1.50ETF期权波动率期限结构指标回测结果：

2016年至2018年,打分1-3低,7-10高

基于牛市价差ratio的skew指标：多空1.56，多3.31，空0.079，最大回撤11.59%,wr=52.24%

Ratio\_skew+diff\_iv: 多空1.56，多3.22，空0.21，最大回撤10.75%,wr=51.14%

主成分skew+diff\_iv：多空2.1，多头4.1，空头0.72，最大回撤10.90%,wr=52.22%

2.价值因子去杠杆化调整

BP,EP,SP,EBITP等因子，其分子对应的是公司总体指标，与之对应的是债权人和股东，但是分母都是股价，仅对应股东。从分子分母相匹配的角度，上述因子没有考虑到经营杠杆。因此在计算式，可以考虑如下调整：B/(所有者权益（股价）+负债-部分非核心资本？)，调整后因子的平均IC值和回测收益均有所提高。

3.盈余质量因子

公司管理层对公司盈余进行过度过量的管理，必然会影响盈余质量（盈余对公司业绩的反应程度，以及当下盈余与未来盈余的相关性），使得盈余质量低下。但还有其他原因会使得盈余质量低下。

一般衡量上市公司盈余管理的因子有很多，比如说：营业利润波动率/经营现金流波动率，因为前者容易操纵，后者不易操纵；或者通过同行业公司回归出一般的盈余和公司资产、收益增量等因素的系数，然后讲系数带到具体公司中计算残差，即为估计的盈余操纵变量。

4. 基于噪音交易者的行为偏差(光大)

Step 1. 光大使用雪球帖子数据，构造投资者行为指数(MDI)。指数编制规则如下：筛选过去一年内帖子最多的前10只个股，按流通市值加权，每年年初调整。

Step 2. 横截面回归CAPM模型中的βCi，其中rft表示无风险利率，rmt为市场/基准的收益率

rit−rft=αi+βCi(rmt−rft)+ϵit

Step 3. 横截面回归BAPM(Behavior Asset Pricing Model)中的βBi，其中rft表示无风险利率，rBmt为"投资者行为指数"(也称为MDI)的收益率

rit−rft=αi+βBi(rBmt−rft)+ϵit

Step 4. 行为偏差BE，即为CAPM与BAPM中的beta差，也称为噪音交易者风险的代理变量

BEi=βCi−βBi

Step 5. 构造衍生因子，

行为偏差波动因子(BE\_STD)，取行为偏差(BE)在过去6个月的标准差

行为偏差因子(BE\_MEAN)，取行为偏差(BE)在过去6个月的平均值。

5.FM五因子

2015年，Fama和French发表了一篇新的五因子模型的文章，在文章里，他们加入了盈利能力和投资两个因子。与三因子类似，新加入的因子也是用组合收益率之差来表示的，盈利能力因子表示高盈利组合股票与低盈利组合股票收益率之差，而投资因子（扩大再生产的能力）则表示投资水平高的股票组合与投资水平低的股票组合收益率之差。其中盈利水平风险是指，盈利能力较高的行业一般会伴随着更高的风险。我们用ROE来衡量盈利水平。记做 E(RMW)，其计算方法和E(SMB)、E(HMI)类似（也是将股票分成三份，然后计算高/低盈利水平的股票期望收益率之差）。而投资水平则用企业的资产增长与之前年度的总资产之比来表示  
 不过，他们发现五因子模型同样有一个缺陷，就是对于那些高投资比例和低盈利能力的小盘股而言，五因子没有很好的解释能力，对此，Fama他们的看法是，CAPM最严重的问题就在于小盘股身上。另外他们还发现，其实在五因子模型里，市值因子的解释能力已经被另外四个因子代替了，如果不是为了逻辑的完整性以及加上这个因子并不会对结果产生影响，我们是可以把市值因子去掉的。  
 而在对三因子和五因子模型在中国的适用能力的研究上，学者似乎还没有达成共识。赵胜民、闫红蕾和张凯等人在2015年的研究认为，三因子模型比五因子模型更适合用来分析中国的A股市场，由于投资者更关注估值水平而非成长性，这导致投资因子和盈利因子在A股市场的功能并不显著。而李志冰等人在2017年的研究却发现，五因子模型的解释性更好，但是他们同时发现A股灾不同时期，比如股改前后因子的作用机制是不同的，而五个因子对于一些组合的解释能力同样不足，这意味着，对于A股市场来说，五因子模型可能还不足以揭示市场的定价方法。对于A股的研究者而言，如何寻找具有中国特色的股票因子，和认识A股市场一样，仍然是一条非常漫长的道路

