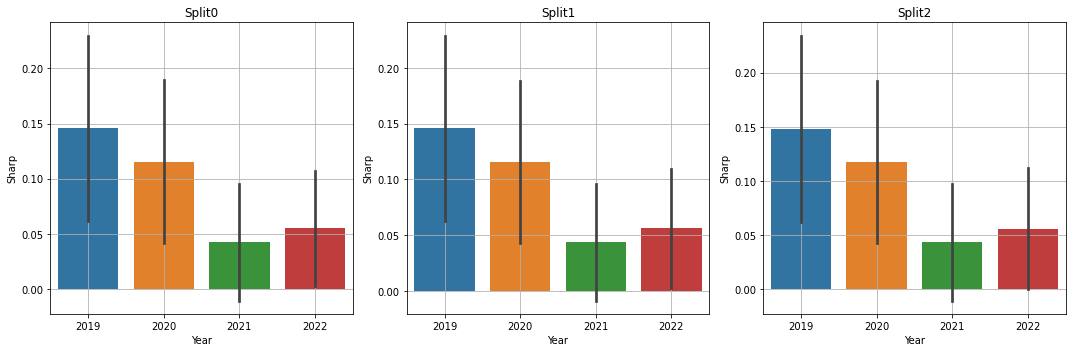
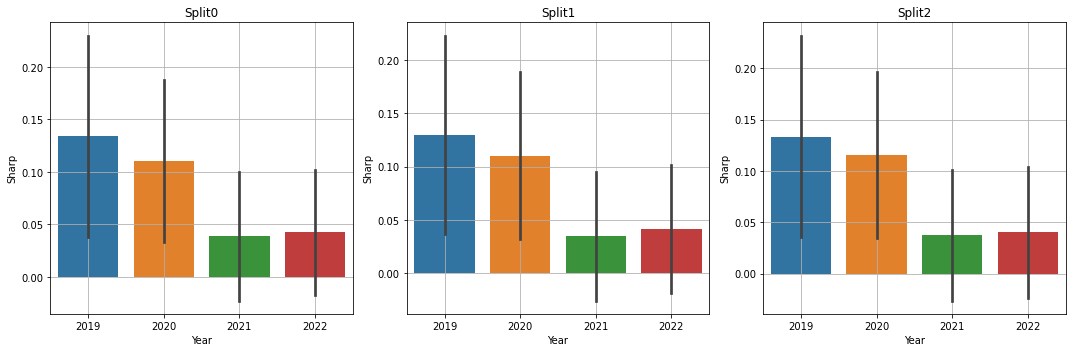
# 2024.01.10

## 查看2700因子的分年表现并以此审视因子筛选方法

之前的所有模型都遇到了2022年表现骤降的现象，并且我的因子方法在三个均分的chunk上得到了相差较大的表现。因此需要审视一下为什么会发生这种不均匀的现象，以重新审视因子筛选的方法是否正确。

