -
3.3 AdaBoost ....- 18 -
3.4 ....- 19 -
4 ....- 21 -

....- 4 -
....- 6 -
....- 7 -
....- 8 -
....- 8 -
....- 9 -
....- 9 -
....- 10 -
....- 10 -
..- 11 -
..- 11 -
....- 12 -
....- 13 -
....- 14 -
....- 14 -
....- 15 -
.....- 15 -
.....- 16 -
.- 16 -
.- 16 -
.....- 18 -
.....- 18 -
-
19 -
-

19 -
..- 20 -
..- 20 -
.....- 20 -

p_k

$$Gini(\mathrm{p}) = \sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k=1}^{K} p_k^2$$

C_k

$$Gini(\mathrm{D}) = 1 - \sum_{k=1}^{K} (\frac{C_k}{D})^2$$

D_1 D_2

$$Gini(\mathrm{D},\mathrm{A}) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2)$$

D、D_1 D_2

1. D

2. D

3. A a D

4. A a
D_1 D_2

5. 1-4

X

X

sklearn Entropy

Information Gain

p_k

$$Entropy(D) = -\sum_{k=1}^{K} p_k log_2(p_k)$$

$$Entropy(D) = -p_k log_2(p_k) - (1 - p_k) log_2(1 - p_k)$$

D A

A

$$Gain(A) = Entropy(D) - Entropy(D, A)$$

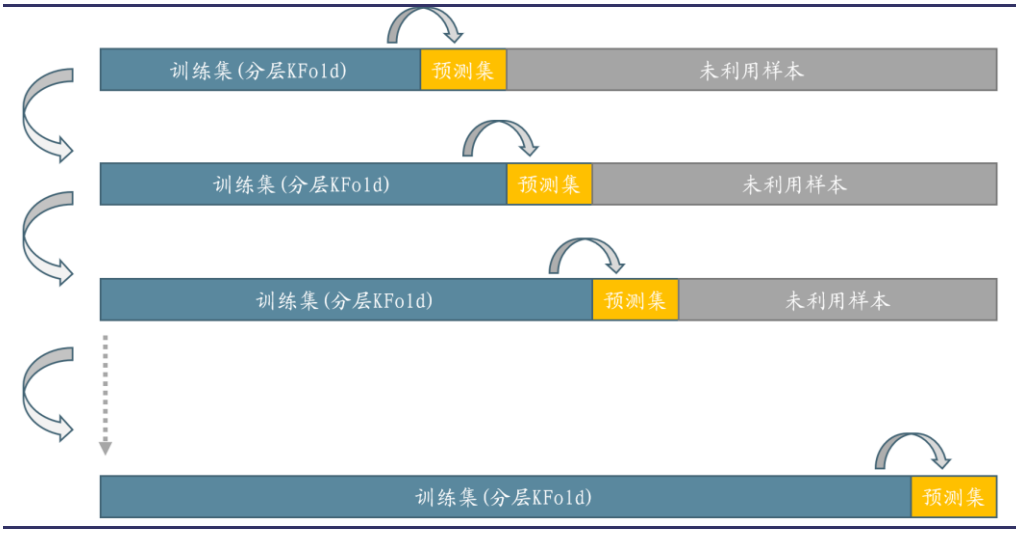
D、D_1 D_2

1.2

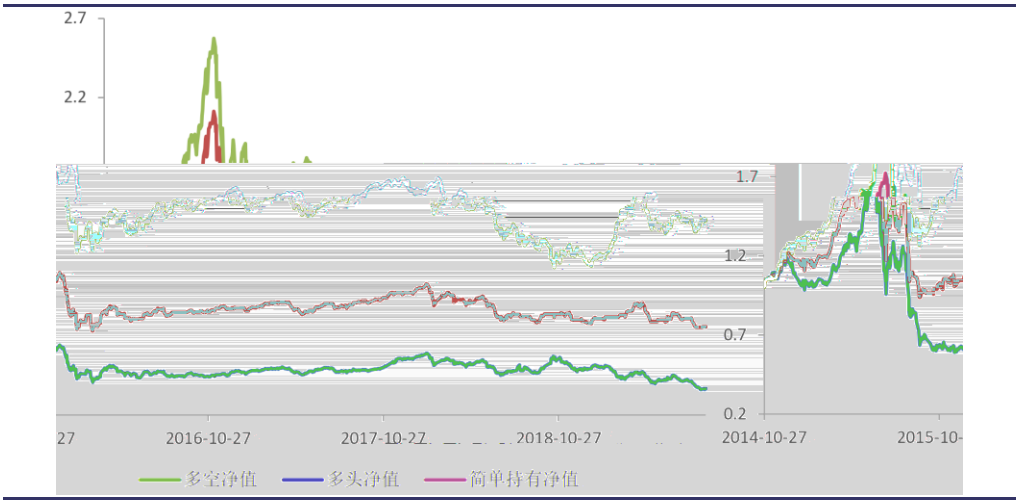
Wind,

1.3

1.3.1

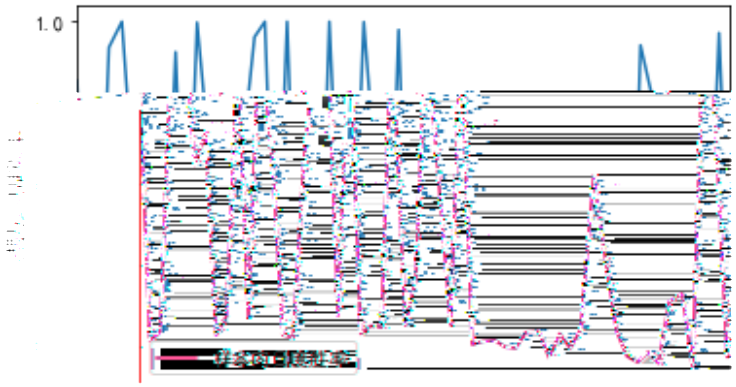

训练集(分层KFold)
预测集
未利用样本
训练集(分层KFold)
预测集
未利用样本
训练集(分层KFold)
预测集
未利用样本
训练集(分层KFold)
预测集

1.3.2


2.7
2.2
1.7
1.2
0.7
0.2
27
2016-10-27
2017-10-27
2018-10-27
2014-10-27
2015-10-
多空净值
多头净值
简单持有净值

