

基于时间序列动态调整的网格策略

张宇萌 姜春妮

摘要

在使用网格策略进行投资时，往往设置网格线在交易的过程中保持静止不变。本文以股指期货跨期套利组合为例，使用ARIMA、VAR等时间序列方法，使用组合价差的历史数据和其他外部因子预测未来价差，从而根据预测值动态调整网格。

本文首先对ARIMA、VAR方法的理论基础和运用在网格策略上的调整作出介绍，然后分别阐述ARIMA和VAR的研究过程，包括因子选取、策略设计、结果分析，对模拟交易的结果进行展示，最后为下一阶段可行的研究空间提出建议。该策略已完成部署。

目录

1. 摘要
2. 基于股指期货的网格策略阶段成果与概述
3. 研究方法介绍
 - a. ARIMA
 - b. VAR
 - c. 集成策略
4. ARIMA研究过程
 - a. 解释变量与被解释变量选取
 - b. 数据分析
 - c. 可执行区间
 - d. 预测结果分析
5. VAR研究过程
 - a. 建立调参框架
 - b. 建立外部因子
 - c. mapeStrategy
 - d. 回测结果
6. 集成策略
7. 部署方式
 - a. 输入数据与输出数据
 - b. 代码说明
8. 未来可以展开的工作
 - a. ARIMA
 - b. VAR
 - c. 其他
9. 附录

一、基于股指期货的网格策略阶段成果与概述

本文研究对象为沪深300股指期货买入远月、卖出近月的股指期货组合，对两合约价差即投资组合的价格进行网格策略交易。主要研究内容为网格线的设置。

第一阶段中，python复现网格交易策略，得到上半年6-3，3-12合约应采取关于0对称、间隔为3的网格线，即 $[-9, -6, -3, 0, 3, 6, 9]$ ，下半年9-6、12-9合约应采取整体向下偏移3的网格线，即 $[-12, -9, -6, -3, 0, 3, 6]$ 。此外，上述两种网格线在交易过程中保持不变。

第二阶段，研究交易过程中根据市场情况的变化，如波动率、持仓量等，对网格线进行动态调整。本文对该阶段的理论模型、成果进行阐述。

二、研究方法介绍

1. ARIMA

差分自回归移动平均模型（ARIMA）将非平稳时间序列转化为平稳时间序列，然后将被解释变量对其滞后值、解释变量进行回归，根据得到的关系式和解释变量预测被解释变量的未来值。ARIMA(p,d,q)中，p为自回归项数，d为使时间序列平稳的差分次数，q为滑动平均项数。

ARIMA的核心思想在于，一个变量的变化受通过其过去的值、其他变量影响，这符合股指期货的特点：价格连续波动、受前几期的价格影响，受波动率的影响。

具体过程为，t1~t2的训练期内，使用解释变量前几期的值、被解释变量前几期的值回归被解释变量当前的值，得到各参数的设置。t2~t3的预测期内，使用解释变量在t2~t3的值、被解释变量前几期的值、t1~t2训练期得到的参数，预测被解释变量在t2~t3取值序列。

具体细节通过python中的pmdarima.arma.auto_arma实现，不需要人工建模或调参。

2. VAR

向量自回归（Vector Autoregressive，简称VAR），是一种多元时间序列方法，将一组时间相关变量表示为它们自己的过去值以及该组中所有其他变量的过去值的线性函数。

之所以使用这种方法是因为，首先价差变动很大程度上是自相关的，在预测未来变动情况时应当考虑自身滞后几阶的情况，其次在之前的研究中得出的结论是价差变动除了自身影响以外可能受到一个乃至多个外部因子的共同影响，而向量自回归可以很好地刻画这种影响关系。

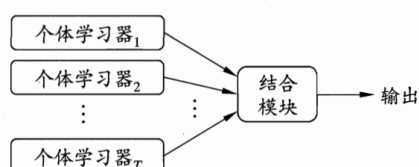
VAR的基本形式如下图所示。该图中可以认为 Y_1 是价差spread， Y_2 和 Y_3 是两个外部因子，可以看到三者是相互影响的，当前时刻t的 Y_1 值既受到自身滞后阶的影响，又受到外部因子滞后阶的影响。

$$\begin{aligned} Y_{1,t} &= \alpha_1 + \beta_{11,1} Y_{1,t-1} + \beta_{12,1} Y_{2,t-1} + \beta_{13,1} Y_{3,t-1} + \beta_{11,2} Y_{1,t-2} + \beta_{12,2} Y_{2,t-2} + \beta_{13,2} Y_{3,t-2} + \epsilon_{1,t} \\ Y_{2,t} &= \alpha_2 + \beta_{21,1} Y_{1,t-1} + \beta_{22,1} Y_{2,t-1} + \beta_{23,1} Y_{3,t-1} + \beta_{21,2} Y_{1,t-2} + \beta_{22,2} Y_{2,t-2} + \beta_{23,2} Y_{3,t-2} + \epsilon_{2,t} \\ Y_{3,t} &= \alpha_3 + \beta_{31,1} Y_{1,t-1} + \beta_{32,1} Y_{2,t-1} + \beta_{33,1} Y_{3,t-1} + \beta_{31,2} Y_{1,t-2} + \beta_{32,2} Y_{2,t-2} + \beta_{33,2} Y_{3,t-2} + \epsilon_{3,t} \end{aligned}$$

因此，VAR策略的基本想法就是选定1~2个外部因子后（选择更多的外部因子也能实现，但是要注意因子过多可能导致的噪声较大问题），用过去一段时间的历史建立上述形式的向量自回归模型，然后使用该模型计算并预测未来一段时间的价差spread，并根据spread调整网格。

3. 集成策略

机器学习中“集成学习”的基本思想是通过构建并结合多个学习器来完成学习任务，下图是集成学习的一般结构，即产生一组“个体学习器”，然后使用某种策略将它们结合起来，这样做的好处在于，假如个体学习器“好而不同”，兼具准确性和多样性，其得到的结果将会更加可信。



我们设计的网格策略参照“集成学习”思想，最后将上文中的ARIMA策略和VAR策略进行集成。

三、ARIMA研究过程

1. 解释变量与被解释变量选取

本文在使用ARIMA时，令解释变量为期货价格波动率的滞后项，被解释变量为期货价差。

之所以使用这种方法是因为，在数据分析中，我们发现期货价格与标的波动率（即自身波动率的滞后）关联性极强，因此使用ARIMA可以根据期货波动率的滞后对期货价格进行预测。