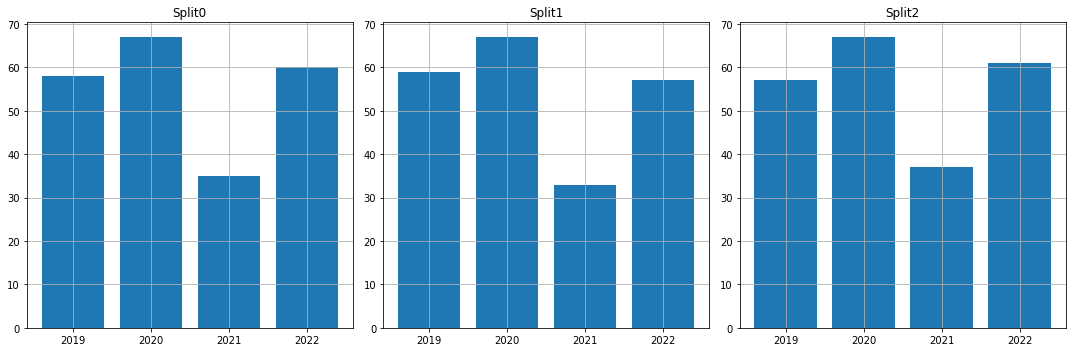
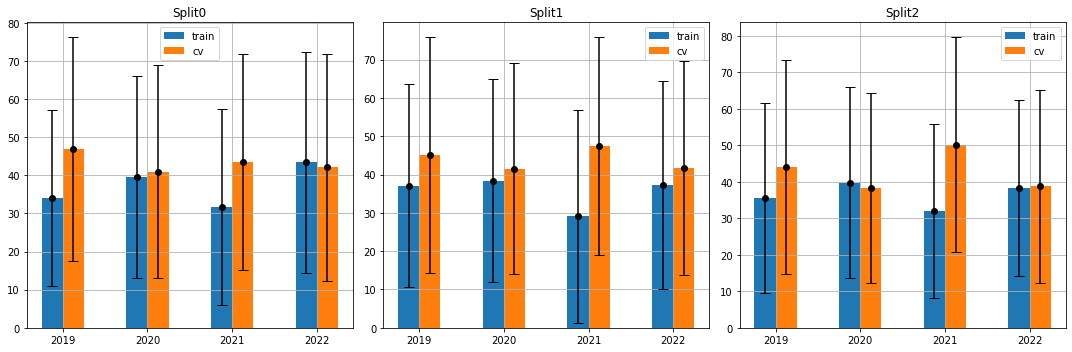
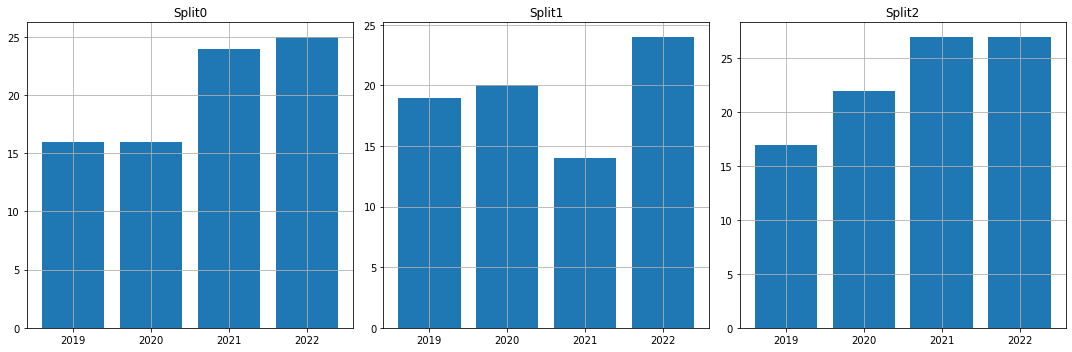
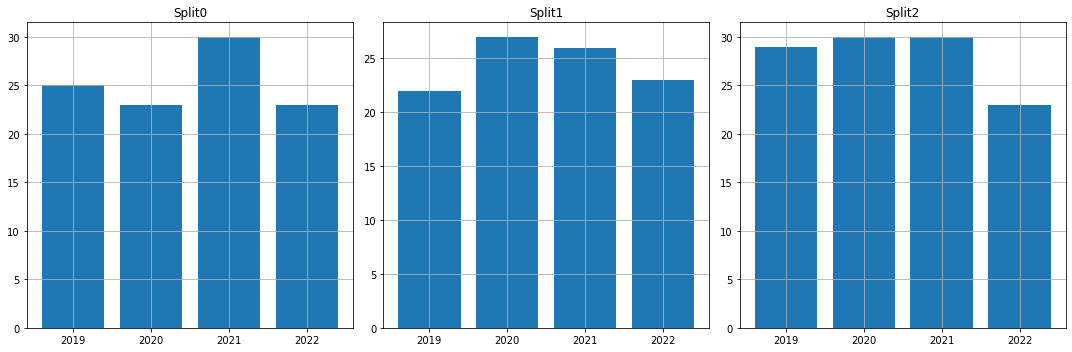