Wind,

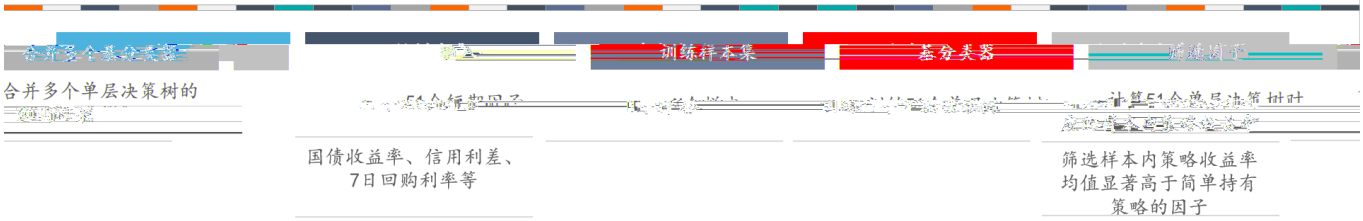
Wind,

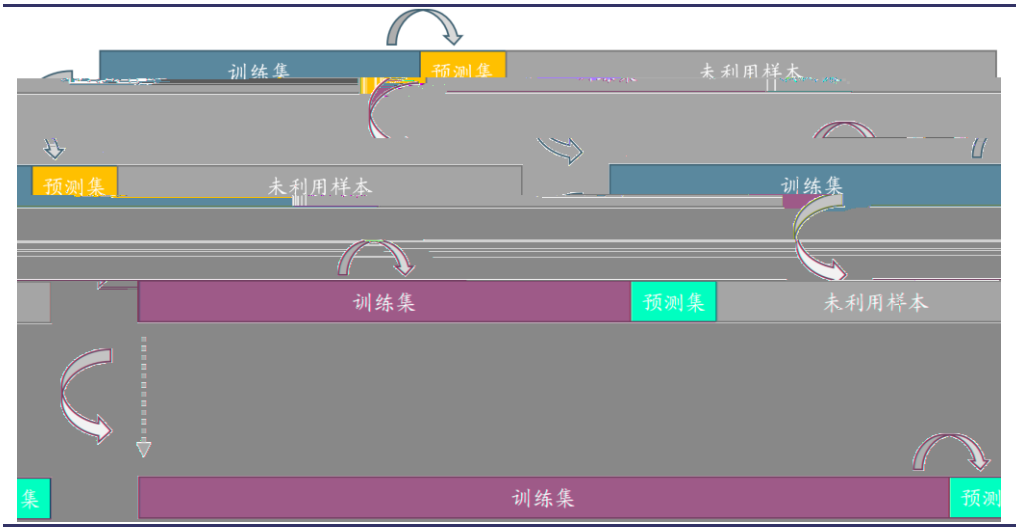


Wind,

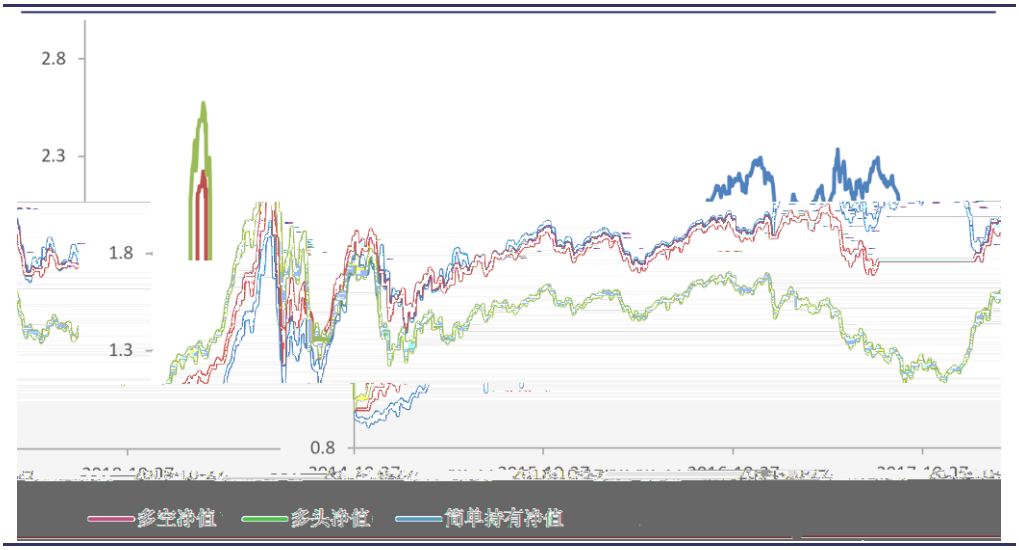
1.4

1.4.1


合并多个单层决策树的
训练样本集
基分类器
国债收益率、信用利差、
7日回购利率等
筛选样本内策略收益率
均值显著高于简单持有
策略的因子

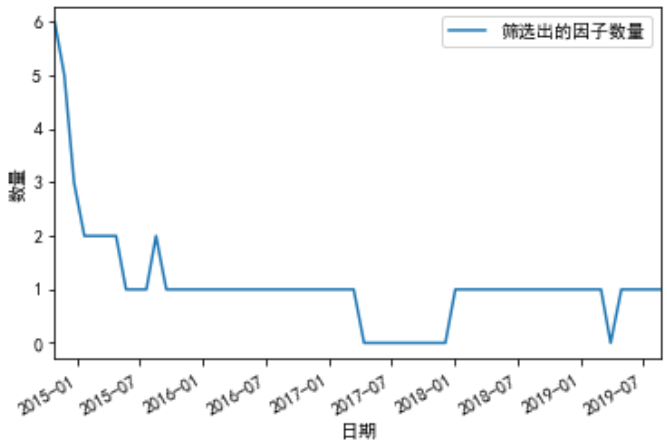

训练集
预测集
未利用样本
预测集
未利用样本
训练集
训练集
预测集
未利用样本