采用期货波动率的滞后项是因为，在预测被解释变量时，需要用“未来的”解释变量对被解释变量进行回归，但这显然是不行的。因此，考虑到期货波动率的滞后和标的波动率基本一致，可以采用期货波动率的滞后作为解释变量，这样在预测 $t_2 \sim t_3$ 的被解释变量时， $t_2 \sim t_3$ 的解释变量“滞后的期货波动率”已知，就可以作为xreg回归被解释变量了。

此外，被解释变量选取为价差而非价格的原因在于，最终网格交易的触发对象是价差而非价格，且价差与价格关联紧密、实际验证中也效果较好，因此选取被解释变量为价差。

2. 数据分析

2.1 波动率

期货价格波动率取240分钟内价格的标准差，因为这种波动率计算方式得到的波动率和价格的pearson相关性最高。

2.2 训练长度和预测长度的选取

试验发现，训练所取的时间段T1较长会出现预测时间序列过于平稳的问题，预测所取的时间段T2较长会出现因为没有及时更新真实数据而预测不准的问题。因此，本部分对T1、T2的选取进行优化、以最准确地预测未来一段时间内的价差变化趋势。将预测出来的趋势与实际趋势对比，就可以知道预测趋势是否准确。也就是说，本部分研究的是，多长的历史数据可以更精确地识别出价差变化的趋势。

常规情况下，观测到spread出现上升或下降，并不能直接根据上升或下降的“苗头”来调整网格偏移量。这是因为一定幅度下的上升下降可能只是源自于价格的扰动、白噪声。那么，如何定量“能够确认市场出现上升或下降趋势”的“一定幅度”就是时间序列解决的问题。例如，当价差出现上扬趋势时，利用这里的方法可以判断当前的上扬趋势是否是“显著”的，若是显著的，则需要上调网格偏移。

最终选取的 $T1=180\text{min}$ ， $T2=10\text{min}$ 。

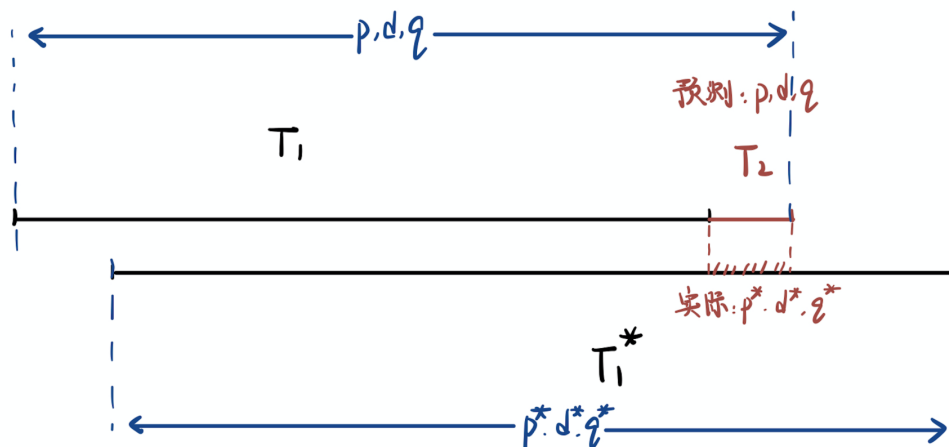
3. 可执行区间

3.1 思路

对每一个训练区间T1而言，ARIMA得到的模型都是不一样的，主要体现在p、d、q的取值不同，而p、d、q直接决定的是T2期间的预测结果。

ARIMA预测的基础是，T1和T2遵从相同的ARIMA模型，包括相同的p、d、q。因此，如下图所示，如果相邻训练期T1、T1*的p、d、q相差太大，说明两段时间适用的ARIMA模型差别太大，说明根据T1训练出来、应该适用于T2的模型与T2实际符合的模型差别太大，也就不能用p,d,q代替p*,d*,q*预测T2阶段的被解释变量。

换言之，这里p,d,q衡量的是一段时间内的“特征”，如果T1和T2的p,d,q相差很大，说明两段时间所符合的时间序列特征是不一样的，使用T1阶段得到的p,d,q去预测T2阶段的结果是十分不准确的。



需要注意的是，“两段时间的 p, d, q 相等”是允许一定误差的，也就是说只要两段时间的 p, d, q 差距在一定范围内，我们可以认为他们是可以公用 p, d, q 、即符合相同时间序列特征、可以用前一段时间训练出来的模型预测后一段的。这是因为， p, d, q 的选取是通过调包实现的“最优”选择，而由于计算误差、最优准则选取的不同等误差，最优选择附近的“较优”选择都是可以接受的。因此，几个训练时间段得到的几组 p, d, q 只要在一定范围内波动，就可以认为这几个时间段内模型保持稳定，可以进行ARIMA预测。我们称这样的时间段为可执行区间。

3.2 可执行区间的获取

本部分，在所有合约中找出共性的可执行区间，作为实际交易中进行ARIMA调整网格的时间区间。

具体方法为，对每个合约而言，首先每30分钟划分为一个时间段看作整体 T ，根据时间先后打上time tag，如第一个30分钟内的所有数据的time idx=1，第二个30分钟的所有数据的time idx=2，以此类推；然后，对每分钟到向前 T_1 的时间段计算 p, d, q ，得到 p, d, q 后填入该分钟的数据栏；然后，计算每 $5 \cdot T$ 的时间段内 p, d, q 各自的标准差，小于等于1.5的时间段认为是平稳的可执行区间。

各合约的可执行区间结果如表所示：

IF2012-2009		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6 ~ 10	3~5	58~63 171 ~ 177 423 ~ 433 446 ~ 454
10 ~ 20	5~10	65~79 180 ~ 190 195 ~ 206 292 ~ 305
20~	10~	85 ~ 114 120 ~ 148 212 ~ 241 245 ~ 284 311 ~ 337 374 ~ 418

IF2009-2006		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6 ~ 10	3~5	98~106 305~312
10 ~ 20	5~10	16~29 50~63 220~235 399~409
20~	10~	68~95 108~179 185~216 245~280 324~349 369~392