6.多重共线性

a.逐步回归，向前或者向后，如果新加入变量之后之前的变量不显著了（F检验），则说明有多重共线性，则不选取该指标。

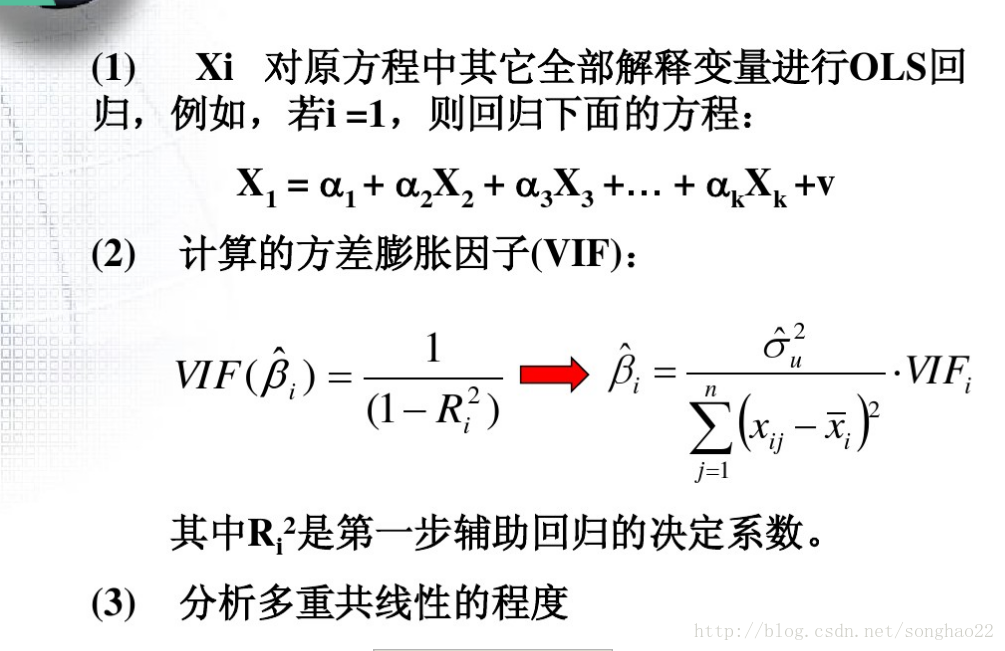
Forward selection: 首先模型中只有一个单独解释因变量变异最大的自变量，之后尝试将加入另一自变量，看加入后整个模型所能解释的因变量变异是否显著增加（这里需要进行检疫，可以用 F-test（是否系数都为0）， t-test 等等）；这一过程反复迭代，直到没有自变量再符合加入模型的条件。

Backward elimination: 与 Forward selection 相反，此时，所有变量均放入模型，之后尝试将其中一个自变量从模型中剔除，看整个模型解释因变量的变异是否有显著变化，之后将使解释量减少最少的变量剔除；此过程不断迭代，直到没有自变量符合剔除的条件。

Bidirectional elimination: 这种方法相当于将前两种结合起来。可以想象，如果采用第一种方法，每加入一个自变量，可能会使已存在于模型中的变量单独对因变量的解释度减小，当其的作用很小（不显著）时，则可将其从模型中剔除。而第三种方法就做了这么一件事，不是一味的增加变量，而是增加一个后，对整个模型中的所有变量进行检验，剔除作用不显著的变量。最终尽可能得到一个最优的变量组合

b.方差膨胀因子VIF：

表**示**某一个因子有百分之多少由其他因子表示



c.增加样本容量

d.变量差分，可能会消除多重共线性，但残差可能出现自相关，因为差分了s3-s2与s2-s1

e.主成分pca分析

f.PSM方法

7.MA方法

a. 等权重移动平均SMA；

b. [n.n-1,,…,1]权重WMA，加权移动平均，当然还有其他加权方法，比如交易量加权等

c. 指数平滑移动平EWMA，权重：[1,1-a,(1-a)^2…]，可以指定区间长度n，a=2/(n+1)。此时a使得区间外的权重不明显，占比小于千分之一；或者可以直接指定a的值。

在python计算中使用pd.ewm进行计算，在数据有限的情况下adjust=True，是按照公式严格计算，在数据量较大的情况下adjust应该=False，是化简了的公式：EWMAt=aSt+(1-a)EWMAt-1

d. KAMA平均