

说明

针对文章的研报的说明：

1. 研报第7页，step1中应该是对 $\beta = 0$ 假设来估计。
2. 研报使用了行业的市盈率的时间序列，我这边没有该数据。依然使用原始的价格序列作为时间序列。
3. 文中策略部分是使用周二的时间序列来估计，在周三进行操作。

代码说明

bubble_fast是check的代码。用以验证自己矩阵写的t检验统计量是否正确。boosting500次，如果使用statsmodels的ols包，会出现大量的计算导致速度很慢。因此，需要将回归改写成tensor计算的方法。

s1_check_alg.py展示了在使用一种行业指数时的情景。

具体跑代码，只需要跑

s2->s3即可。

策略说明

文中对回测的逻辑说的很不清楚，为此我们对回测逻辑进行我认为的合理描述：

取150周作为训练样本，大约三年时间。按照文中对MADF的描述，计算

$$MADF_{\tau} = \max \{ADF_{\tau_2}^{\tau_1} : \tau_2 = \tau \text{ and } \tau_1 \in [1, \tau - \tau_0]\} \quad (1)$$

得到该150周期真实的MADF。

通过Wild-boosting计算MADF的分布。计算真实MADF的分位数值。

关于boosting的计算中，对此进行一个说明：

使用boosting计算了 Δy_t^b ，但是没有计算 y_t^b 。原因是，不知道 y_t^b 具体会取代回归方程(2)哪个值：

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^K \Psi_i \Delta y_{t-i} + \varepsilon_t \quad (2)$$

关于wild-boosting可以参考[boosting regression](#)，我理解的boosting应该对方程左边的 Δy_t 抽样改变，但是不应该改变方程右边的值。因此，boosting抽样算出来的 Δy_t^b 可以直接作为方程(2)右边估计替代 Δy_t 。而计算的 y_t^b 只能用以替代方程右边的 y_{t-1} 的估计。在不改变只替代方程左边的前提下，直接使用研报第8页的boosting公式，即公式(3)进行替代估计。

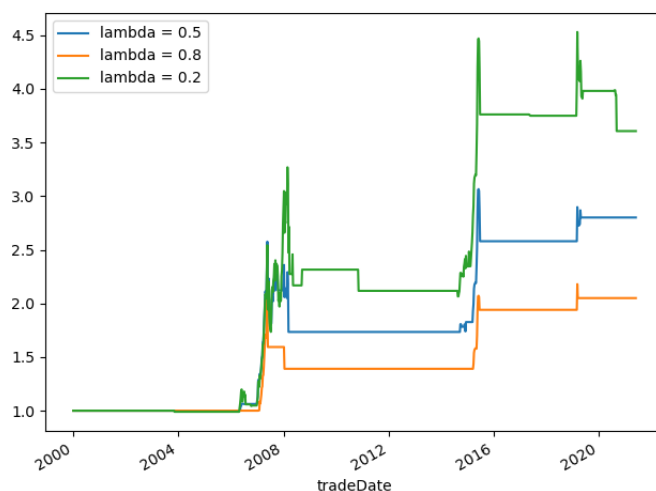
$$\Delta y_t^b = \hat{\alpha} + \sum_{i=1}^K \hat{\Psi}_i \Delta y_{t-i}^b + e_{t-K}^b, \quad t = K+2, \dots, T \quad (3)$$

MADF

ADF作为单位根检验，衡量的是一条时间序列不稳定的程度， t 统计量越大，说明公式(2)中 $\beta = 0$ 的概率越小。说明了此时时间序列的不稳定。MADF说明的是从前面某一个时刻开始，到当前时间段 t 不稳定的程度。 t 值越大，说明了向上上涨的不稳定程度越大。这也是文章使用 t 统计量进行策略的主要逻辑。

研报在实验的过程中使用了较大的分位数，三种分位数并不是很适用在我们的实验中，我们使用以下三种分位数来进行说明。

从s1.py的代码中，我们使用0.2，0.5，0.8三种阈值，并给出回测的收益情况。我们算法完整的捕捉到了08年，15年的两波大牛市。



为什么0.2，0.5的回测效果会好于0.8的回测效果？

价格的不稳定可能也是文章中并没有使用价格序列而使用市盈率序列作为时间序列判定的原因。按照逻辑来看，选择大的阈值是更能保证泡沫发起点的。但是，如果波动较大，说明adf的检验基本都是不会通过的。在wild-boosting的抽样过程中，有一个残差乘以一个正太分布来减小异方差影响的过程。抽样的序列绝大多数绝对值都是小于1的，这就会让boosting得到的adf值偏向平稳。就是说绝对值较小。那么，我们得到的阈值也应该可以相对更小。中国股市的波动较成熟市场的波动是更大的，异方差更大就会让boosting的adf值偏的更小。所以，我们在波动大的市场选择更小的阈值是合理的。这里解释了为什么0.5的效果会好于0.8的结果。

0.2的逻辑出了上述的几点之外，还有一点是在，我们选择的时间维度是足够长的。在这个长的时间维度上有上涨的趋势（中国股市波动大，但是长期看依然是上涨的），再加上boosting造成的更平稳的序列，boosting的adf值大于0的部分是更多的，且绝对值更小，且！这个boosting是在假设平稳的情况下的boosting，即 $\beta = 0$ 对应研报说明第一条，在这个情况下模拟的数值更加平稳。这也就说明了为什么选择0.2这个阈值依然还是有绝大多数的点是没有投资的。

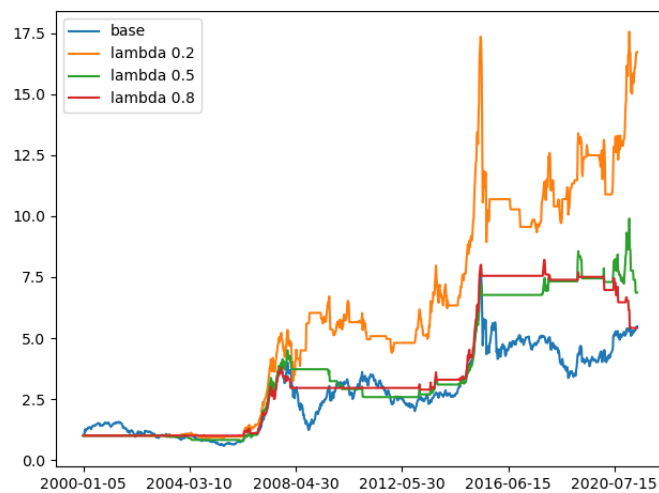
可以通过s1.py过程看到，大于0.2部分的madf值还是挺大的。这也就说明了，这个阈值并不是那么准确。阈值只能说明对风险的接受程度，并不能准的代表泡沫概率。但是可以体现泡沫概率。

加上中国市场整体向好，就出现了以上的回测结果。

回测结果

我们回测了0.5，0.8两个阈值对应的策略1，以及0.2阈值对应的策略2。

0.5与0.8作为避险策略，对应论文的策略一会得到更好的效果。0.2作为一种有机会投资就进行投资的策略，对应文章策略2会有更好的效果。



数据说明

数据来自优矿行业指数。