IF2006-2003		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6 ~ 10	3~5	139-147
10 ~ 20	5~10	91-104 170-186 242-254 318-331 389-401
20~	10~	32-65 117-137 192-224 260-288 294-314 355-385

IF2003-1912		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6 ~ 10	3~5	21-28
10 ~ 20	5~10	4-18
20~	10~	34-57 81-178 184-222 243-278 284-342 347-390 396-424

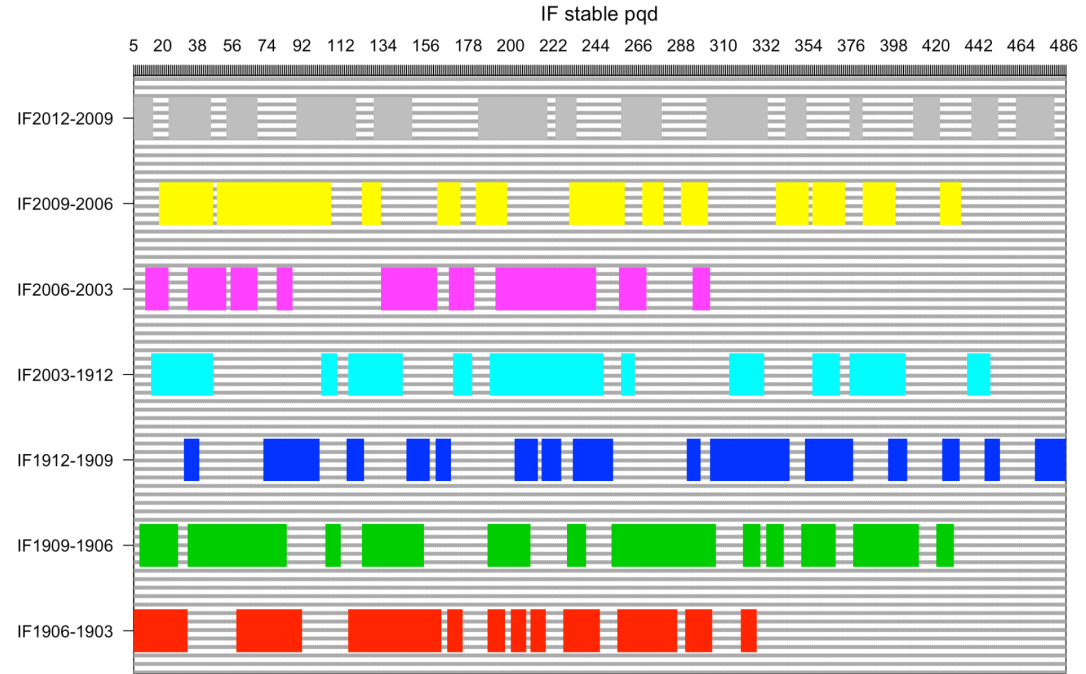
IF1912-1909		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6 ~ 10	3~5	106-116 133-140 145-152 156-165
10 ~ 20	5~10	84-103 384-395 398-408 419-433 455-465
20~	10~	4-79 199-223 226-247 251-378

IF1909-1906		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6 ~ 10	3~5	171-176 184-191 288-295
10 ~ 20	5~10	197-209 215-234
20~	10~	4-28 34-62 68-97 109-158 239-283 300-349 355-419

IF1906-1903		
n个T	持续时长(h)	区间(自有价格以来) 30min为单位1
6~10	3~5	67-73 82-89 94-102 106-113
10~20	5~10	207-225
20~	10~	16-41 116-137 141-188 235-278 284-305

3.3 每份合约的可执行区间

3.2中各合约的可执行区间如图所示：



从上图可以发现，不同年份、相同月份的合约遵从相似的可执行区间。各年份的同月份合约可执行区间取交集，得到各类合约最终的可执行区间如下表所示（单位为从该投资组合有数据开始的每30分钟为一单位的time tag）：

IF03-12		IF06-03		IF09-06		IF12-09	
start	end	start	end	start	end	start	end
10	40	0	70	10	80	90	100
90	140	110	150	250	300	180	240
180	240	180	240	340	400	300	350
350	400						

在上表的可执行区间内才使用ARIMA进行价差预测、网格调整。

4. 预测结果分析

4.1 思路

代码中模仿实际每次传入一定长度的更新数据，滑窗读入数据表格，根据ARIMA预测出来的结果调整网格，然后模拟用调整后的网格进行交易，直到下次网格更新。得到动态网格更新下的策略收益情况。

下面详述如何根据ARIMA的预测结果调整网格：

ARIMA预测得到长度为更新频率10min的价差数据，以这一段数据为依据调整网格。

其一，若预测出来的价差数据段的均值比当前的价差大，则上调网格，反之下调网格。具体来看，预测数据的均值与当前价差的差值乘以放大倍数（本文取0.1），然后得到的值四舍五入到0.2的倍数，就是网格上下便宜的量。这里存在放大倍数这一超参，该参数反映了网格上下偏移对预测的敏感程度，若放大倍数较大，网格根据预测变化的程度更大，同时预测有误差带来损失的风险也更大；若放大倍数较小，则只有当预测变化明显的时候才调整网格的上下偏移，风险更小。

其二，可以根据预测出来的价差数据段方差调整网格间隙。一般来说，方差越大，预测未来10分钟内价差波动越大，网格间隙应调大。这一部分内容较多，可拓展的内容也很多，本文暂时没有进行网格间隙的调整。

4.2 结果

对各合约，采用不调整网格、可执行区间内使用标的资产即沪深300指数波动率、可执行区间内采用主力合约价格波动率的滞后、所有时间下都采用主力合约波动率的滞后，得到收益情况如下：

合约	初始（不调整网格）	IF300波动率	期货主力合约波动率	所有时间下都执行期货主力合约波动率
IF2012-2009	收益：320661 最大回撤：60849 胜率：0.42 sharp比：2.41	收益: 424020 最大回撤：75060 胜率：0.52 sharp比：2.68 收益提升比：32%	收益: 387120 最大回撤：55140 胜率：0.52 sharp比：2.18 收益提升比：21%	收益: 399769 最大回撤：54780 胜率：0.53204 sharp比：2.25 收益提升比：25%
IF2009-2006	收益：75096 最大回撤：14336 胜率：0.34 sharp比：-1.34	收益: 99180 最大回撤：7380 胜率：0.5 sharp比：-0.062 收益提升比：32%	收益: 149820 最大回撤：14040 胜率：0.50 sharp比：2.77 收益提升比：100%	收益: 162120 最大回撤：18480 胜率：0.51 sharp比：2.58 收益提升比：116%