集
训练集
预测

1.4.2



Wind,

Wind,



Wind,

2 AdaBoost

2.1

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

1. $w_i = \frac{1}{N}, \mathrm{i} = 1, 2, 3, \dots, \mathrm{N}$。

2. $m = 1 \quad M$

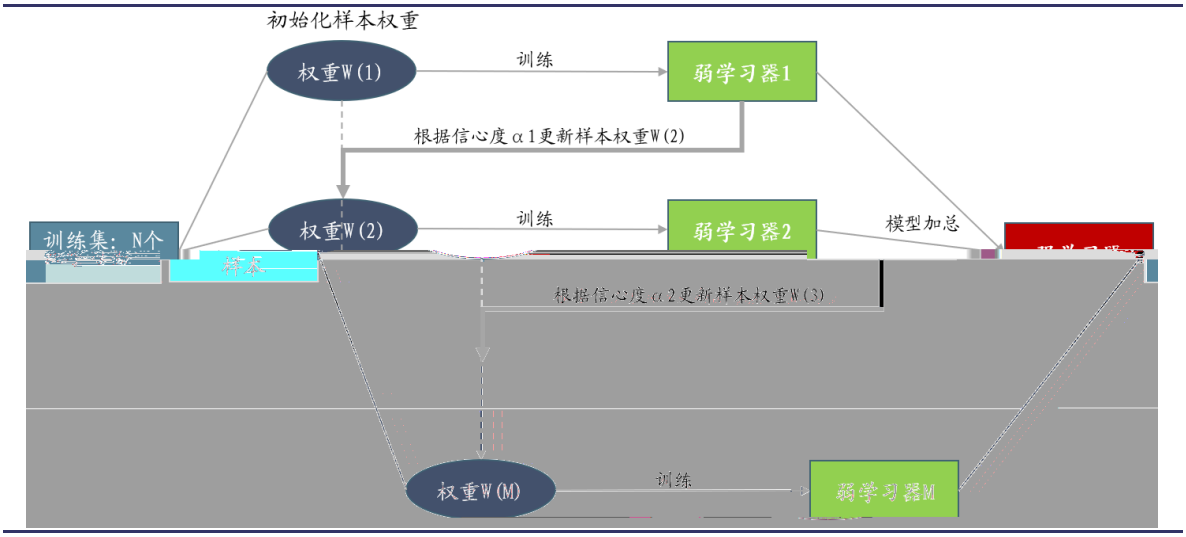
w_i $G_m(x)$

$$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i}$$

