

B 题 光伏建筑一体化板块指数发展趋势分析及预测

第一问题目提示很明显了，移动平均线，还有股票市场中有板块指数（将同一板块的个股按不同的权重方式生成相关指数），它是对该板块走势的整体反映。移动平均就是取前 5 日、10 日、20 日数据的平均值作为下一时间节点的值，这本身就是一种预测算法，附件 1 中有 37 个小表，每个小表中可能大家还会关心交易时间，要取什么时段的数据直接匹配相应字符串即可，注意每个人电脑打开文件日期格式可能不太一样，因此在字符串匹配的时候需要注意下；然后需要注意的是，股票开盘是周一到周五，所以不存在周六周天缺失数据，也会存在部分股票在周一至周五停盘的现象，停盘不是数据缺失，别瞎搞；各股票入市时间也不一样，在识别字符串的时候一定要看有没有对应的字符，如果觉得日期格式麻烦，excel 表中可以将日期格式转换为数值型。

我电脑打开日期格式是 XXXX-XX-XX，有些小伙伴的是 XXXX/XX/XX，以 XXXX-XX-XX 为例，通过 find 函数把后续需要用到的数据位置找到：

```
a=find(X(:,3)=='2019-03-04'); % 2019 年 4 月 1 日前 20 个交易日  
b=find(X(:,3)=='2021-04-30');  
c=find(X(:,3)=='2019-04-01');
```

既然题目说到了 K 线，论文里怎么能没有 K 线图呢，其实绘不绘制 K 线图都不影响结论，K 线是炒股软件中必备的参考，绘制出来显得图更专业点，这不就和其他参赛者拉开差距了吗。



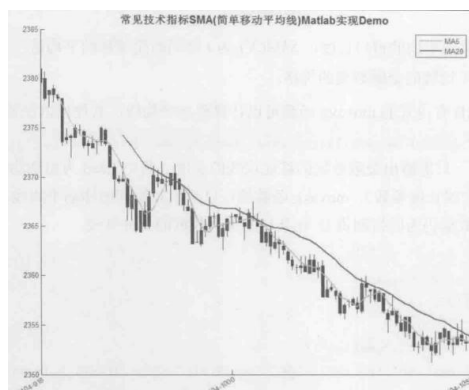
绘制代码样例，是绘制单支股票的，如果文章页输不够，可以绘制几个单个股票图凑凑数

```
X=[];  
i=1;  
[~,X]=xlsread('附件 1.xlsx','Sheet0',num2str(i),'');%很多人这里报错，是因为文件中  
第一个小表命名和其他小表的不一样，自己改下就好了  
X=string(X);  
X=rmmissing(X);%删除<missing>行  
X(2,8)=0;X(2,9)=0;  
a=find(X(:,3)=='2019-03-04');%很多人这里也报错，自己去看看那支股票有没有这个时间
```

的数据，如果是之后才入市的，就换成最新入市时间

```
b=find(X(:,3)=='2021-04-30');
c=find(X(:,3)=='2019-04-01');
Y=double(X(a:b,[4 5 6 7]));%开盘价、最高价、最低价、收盘价
figure
Kplot(Y(:,1),Y(:,2),Y(:,3),Y(:,4))%这里报错是因为 ab 未识别到字符串
ylabel('成交价')
title([X(2,2)+'--K 线图'])
xlim([0,length(Y)+1])
XTick=[1:length(Y)-1];%
XTickLabel=['2019-03-04'
            '2021-04-30'];
XTickLabel=string(XTickLabel);
set(gca,'XTick',XTick);
set(gca,'XTickLabel',XTickLabel);
```