IF2006-2003	收益：149971 最大回撤：5704 胜率：0.37 sharp比：0.69	收益: 122460 最大回撤：6007 胜率：0.51 sharp比：-0.15 收益提升比：-18%	差	收益: 150000 最大回撤：11767 胜率：0.50 sharp比：0.91 收益提升比：0.02%
IF2003-1912	收益：120131 最大回撤：5820 胜率：0.41 sharp比：1.25	收益: 113700 最大回撤：5340 胜率：0.50 sharp比：1.11 收益提升比：-5%	差	收益: 146580 最大回撤：3902 胜率：0.52 sharp比：1.80 收益提升比：23%
IF1912-1909	收益：135202 最大回撤：8097 胜率：0.40 sharp比：2.44	收益: 238320 最大回撤：10860 胜率：0.52 sharp比：3.97 收益提升比：76%	收益: 260460 最大回撤：13921 胜率：0.51 sharp比：4.09 收益提升比：93%	收益: 252480 最大回撤：12600 胜率：0.50 sharp比：3.91 收益提升比：87%
IF1909-1906	收益：197159 最大回撤：8887 胜率：0.41 sharp比：2.22	收益: 322500 最大回撤：10496 胜率：0.52 sharp比：3.20 收益提升比：63%	收益: 361380 最大回撤：16140 胜率：0.51 sharp比：3.38 收益提升比：83%	收益: 353700 最大回撤：14641 胜率：0.51 sharp比：3.34 收益提升比：79%
IF1906-1903	收益：185318 最大回撤：16380 胜率：0.39 sharp比：3.51	收益: 91327 最大回撤：26220 胜率：0.49 sharp比：-1.19 收益提升比：-51%	差	收益: 205562 最大回撤：23040 胜率：0.51 sharp比：3.71 收益提升比：11%
IF1903-1812	收益：803986 最大回撤：16852 胜率：0.44 sharp比：3.74	收益: 704160 最大回撤：16800 胜率：0.52 sharp比：3.83 收益提升比：-12%	差	收益: 700740 最大回撤：23108 胜率：0.51 sharp比：3.70 收益提升比：-13%
IF1812-1809	收益：417942 最大回撤：21167 胜率：0.43 sharp比：3.35	收益: 1489200.0 最大回撤：15592 胜率：0.52 sharp比：5.04 收益提升比：256%	收益: 1539660 最大回撤：21960 胜率：0.54 sharp比：4.63 收益提升比：268%	收益: 1300680 最大回撤：22080 胜率：0.53 sharp比：4.54 收益提升比：211%
IF1809-1806	收益：1111353 最大回撤：14040 胜率：0.40 sharp比：5.12	收益: 1401540 最大回撤：18000 胜率：0.52 sharp比：5.08 收益提升比：26%	收益: 1418100 最大回撤：21120 胜率：0.52 sharp比：4.58 收益提升比：28%	收益: 1472340 最大回撤：20280 胜率：0.52 sharp比：4.68 收益提升比：32%

根据上表结果，有以下结论：

1. 使用主力合约波动率、使用指数波动率都可以进行网格的ARIMA动态调整，为计算简便可以直接使用主力合约波动率
2. 在可执行区间内使用主力合约波动率的方法，只适用于12-09&09-06两种合约，在上半年的两种合约上表现差
3. 在所有时间下都使用主力合约波动率的方法，可以全年使用。这也是最终部署到网站上的算法
4. 由2.3.可知，可执行区间选取的是“执行算法后效果较好”的时间段，带来的优点是可执行区间内的效果好、收益高，缺点是由于可执行区间的选取过于严格、可能放过一些需要调整网格的潜在可执行区间。因此，下一步更好的算法，应该是放宽可执行区间的选取标准，如设置p,d,q的方差在一个更大的数的范围内，都属于可执行区间。
5. 最终选取的任何时间下都使用主力合约的波动率，本地模拟的交易效果显示，绝大多数情况下都有较好收益，只有IF1903-1812收益不好，但这份合约在其他动态调整网格下的收益情况都不好，因此可能是这份合约本身数据的问题，不再考虑了。

四、VAR研究过程

1.建立调参框架

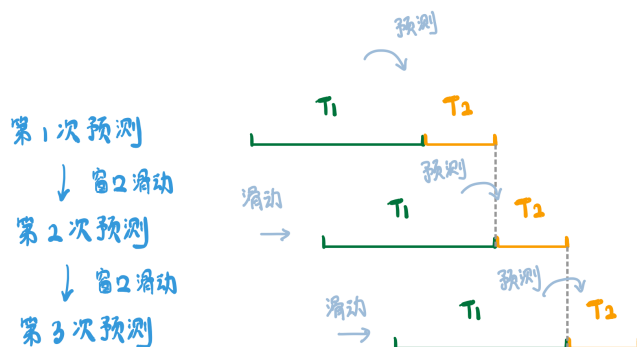
正如研究方法介绍中对于VAR原理的阐述，除了价差spread本身以外，建立模型需要1~2个外部因子，为了验证选取什么外部因子建模的预测效果最好，建立了一个调参框架var_for_grid（该框架不在最终策略中，将放在附录中的github仓库里）。

框架的具体功能如下：

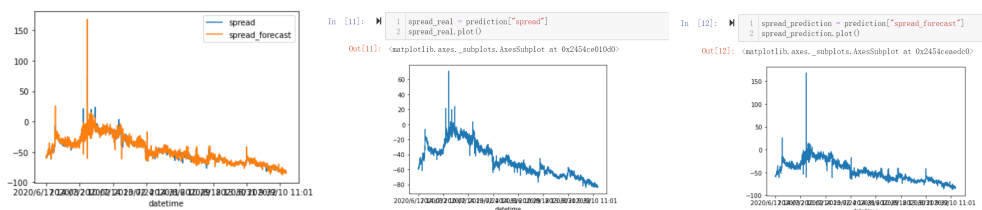
- ① 设置合约名称name、建模因子factors（包括价差spread自身和其他任意的外部因子，只要数据中有）、训练时长t1、测试时长t2（即最终策略中的frequency）

```
1 name = "IF2012-2009"  
2 factors = ["spread", "hs300"]  
3 t1=180  
4 t2=10 # 即frequency  
5
```