$$\alpha_m = \ln\left(\frac{1-err_m}{err_m}\right)$$

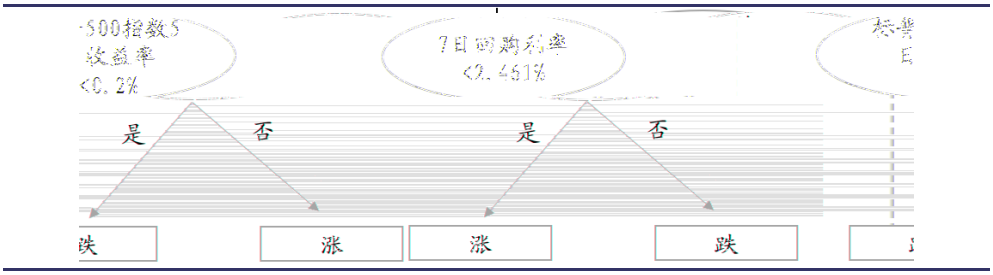
$$w_i \leftarrow w_i \cdot \exp\left[\alpha_m \cdot I\left(y_i \neq G_m(x_i)\right)\right], i = 1, 2, 3, \ldots, N$$

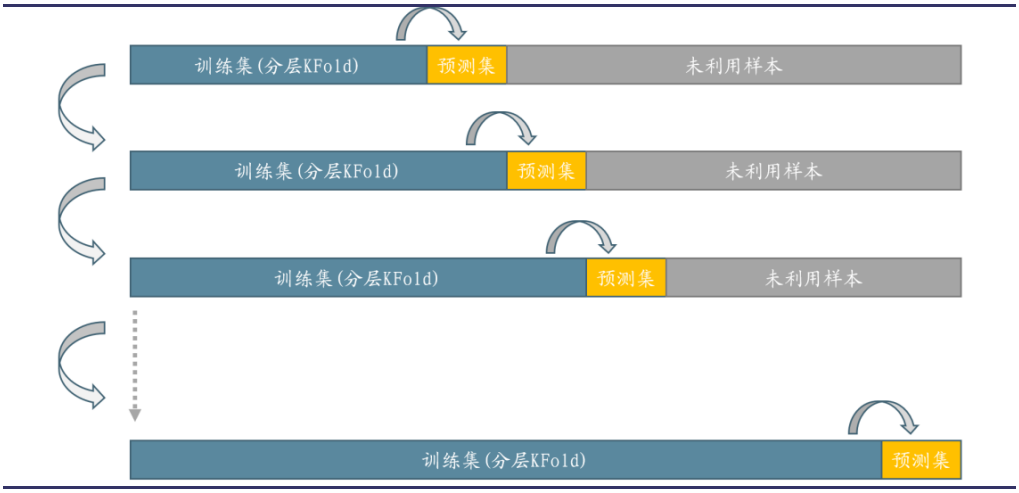
$$G(x) = sign[\sum_{m=1}^{M} \alpha_m G_m(x)]$$



