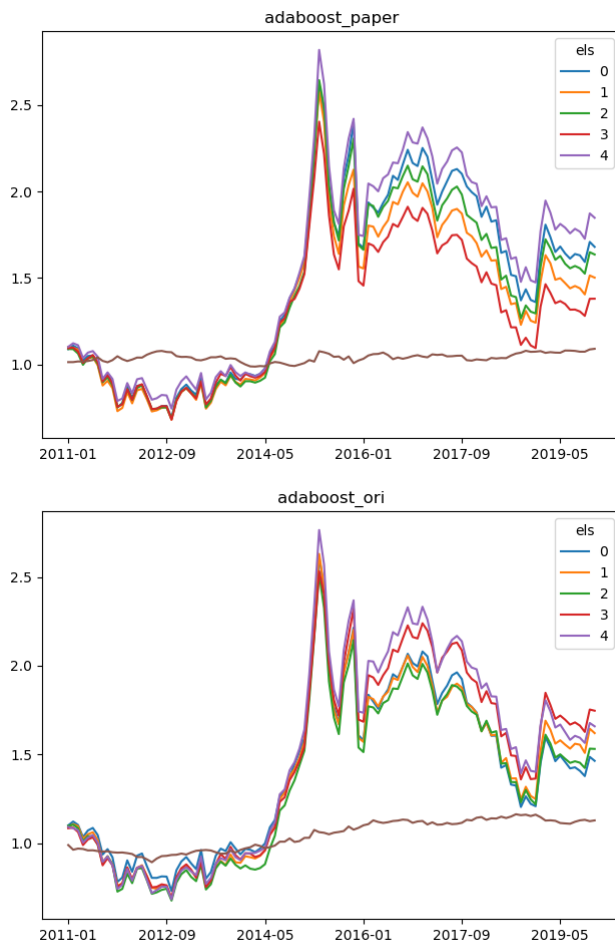


研报说明

完全按照研报算法，结果基本坏掉。无法使用。检查算法没有问题的情况下，完全按照传统的adaboost实验，效果也坏掉。由此，我们基本判断，adaboost方法不适用于优矿的因子数据集。



据此，我们尝试使用更为高级的树方法，GBDT对优矿因子重新建模。由于GBDT运行速度慢，因子过多会引起过拟合情况。因此，依照之前研报（ela）的因子筛选方法对因进行筛选。

由于，

其具体步骤如下：

1. 使用ICIR指标对因子先进行筛选。
2. 使用本篇研报提出的adaboost方法筛选因子。
3. 最后使用筛选的因子，使用GBDT模型建模。

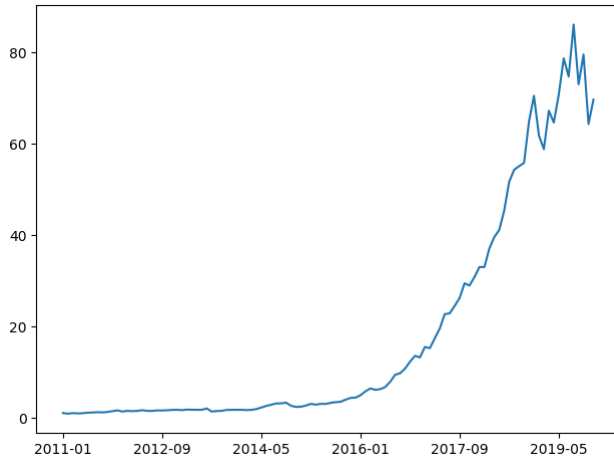
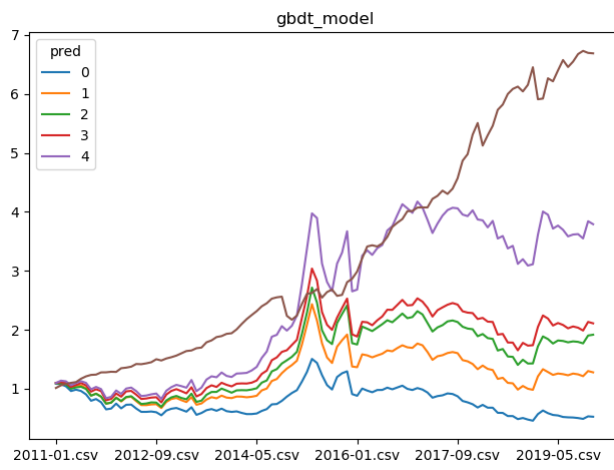
解释：

1. 该篇论文提出的方法是否可以因子筛选？

可以，这篇论文提出的方法非常适合因子筛选，原因是，根据某一次迭代选择一个因子，该次对样本加权后，与选择因子相关性高的因子在后续的选择中反而不会被选中（该特点在研报中有提及）。该模型对加法模型的因子选择有效。且其运行速度较复杂模型快。

2. 最后主模型选择GBDT的原因：

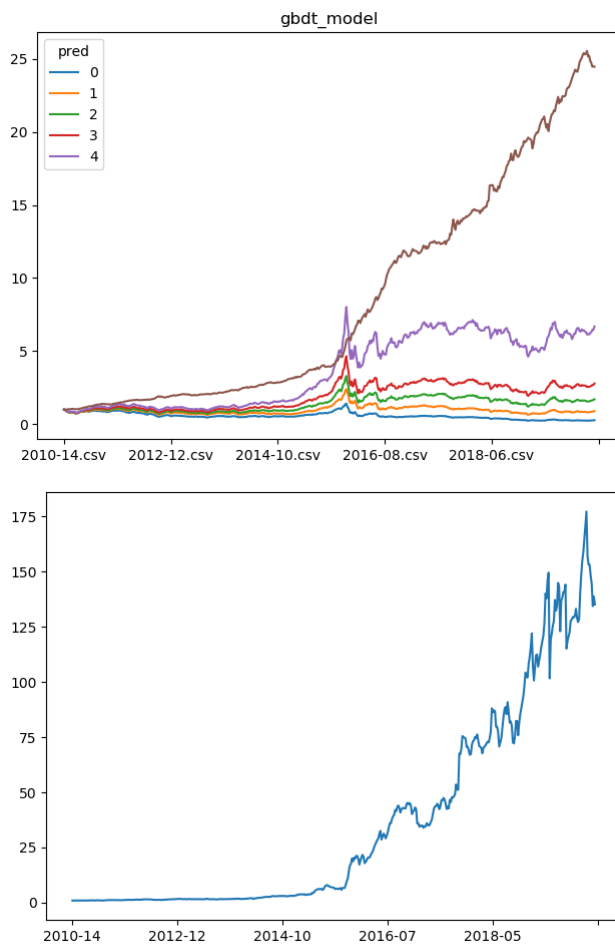
Adaboost模型在一定程度上可以理解为加法模型。所以，通过加法模型选择的因子更加适用于基于加法模型的算法。由此，我们选择高级的加法树模型GBDT作为主模型。



图片的说明：4号标识，即紫色曲线为分组的多头组合收益。棕色曲线标识多空组合结果。

最后一张图片表示做多30只评分最高股票与做空30只评分最低股票的组合收益率，千分之五手续费。

在week上的表现



由于手续费的存在，按月的低频交易波动更小。

代码说明：

ada_month.py 为按照研报算法的计算程序。

ada_ori_month.py 为按照原始的adaboost的计算程序。

gbdt_update.py 为以研报中的adaboost为删选步骤的计算程序。

gbdt_week.py 为以星期为频率计算的程序。

alg文件夹中的adaboost.py为研报方法。