- ② 读入合约数据，根据上面设置的参数来进行建模、训练和预测，窗口滑动的具体过程如下图所示。



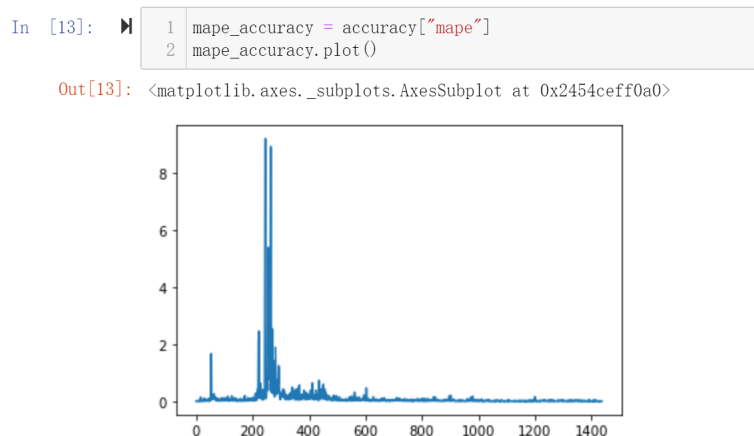
- ③ 将每次预测的t2时间段预测结果存储并拼接，即得到用VAR模型预测的整个合约时间段的spread，画出预测结果spread_prediction、真实数据spread_real以及两者重叠的图，比较预测结果大致的准确性，如下所示。



④ 计算每个t2时间段预测值和真实值的mape值并画图，mape值指的是mean absolute percentage error，即平均百分比误差，用于衡量预测结果的准确程度，mape值越低说明预测结果越准确。具体计算方式如下所示。

```
In [5]: # 计算mape 预测的准确性
def forecast_accuracy(forecast, actual):
    mape = np.mean(np.abs(forecast - actual)/np.abs(actual)) # MAPE
    return mape
```

画图结果如下所示，可以非常直观地看出，在这个合约中，什么时候VAR策略的预测结果是准确的，以IF2012-2009为例，可以看到在前500次预测中mape值相对较高，说明预测结果和真实结果相差较大，而从500次开始到最后结束，mape值基本稳定在0附近，说明预测值相对比较准确。



⑤ 将预测结果和mape值存储在csv中，方便研究和回顾。比如如果选择合约是IF2012-2009，t1为180，t2为10，建模因子选择spread和hs300，结果存储的形式就如下图所示。

IF2012-2009_t1=180_t2=10_['spread', 'hs300']_accuracy.csv

IF2012-2009_t1=180_t2=10_['spread', 'hs300']_prediction.csv

	A	B	C	D	E	F
1		train_start	train_end	test_start	test_end	mape
2	0	2020/6/17 9:32	2020/6/17 14:02	2020/6/17 14:03	2020/6/17 14:12	0.012882
3	1	2020/6/17 9:42	2020/6/17 14:12	2020/6/17 14:13	2020/6/17 14:22	0.008707
4	2	2020/6/17 9:52	2020/6/17 14:22	2020/6/17 14:23	2020/6/17 14:32	0.007328
5	3	2020/6/17 10:02	2020/6/17 14:32	2020/6/17 14:33	2020/6/17 14:42	0.011765
6	4	2020/6/17 10:12	2020/6/17 14:42	2020/6/17 14:43	2020/6/17 14:52	0.017483
7	5	2020/6/17 10:22	2020/6/17 14:52	2020/6/17 14:53	2020/6/18 9:32	0.027788
8	6	2020/6/17 10:32	2020/6/18 9:32	2020/6/18 9:33	2020/6/18 9:42	0.016768

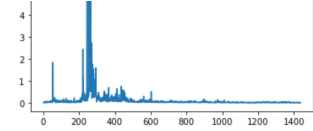
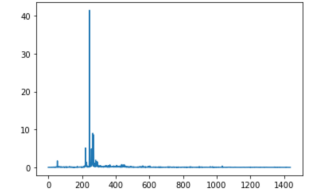
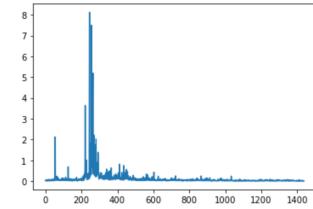
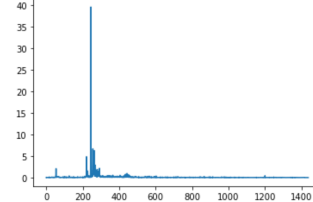
	A	B	C	D	E	F
1	datetime	spread	hs300	spread_forhs300_forecast		
2	2020/6/17 14:03	-59	3998.6	-59.2578	3998.488	
3	2020/6/17 14:04	-60.2	3999.11	-59.3454	3998.007	
4	2020/6/17 14:05	-60	3998.85	-59.4852	3997.469	
5	2020/6/17 14:06	-59.8	3999.69	-59.2898	3996.4	
6	2020/6/17 14:07	-58.4	4001.53	-59.2486	3995.526	
7	2020/6/17 14:08	-58.8	4003.25	-59.4005	3994.43	
8	2020/6/17 14:09	-58.4	4003.88	-59.3318	3993.439	
9	2020/6/17 14:10	-58	4005.06	-59.2454	3992.574	
10	2020/6/17 14:11	-58	4005.88	-59.3288	3991.724	
11	2020/6/17 14:12	-58.8	4007.94	-59.2741	3990.732	
12	2020/6/17 14:13	-59.8	4008.95	-58.6308	4009.398	
13	2020/6/17 14:14	-58	4009.04	-58.0246	4010.375	

这个框架的意义有两点。首先，可以较为方便地看到选择不用训练时长、测试时长以及外部因子后，在整个合约时间段上的预测准确程度，方便调参。其次，在最后的VAR策略中很多过程包括建模等可以直接使用框架中的函数，全过程的mape值展现也为后面设计的mapeStrategy提供了启发。