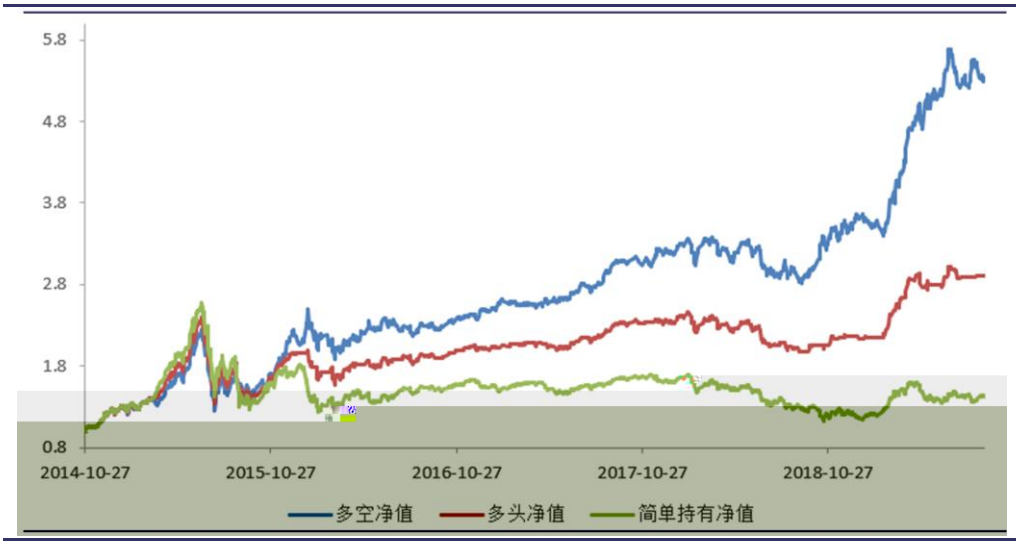
2.2


输入
训练样本集
弱分类器
调整权重
强分类器
51个短期因子
国债收益率、信用利差、
7日回购利率等。
选择预测准确度最高的
提高分类错误的样本权
重


是
否
是
否
涨
涨
跌

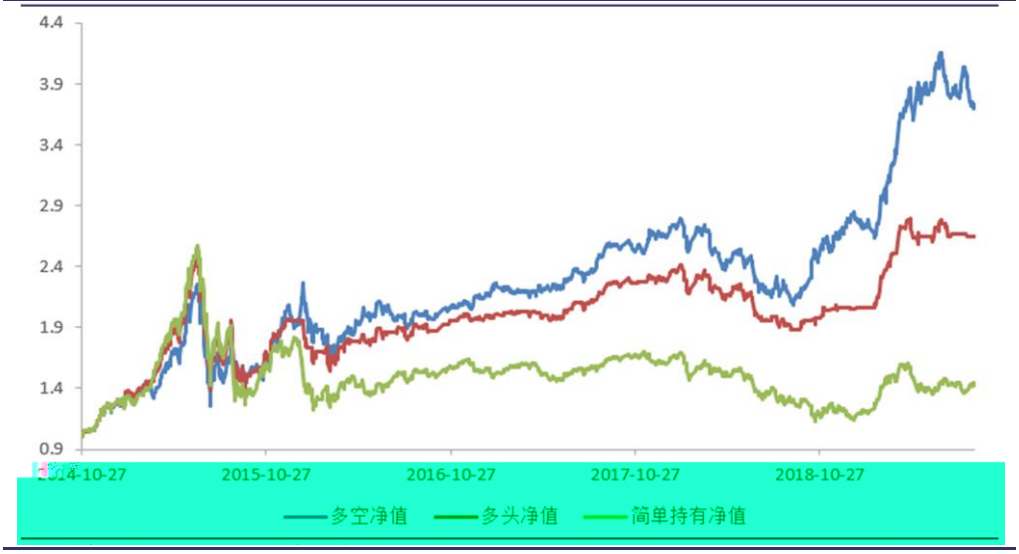

训练集(分层KFold)
预测集
未利用样本
训练集(分层KFold)
预测集
未利用样本
训练集(分层KFold)
预测集
未利用样本
训练集(分层KFold)
预测集