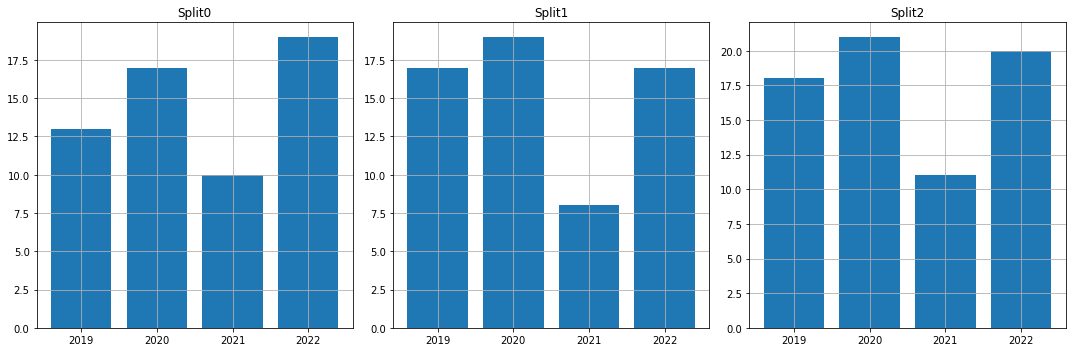
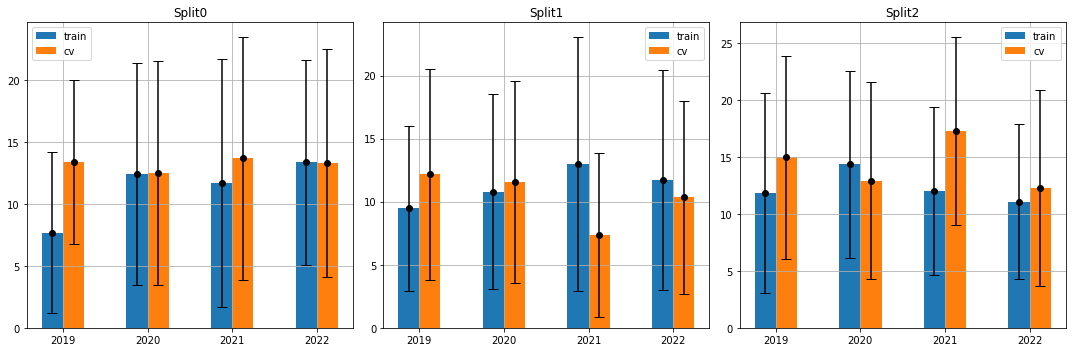
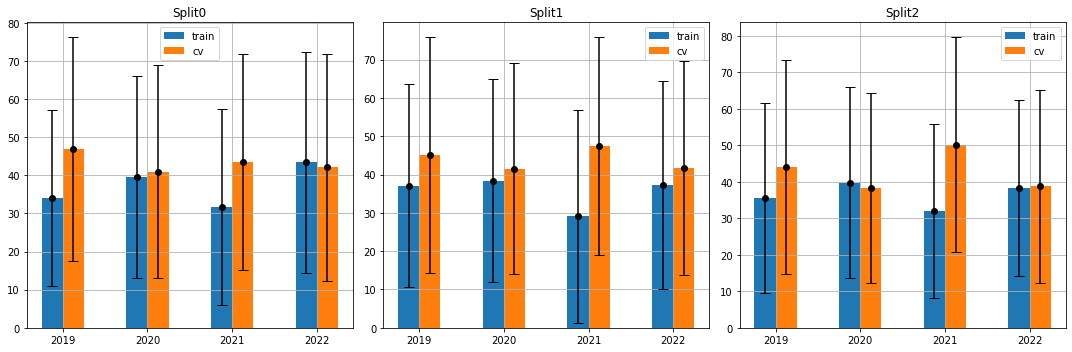
由于我可以根据因子名字读取到单因子的rstat，因此我用单因子的夏普作为单因子表现的衡量工具。