2.选择外部因子

使用上述框架进行调参，以IF2012-2009为例，选择不同的外部因子，计算整个过程mape的均值，结果如下所示。可以看到，使用价差spread自身以及主力合约的交易量（以IF2012-2009为例，使用IF2009volume）的结果mape均值最小，说明预测结果最好。因此**建立VAR模型选择spread和主力合约交易量volume**。

用于建立mape的因子	整个合约的mape均值	整个合约的mape画图
IF2012-2009 spread、hs300	0.119091	
IF2012-2009 spread、IF2009open_interest	0.117033	
IF2012-2009 spread、IF2009volume	0.110432	
IF2012-2009 spread、IF2009amt	0.110588	
IF2012-2009 spread、timeleft	0.140837	
IF2012-2009 spread、IF2012open_interest	0.11664	
IF2012-2009 spread、IF2012volume	0.111147	
IF2012-2009 spread、IF2012amt	0.110945	

		
IF2012-2009 spread、IF2009volume、 IF2009amt	0.149237	
IF2012-2009 spread、 IF2009open_interest、 IF2009volume	0.117006	
IF2012-2009 spread、 IF2009open_interest、 IF2009volume、IF2009amt	0.148586	

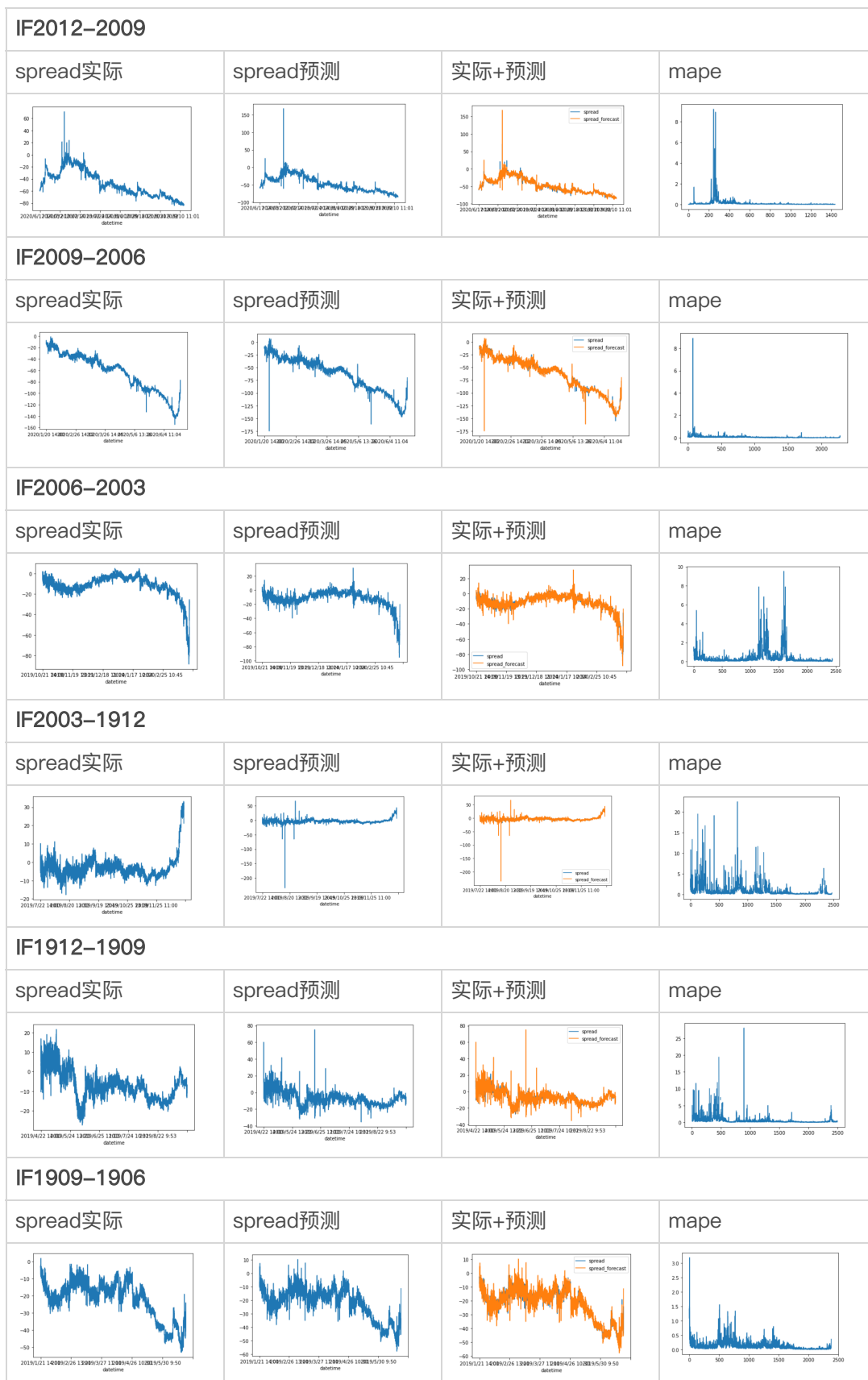
其他条件不变，将外部因子从hs300改为主力合约交易量并进行回测，收益确实提升。

spread+hs300	spread+IF2009volume
收益：421380.0 年化收益率：1.2192708333333333 最大回撤：94083.45852 胜率：0.5237341772151899 sharp比：3.3334829212284562	收益：461820.0 年化收益率：1.3362847222222223 最大回撤：94083.45852 胜率：0.5323529411764706 sharp比：3.2664605468045087