2.3


5.8
4.8
3.8
2.8
1.8
0.8
2014-10-27
2015-10-27
2016-10-27
2017-10-27
2018-10-27
多空净值
多头净值
简单持有净值

Wind,

Wind,



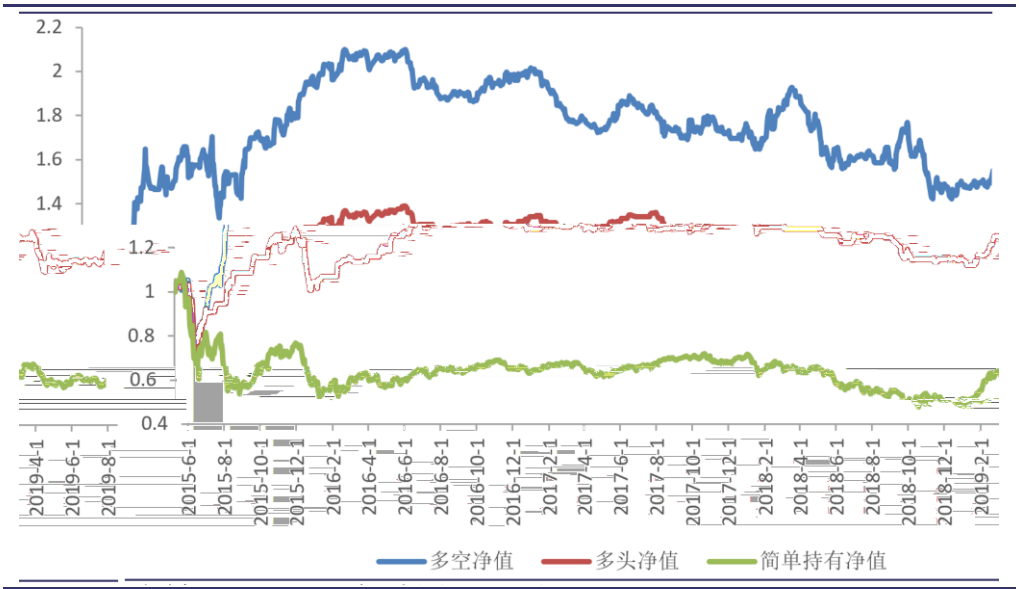
Wind,

Wind,

3

3.1

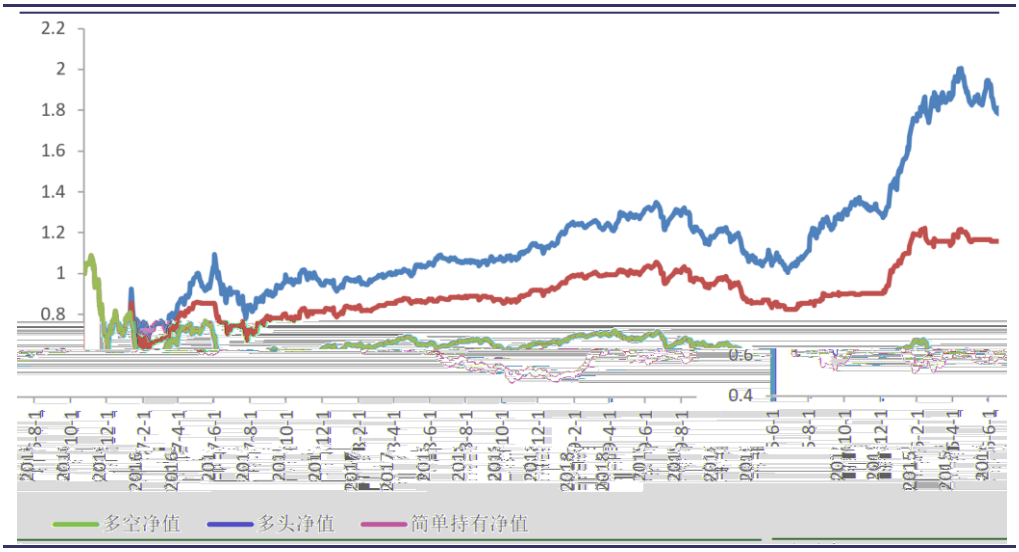
3.2


2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
2019-4-1
2019-6-1
2019-8-1
2015-6-1
2015-8-1
2015-10-1
2015-12-1
2016-2-1
2016-4-1
2016-6-1
2016-8-1
2016-10-1
2016-12-1
2017-2-1
2017-4-1
2017-6-1
2017-8-1
2017-10-1
2017-12-1
2018-2-1
2018-4-1
2018-6-1
2018-8-1
2018-10-1
2018-12-1
2019-2-1
多空净值
多头净值
简单持有净值