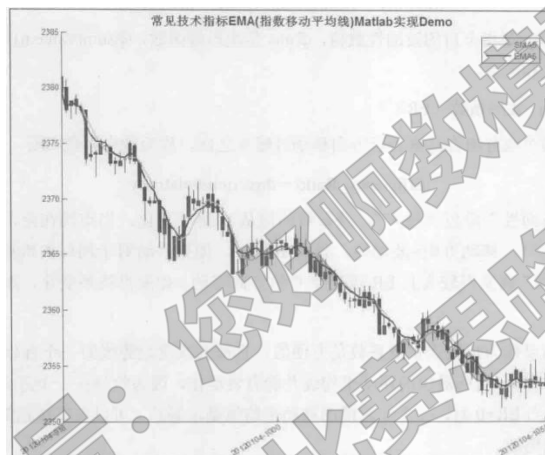
接着说第一问的板块指数，题目有说是将同一板块的个股按不同的权重方式生成相关指数，附件中的开盘、最高最低、收盘数据已经是指数了，这是股票指数，不需要你再去算什么，板块指数又是另一码事，是通过加权和之后乘以基点数算出来的，基点数股票都是 100，外汇才是 1000，不了解的可以去问问会炒股的。权重我们一般参考成交量或成交额的比例做归一化之和为 1 来定；如果想从算法角度出发，可以通过熵权法求得权重乘以原有数值，加权和之后乘以基点数得到光伏建筑一体化板块指数。板块指数使用收盘价来算，炒股我们业主要看重收盘价，除了收盘价，还是建议绘制出 K 线图，因此开盘价、最高价、最低价的板块指数可以采用同样的方式进行计算。通过四个特征指标的指数值，可以绘制出 K 线更直观的体现出趋势，如果觉得 K 线图没必要画，可以就绘制一条板块收盘价指数曲线就行。接下来题目要求做移动平均，数据准备按 2019-04-01 到 2021-04-30 取历史最近 5 个、10 个、20 个取数，每近 5 个、10 个、20 个数据取平均算下一个数，以此类推，最后 hold on 加一条曲线到 K 线图上，例如下图，移动平均又叫简单移动均线法（SMA），加个名称显得有格调一些。



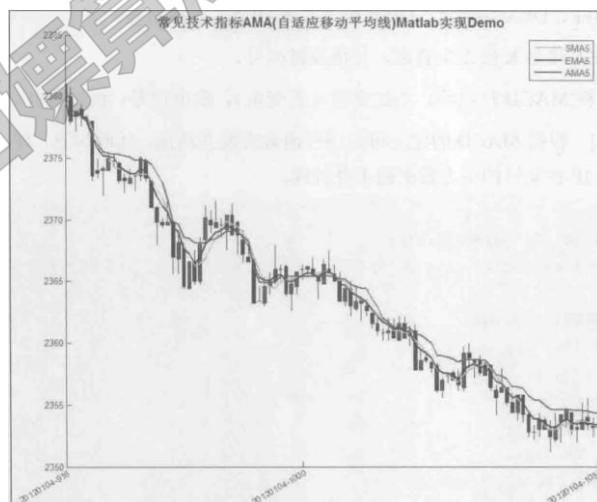
第二问，SMA 是一个参考，也是一个预测模型，作为参考线的时候主要是观察股价是否还有上涨的潜力，做预测时看效果误差肯定不行，误差分析比如说可以做下残差图、误差棒图等等，也可以通过计算 F、t、R 检验值反映误差很大。修正模型很简单，时间序列方法

基本框架都很相似，换一种时间序列方法就行，推荐几个方法及效果，均在《量化投资——以 MATLAB 为工具_李洋-郑志勇》第四章中，第二问采用的新方法要与第一问移动平均做对比，突出第二问模型的优势。

①指数移动平均法（EMA），效果



②自适应移动平均法（AMA），下图中的数据预测的效果要比 EMA 要差点，可以多对比几个算法选最好的用



确定好了新模型，然后预测 5 月 28 日后的 20 个交易日的收盘价数据，如果觉得有绘制 K 线图的必要，可以在将最高价、最低价、开盘价也预测出来，预测难免有误差，如果预测出的最高价低于了开盘价或者收盘价其中一个，就取开盘价和收盘价中最大的一个作为最高价。预测出了 20 个交易日的各指标数据，通过 SMA 算下周月的移动平均数，基于收盘价绘制出每日平均线(日平均线就是原数据)、周平均线(取历史近 5 个交易日取平均)、月平均线(可以取历史近 22 个工作日取平均)，最好是绘制出 K 线做参考。移动平均很好理解，就是依次向后遍历取 n 个数据进行计算。