3.mapeStrategy

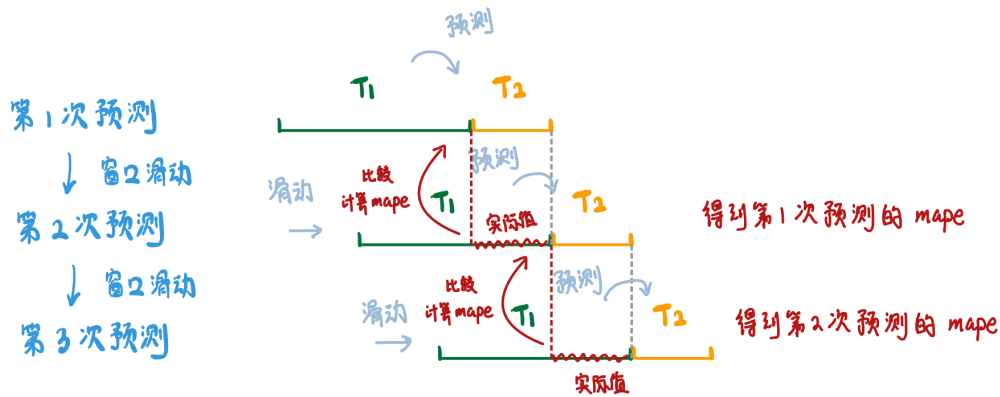
在上面的测试中都使用的是IF2012-2009，大约600次以后预测结果较为准确，所以直接设置的是600次之前网格保持[-12,-9,-6,-3,0,3,6]不动，但是对于其他合约而言这显然是不可行的。在调参框架中使用不同的合约进行预测和mape值计算，结果如下所示，并非所有合约都在后半段得到较好且稳定的预测结果。

可以看到，12-09和09-06的预测结果相对好于06-03和03-12，更适合用基于时间序列预测来调整网格。值得一提的是这个结果和ARIMA策略得出的结果是一致的。



我们希望VAR策略对于任意合约或者任意时间段都能“自己判断预测结果是否可信，即是否要根据预测结果来调整网格”，因此设计mapeStrategy。

策略的关键步骤是在实际使用过程中，无论是否要使用预测结果来调整网格，该策略都会使用当前输入的t1数据来建模并预测t2的spread，将预测结果存储（代码中的全局变量previous_forecast），过了t2时间在下次调用该策略时，上一次预测时间段的实际数据也能够得到了，所以可以计算“上一次预测结果的mape值”，具体如下所示，每一次mape值都进行存储（代码中的全局变量mape_history）。在策略中是存储在全局变量



我们需要设置一个mape的“可接受值”，即如果预测结果的mape值超过了这个阈值，就认为预测结果是不准确的。通过对历史数据预测结果的分析，认为mape的可接受值设定为 ≤ 0.1 是比较合理的。

mapeStrategy的基本思想是：

设置一个全局变量 canshift，取值为0或1，表示当前的VAR策略预测结果是否准确可信，是否可以根据VAR策略的预测结果来移动网格，1表示可以，0表示不可以。

① 从初次调用策略开始到预测次数（turn）不超过10次时，canshift都保持为0，因为一开始的mape值看起来都不太平稳，保险起见策略开始时不进行预测。

在每一次进行预测前更新canshift值。

② 如果 canshift=0，

看之前五次的mape值是否都小于0.1，如果是的话，canshift=1，从本次开始可以根据预测调整网格。

③ 如果 canshift=1，

看之前一次的mape是否大于0.1，如果是的话，canshift=0，认为预测异常，立刻停止根据预测调整网格，网格保持不变。

通过mapeStrategy，VAR策略可以自动识别什么时候预测的准确率较高、预测结果可信，并且根据识别结果来决定是否根据预测对网格进行调整。

保持其他条件不变，使用mapeStrategy再次进行回测，可以看到收益提升，说明了mapeStrategy的有效性。

从600次预测后开始使用预测值调整网格	根据mapeStrategy自动判断是否使用预测值调整网格
收益：461820.0 年化收益率：1.336284722222223 最大回撤：94083.45852 胜率：0.5323529411764706 sharp比：3.2664605468045087	收益：479460.0 年化收益率：1.387326388888889 最大回撤：93903.45852 胜率：0.5218023255813954 sharp比：3.3397833406108184

在使用了mapeStrategy后，只要连续地调用VAR策略或者最后的集成策略，每次预测的mape值就能存储在代码的全局变量中，并自动根据历史mape值判断当前预测结果的准确性，是否根据预测结果调整网格。

4.回测结果

选择和初始代码相同的时间跨度进行回测，结果如下所示。因为12-09和09-06的预测结果较好，而06-03和03-12的mape值波动率较大，使用mapeStrategy判断后对于网格的移动并不多，因此这里只对12-09和09-06进行回测，结果如下，收益都有提升。

合约	初始	VAR Strategy
IF2012-2009	收益：320661 最大回撤：60849 胜率：0.42 sharp比：2.41	收益: 508440 最大回撤：97980 胜率：0.53 sharp比：2.79 收益提升比：59%
IF2009-2006	收益：75096 最大回撤：14336 胜率：0.34 sharp比：-1.34	收益: 185760 最大回撤：21120 胜率：0.51 sharp比：1.62 收益提升比：147%
IF1912-1909	收益：135202 最大回撤：8097 胜率：0.40 sharp比：2.44	收益: 248580 最大回撤：13921 胜率：0.51 sharp比：3.98 收益提升比：84%
IF1909-1906	收益：197159 最大回撤：8887 胜率：0.41 sharp比：2.22	收益: 330180 最大回撤：14700 胜率：0.50 sharp比：3.09 收益提升比：67%
IF1812-1809	收益：417942 最大回撤：21167 胜率：0.43 sharp比：3.35	收益: 1453440 最大回撤：10560 胜率：0.54 sharp比：4.99 收益提升比：248%
IF1809-1806	收益：1111353 最大回撤：14040 胜率：0.40 sharp比：5.12	收益: 1429920 最大回撤：20880 胜率：0.52 sharp比：5.06 收益提升比：29%

和ARIMA策略回测结果进行比较，会发现在上述六次合约的回测中，三次的收益好于ARIMA，另外三次的收益比ARIMA差，因此可以认为ARIMA策略和VAR策略在不同的合约预测的表现不同，可以看作两个“好而不同的个体学习器”，体现了下文中设定的集成策略意义所在。

五、集成策略

在研究方法介绍中已经提及“集成学习”的优点，我们设计的网格策略参照“集成学习”思想，最后将上文中的ARIMA策略和VAR策略进行集成，具体方法是每次预测时同时运行两种策略，将得到的两个网格预测值取均值，并作为最终结果输出。