1. 2700因子分3个chunk，分年表现展示
   1. 
   2. 可以看到，因子虽然在2021年之后表现变差，但是2022年反而是平均表现上升的。
2. 而换成我在这三批上筛选的因子，分年表现则如下：
   1. 
   2. 可以看到，换成我自己筛选的因子后，2022年夏普几乎和2021年持平了。也就是说我的筛因子方法并没有把2022年表现好的因子选择到。

对于2022年模型回撤很大的问题，从因子的角度出发，可能有如下几点：

1. 有可能是整体上因子在2022年变差（已排除）
2. 有可能是2018-2021年好的因子，在2022年表现并不好，于是表现变差
   1. 有可能是整体上所有在2018-2021年好的因子在2022年都很差，而恰好2018-2021差的因子在2022年好，导致我的筛选方法选不出2022年好的因子
   2. 有可能是我的因子筛选方法恰好把2018-2021年好并且2022年差的因子选进去了，而不是整体上这样
3. 有可能是2022年被选中的因子虽然表现还行，但是因子种类比较单一（例如：大部分都是反转，或者大部分都是风险因子），从而导致风险暴露较大
   1. 有可能是因子整体上2022年好的因子种类都比较单一
   2. 有可能是我的筛选方法筛出来的因子在2022年种类非常单一

### 以下是实验：

1. 下图是2700因子分三批不筛选的整体情况展示
   1. 这个图展示的是每年测试时，所用训练年份的 Top100 因子在预测年份中还在 Top100 的个数。
      1. 例如2019年是60，则代表在2014-2018年期间，夏普排名前100的因子在2019年仍然排在前100的个数
      2. 
      3. 可见，2021年是一个行情变化较大的一年，这一年里只有35%的因子在测试年份仍然在 Top100，然而2022年的行情延续性却较好。
   2. 这个图展示的是，每年在训练和测试年份中都位于 Top100 的因子的排名的均值与方差
      1. 例如2019年的训练年份（2014-2018年）均值排名是34名，在测试年份（2019年）的均值排名是46名
      2. 
      3. 可见，因子整体上只有2021年发生了较大的排位变化，2022年的行情延续性仍然是极好的。
2. 下图是当我换成2700因子分三批我筛选过后的整体情况
   1. 这个图是我每批筛选后在每年的训练时间排在前30名的因子在整体不筛选的900个里面仍然在前100的数量
      1. 例如2019年这个数量是16，也就是说我筛选的因子（246个）中，2014-2018年期间的平均Sharp在前30名是否能在900个因子中排在前100呢？答案是16个。（为了证明筛选方法确实是把好的因子挑出来了）
      2. 
   2. 这个图是我每批筛选后在每年的测试时间排在前30名的因子在整体不筛选的900个里面仍然在前100的数量
      1. 例如2019年这个数量是25，也就是说我筛选的因子（246个）中，2019年期间的平均Sharp排在前30名的因子是否能在900个不筛选的因子中排在前100呢？答案是25个。（为了证明筛选方法确实是把好的因子挑出来了）
      2. 
      3. a和b纵向对比，可以得出的结论是，在2021年及以前，我的筛选方法筛出来的因子在训练时间段上不是最出色的，但是在测试时间段上是最出色的。然而2022年，这个情况反过来了。
   3. 这个图是我每年筛的因子的中，每年测试时，所用的因子中在训练时间段内排在 Top30 的因子在测试时间段内仍在 Top30 的个数。
      1. 例如2019年是13，则代表在我筛选的246个因子中，2014-2018年期间夏普排名前30的因子在2019年仍然排在前30的个数为13。
      2. 
      3. 可以看到，2021年和2022年确实是行情转折较大的两年。我的筛因子方法并不能规避这个问题。
   4. 当我们把目光聚焦在我筛的因子中训练和测试都在 Top30 的因子，可以得到下图
      1. 这个图展示的是，在我每年筛选的因子中，在训练和测试年份中都位于 Top30 的因子的排名的均值与方差
      2. 例如2019年的训练年份（2014-2018年）均值排名是7.5名，在测试年份（2019年）的均值排名是13名
      3. 
      4. 此图需要和不筛选因子时的整体情况进行比较
         1. 
         2. 经过比较，可以看到，经过了因子筛选之后，训练和测试都在前列的因子，在训练时间段和测试时间段期间的表现差距变大了，尤其是2022年。