第三问数据连接

<http://www.csindex.com.cn/zh-CN/indices/index-detail/000001>



会模拟点击爬虫就用程序爬取数据，不会就手动统计

本文还是考虑的是 37 支股票的整体，还是基于第一问的结果数据，注意这里是以两个月为一个时间段，该时间段的数据取平均或者是取时间段末数据都可以，数据大小和相关性没什么影响，如果数据变化趋势越相近，相关性就越大，求相关性方法有常用的三大相关系数法(皮尔逊、Kendall、Spearman)可用 corr 函数实现，还有余弦相似度

余弦相似度代码：

$$D = (x \cdot y) / (\text{norm}(x) * \text{norm}(y))$$

或

$$D = 1 - \text{pdist}([x', y'], 'cosine')$$

第四问，首先是对 37 支股票的投资风险进行评估，评估最好是采用最近三个月数据实际中最有效，一种方法是从开盘和收盘价格角度去考虑，当日收益=收盘-开盘，可以通过 Var 风险价值公式进行计算，参考《金融数量分析：基于 MATLAB 编程》第二章；第二种是从数据波动角度去衡量，简单的可以先对数据做平滑处理然后去做误差，相互比较进行排序；还有个方法是股票 Lyapunov 指数法，该方法也是股票界比较认可的方法，Lyapunov 指数值越大，股票风险越高，该方法主要描述股票的混沌离散状态。如果有其他能描述数据的混沌情况都可以。

算法步骤

用小数据量法计算最大 Lyapunov 指数的步骤如下所示：

(1)对上证上证指数时间序列 $\{x_n\}$ 进行快速傅里叶变换以求出平均序列周期

P ;

(2)根据前文计算所得的延迟时间间隔 τ 和嵌入维数 m 进行相空间重构，得到新的时间序列 $\{y_j\}$;

(3)在重构的相空间中，选取一个参考点 y_j ，寻找其最近的邻点 y_{j_1} ，并限制其短暂时分离，即 $d_j(0) = \min \|y_j - y_{j_1}\|, |j - j_1| > P$;

(4)对参考点 y_j ，计算与其最近的邻近点 y_{j_1} 在离散 i 步之后的距离：

$$d_j(i) = \|y_{j+i} - y_{j_1+i}\| \quad (4.12)$$

这一步的前提是假设该序列为混沌系统，隐含的关系为：

$$d_j(i) = C_j \exp^{\lambda(i \Delta t)}, \text{其中 } C_j = d_j(0) \quad (4.13)$$