该集成策略还有一个优点在于其保险性，假如在某个时刻，两个策略其中之一对于网格的预测产生较大偏差，和另一个策略取均值后可以有中和的作用，使策略整体更加稳健。

六、部署方式

1. 输入数据与输出数据

输入数据：

名称	数据类型	释义
data	dataframe	长度为250，每10分钟更新一次，传入前250分钟至当前的数据

名称	数据类型	释义	示例
datetime	\	时间	2019/6/20 14:28
远月合约名 如IF2012	double	远月合约价格	3771.8
近月合约名 如IF2009	double	近月合约价格	3825.0
spread	double	远月价格减近月价格价差	-53.2
volume	int	近月合约交易量，如IF2009的交易量	44
avg3	double	spread的三日均线	-52.8

输入示例：

datetime	IF2012	IF2009	spread	volume	avg3
2020/6/23 9:44	3970	4000.2	-30.2	126	-37.4
2020/6/23 9:45	3962	3995.2	-33.2	137	-37.4
2020/6/23 9:46	3958.6	3990	-31.4	187	-37.2
2020/6/23 9:47	3963	3995	-32	135	-37.2
2020/6/23 9:48	3961.8	3994.2	-32.4	140	-37.2
2020/6/23 9:49	3962.6	3994.4	-31.8	54	-37.2
2020/6/23 9:50	3960.4	3993	-32.6	41	-37.2
2020/6/23 9:51	3961.2	3992.4	-31.2	74	-37.2
2020/6/23 9:52	3965.8	3997.4	-31.6	67	-37
2020/6/23 9:53	3964	3995.2	-31.2	75	-37
2020/6/23 9:54	3963	3997.6	-34.6	95	-37
2020/6/23 9:55	3963.6	3995.2	-31.6	62	-37

输出数据：

名称	数据类型	释义
grids	double list	价差与三日均线只差可以交易的网格线

输出示例：

[-12.2, -9.2, -6.2, -3.2, -0.2, 2.8, 5.8]

2. 代码说明

2.1 文件名称及对应策略

文件名称	对应策略
timeseriesGridsStrategyARIMA.py	ARIMA策略
timeseriesGridsStrategyVAR.py	VAR策略
TSGridsStrategy.py	集成策略

2.2 函数或类名称及功能

函数或类名称	功能	所属策略
get_volatility	计算主力合约的波动率	ARIMA
do_time_series	判断是否进行ARIMA预测	ARIMA
gridSummary	建立ARIMA模型并根据预测结果调整网络	ARIMA
ARIMA1	上半年合约进行网格调整	ARIMA
ARIMA2	下半年合约进行网格调整	ARIMA
timeseriesGridStrategyARIMA	调用ARIMA策略的函数	ARIMA
timeseriesGridsStrategyVAR	调用VAR策略的函数	VAR
do_time_series2	是否使用VAR策略	VAR
forecast_accuracy	计算上一次预测结果的mape值	VAR
mapeStrategy	更新canshift, 判断当前预测准确度	VAR
select_p	建立VAR模型时选择滞后阶数p	VAR
invert_transformation	将一阶或二阶差分后的数据还原	VAR
gridSummary2	建立VAR模型并根据预测结果调整网格	VAR
TSGridsStrategy	将ARIMA和VAR策略集成	集成

2.3 全局变量

全局变量名称	释义	所属策略
T1	训练时间段长度	ARIMA、VAR
freq	预测时间段长度	ARIMA、VAR
grids	集成策略中将ARIMA和VAR的网格取均值后得到的最终结果	集成
volatility	主力合约波动率	ARIMA
grids1arima	ARIMA策略的网格（06-03 03-12）	ARIMA
grids2arima	ARIMA策略的网格（12-09 09-06）	ARIMA
canshift	预测结果是否准确可信，是否要根据结果来调整网格	VAR
mape_history	之前所有预测的mape值	VAR
previous_forecast	上一次的预测值，用于在窗口滑动后得到真实值，计算mape值	VAR
turn	预测次数	VAR
grids1var	VAR策略的网格（06-03 03-12）	VAR
grids2var	VAR策略的网格（12-09 09-06）	VAR

注：这里的全局变量仅讨论TSGridsStrategy.py，在策略单独的代码文件中有些许不同，比如TSGridsStrategy.py中的grids_var在timeseriesGridsStrategyVAR.py中的名称即为VAR，因为基本相同所以不再重复说明。

七、未来可以展开的工作

1. ARIMA

1.1 外部因子的选取

本文选取的是主力合约的波动率作为外部因子，可以考虑其他外部因子，或增加外部因子的个数。

1.2 波动率的计算方法

本文波动率的计算是简单的方差，可以使用加权重、EWMA、ARCH、GARCH等方法计算波动率，会更加准确。

1.3 可执行区间的选取

可执行区间选取的是“执行算法后效果较好”的时间段，带来的优点是可执行区间内的效果好、收益高，缺点是由于可执行区间的选取过于严格、可能放过一些需要调整网格的潜在可执行区间。因此，下一步更好的算法，应该是放宽可执行区间的选取标准，如设置p,d,q的方差在一个更大的数的范围内，都属于可执行区间。

2. VAR

2.1 外部因子的选取

在VAR Strategy中选择建模的因子是价差spread和主力合约交易量volume，可以考虑其他外部因子，在合理的范围内也可以增加外部因子的数量。

2.2 mape的可接受值

VAR Strategy认为mape的可接受值是0.1，即预测结果和真实值的mape一旦超过0.1即认为预测不准确，不再继续使用预测值调整网格。这个可接受值可以作为参数调整，如果希望策略更加保守，可以降低，如果认为可以接受更高的mape值，则可以提高。

2.3 mapeStrategy

mapeStrategy设定预测开始10次以内都不根据预测值移动网格，5次预测值都在可接受范围内则认为预测较为准确，只要出现1次预测值超出可接受值即认为预测出现较大偏差，这里的10次、5次、1次都是可调参数，可以根据选择更加保守或冒险的策略来进行相应调整。

除此之外，可以根据历史mape值设计更好的mapeStrategy。

3. 其他

1. T1、T2的选取，即训练时长、预测时长（即数据更新频率）的选取。
2. 网格间隙的调整。
3. 其他市场，如IC，IH。

八、附录

github仓库：

运行代码：<https://github.com/JennyJiang118/grid.git>

部署代码：<https://github.com/JennyJiang118/timeSeriesGrids.git>