# 目前的策略逻辑-7月24日

**描述：**

PS：这个不一定按我从先到后的逻辑排。

首先，我要做的是一个黄金跨境套利策略（是基于沪金和纽约金的）。

对于跨境套利：当两种黄金的价格进行波动的时候，一旦它们的价差或者是价值比超出无套利区间，我们就有利可图。（我这边是没有考虑交易手续费的，所以无套利区间应该还要再大一点）这种有利可图的稳定性是基于，我们基本可以确定黄金的价差经过一段时间是一定会收敛的，也就是说，在大部分情况下，两种黄金的价格比（我用的是比值）是在某一个范围波动的。（我们可以把两种黄金的比值抽象成一个市场）

因为黄金期货不大可能退市，网格交易成为了一种适应振荡的好方法。网格交易适合振荡，那么在比值有较大上行或者下行的情况时呢？（虽然比较少）我们就可以试着使用一些较低频，更加依赖历史数据分析的方式进行交易。

由于我们是量价分析，所以我们不能使用除了金价，关税，运费以外的任何基本面因子。（作为提醒我自己的话）

**首要的交易逻辑：如果价格比值落在了无套利区间里，平仓。同时，在无套利区间不开仓。**

由于黄金的比值是会收敛到无套利区间里的，只要在区间外正确方向开仓的到区间内交易就可以赚钱，然而，由于这种日子不是特别多，还有，在区间外也不一定能够快速的回来，这可能会导致很长一段时间的亏损。这些就是我们需要改进的地方。

**网格阶段：**

对于网格策略，首先必要的是，我们要确定网格的大小和上下界。

我们知道，黄金价格比也是会有极端值的（极大或者极小的快速变化），如果将这部分放入网格，很有可能会被套牢，或者是爆仓。那我们如何确立进行网格的范围？根据比值收敛，我们可以认为，在靠近无套利区间的一定区间内，价格主要是进行振荡的，我可以拿这个作为网格的上下界。

这里，我使用了**核密度估计**（Kernel Density Estimation，即**KDE**），测算出了大部分的比值（可以通过改变密度参数td来改变需要判断的精度，td=0即是min-max的区间，td=1即为平均值的一个点）都落在哪个区间内，以此来减小极端值的影响。（因为极端值毕竟是少数）

现在，我们需要网格大小，我使用训练集的数据，将整个过程的**平均斜率**计算出来，我们把它作为网格大小。（这个是一个问题，因为平均斜率只能表示每天比值大致会偏移多少，并不能作为合理的网格大小的依据）对于上下界不能被网格大小整除的情况，我让最后一个余数单独成为一层网格。（这个影响应当是不大的，毕竟我现在取的网格大小相当小，几乎是每天都有交易）

然后，对于每一个网格，我交易量是多少？

我目前使用的交易量是**（1/网格数量）\*金字塔因素\*info\_diff** 。以下我一点一点解释：

1. **金字塔因素**： 核心思路是，在较低的市场价格（这里是ratio）开仓，所承受的风险就相对越小，因此，在低位我们应该交易（开仓）更大的数额。于是，该参数的值可以用这样子计算：2 \* (KDE\_upper-prev\_ratio\_value) / (KDE\_upper - KDE\_lower) （比值在0到1之间波动，越高开仓越小，以减少风险）
2. **Info-diff**：

首先介绍一下KDJ指标，它是根据随机指标（Stochastic Oscillator）演变而来的，取值范围通常在0到100之间。当K线从下方穿越D线时，视为买入信号；当K线从上方穿越D线时，视为卖出信号。J线的数值越大，代表价格走势的强势程度越高；J线的数值越小，代表价格走势的弱势程度越高。

我在这边仅使用了J指标，用来判定价格走势的强弱势（或者说是预测斜率）。因为J的数值在0-100之间，且是对单一商品而言，所以还是要对这个数据进行一定的处理的。

我们分别算出AU和NAU的j，并且分别对其求前一项平均，再取对数处理，然后把这个数值做差，再加上1（这样处理之后就是在1左右了，原来是在0左右）。这样，我们就得到了Info-diff。

但是在这个阶段一直遇到问题，从2014-2018左右的时间，在网格上能够逐渐收获利益。可是在之后就不大行了，收益率基本保持不动。

（我需要补充一点的是，我之前采用了一个不是网格的逻辑来处理在网格区间内的策略，但这个方法效果很好。具体操作是，判断上上天的值是否不在上天的值的左右0.0012，如果是，我就认为它是穿出网格线，根据穿出方向，以change作为改变量大小进行交易。Change即是由比值的RSI进行对数处理后再由金字塔比值和KDJ中j线变化率处理的值，这个就先不展开讲了。）

**上/下行阶段：**

在网格策略的区间之外，也就是不常见的，有极端值的区间，我们还需要进行操作。因为在这个区间，比值基本上是不会在某一个值上下振荡的，而是一个大趋势脱离，再一个大趋势回归，在这里我们就需要尽量的抓住较高和较低值。

这里，我引入了一个计算值，叫做”相对位置“，计算方式是：对第i-1的值，考虑0-i-2中有多少个比i-1的值大和小的数目，再除以现在的个数。当然，我采用了基于与0-i-2的值的均值的比较，将相对位置放到了-1（最小）到1（最大）之间。在判断是否开仓时，我认为，在数据充足的情况下（i>150时，不然可能会一直涨就一直1），相对位置的绝对值大于0.995就能说是极端值了。当然，我们不可能仅仅按照绝对值大于0.995就开仓，我们还使用之前的斜率进行判断，当之前五个斜率的平均值的绝对值够大的时候， 我认为它属于急速变化的情况，用以进一步验证。（因为我们知道只要一直上升就一直是1）。

在其他地方，我们不进行交易，以免不必要的损失。整个交易思路差不多就是这样子了。

**问题（主要是网格部分）：**

1. 网格大小的计算和定义还存在逻辑上的不完备，我需要找到更有说服力的计算方法。（虽然我已经把过百分之一的异常值替换，重新计算网格大小，但是效果不明显）
2. 对于网格交易并不赚钱的情况，我一直认为是，由于无套利区间在网格处理的区间内部，有一部分的交易被拦腰砍断了。这导致，在下方的部分没有拿到应有的收益，在上方的部分被逼迫高位开仓。目前我还没有能够很好处理的方法。
3. 关于网格交易数目的确定，我认为目前的情况还是不够合理，虽然1/网格个数是很平凡的一种处理办法，但是它的效果并不好（我认为这个网格太小了，合适的交易频率应该在10-15天才进行一次交易）
4. 在0-150的这个区间内（也就是数据不完全的情况下），我没有什么办法能够找到机会，此外，如果150个都是振荡而后面有一个攀升或者下降的情况，效果会很不好。（因为前面一直没有更大的，所以会很快的开仓，失去机会）