Wind,

Wind,

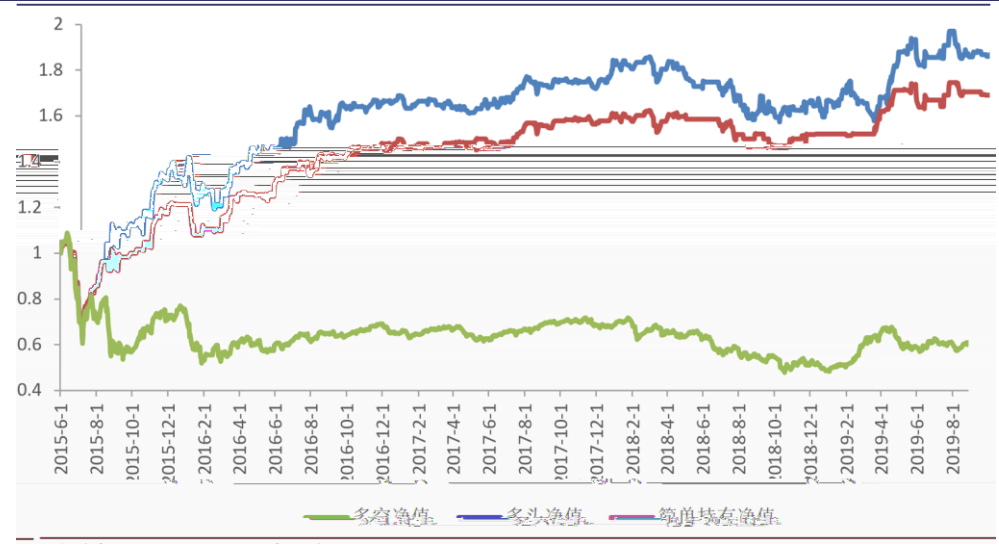
3.3



Wind,

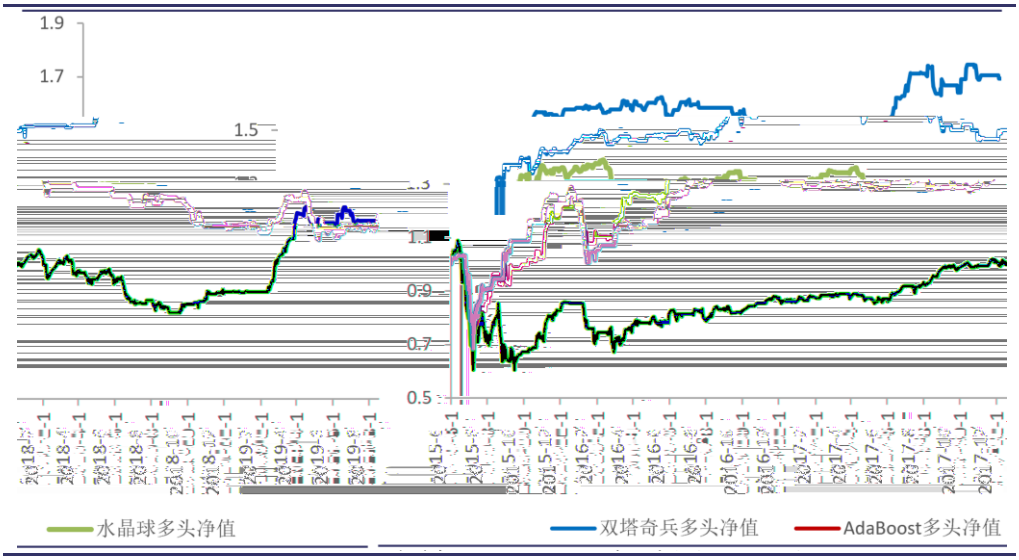
Wind,

3.4



Wind,

Wind,



Wind,

4

“ ”