对上式两边同时取对数得到：

$$\ln d_j(i) = \ln C_j + \lambda(i \cdot \Delta t) \quad (4.14)$$

(5)对(4.14)式利用最小二乘法做线性回归得到的斜率即为该系统的最大 Lyapunov 指数。

利用 matlab 编写程序得到上述回归方程和图像如图 4-6：

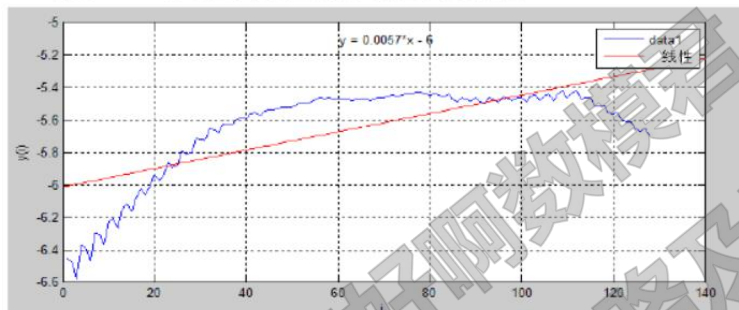


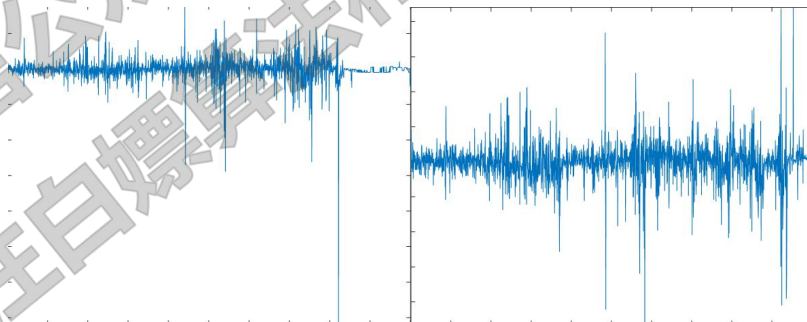
图 4-6 小数据量法计算最大 Lyapunov 指数

由图 4-6 可以看出，对其进行线性回归得到回归方程为：

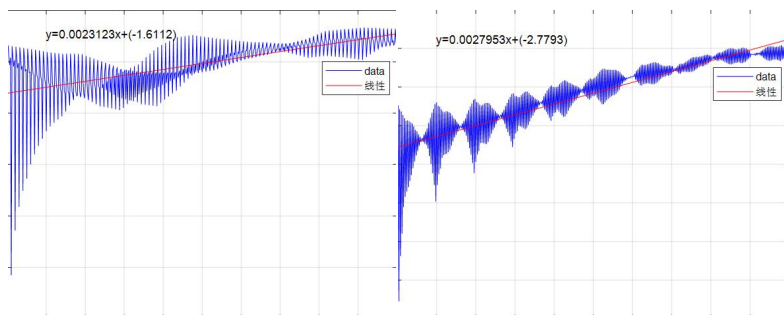
$$y = 0.0057 * x - 6 \quad (4.15)$$

指数值为最后回归方程的 x 系数

效果图：案例程序见群 322273643



如图两个股票某指标的变化曲线，Lyapunov 指数结果如下



从结果来看后者风险要高于前者

股票为 T+1 模式，第一天买第二天开盘才可以卖，如果要做的理想化可以假设卖出价格为第二天的开盘价。胡总和假设当天收盘价买入或者是最低点买入，第二天最高点抛出，怎么来自己决定。

在实际炒股中，一般做投资不超过本金的 15%，至多不超过 20%，每从一支股票买入 1 股是 100 个点，不一定需要购买整数股，买 0.01 股也行，买入股的花费=股数*单价*100，100 为基点数，就比如南玻 A 在 2021-05-27 最高价买入 1 股，单价为 10.22，那么就需支付 1022 元。

一个月投资为短期投资，上一步已经确定好了股票排序，接下来在预测下六月份每个交易日的开盘价、收盘价、最高价、最低价，投资固有策略：拿本金的 15%出来选择当前最好的 5 支潜力股，这里的潜力股，第一天就以排名来看，资金分配比例可以参考各股的日涨幅、收益率等比例进行划分，股票不可能一直盯着数据看，也不可能有跌就立即卖出，这里有一个平仓线概念，就是说设置一个涨幅线和跌幅线，涨了 15%或跌了 15%，直接卖掉，涨幅和跌幅线可以自行设置。当有卖出决策发生时，可以视为股票发生明显，需要重新排名分析，重新取近 3 个月数据计算 Lyapunov 指数重新排名。投资还有一个衡量股票存在潜力的指标是 MACD，见《量化投资——以 MATLAB 为工具》李洋-郑志勇》第四章，该指标主要衡量的是买入和卖出的趋势，侧面可以反映出涨跌的增长率，当大量投资者买入股票时，股票价格会增长，同样的大量投资者卖出，那么股票价格会降低，MCAD 与 DEA(指数移动平均线)的交点被投资者们作为买入和卖出的切入点。

第四问思路是实际中的决策，就别用全局用优化算法寻优了。构建好规则后就按时间模拟就行，唯一需要做优化设计的就是涨跌两条平仓线，一般范围在 10%-20%之间，决策触发点要么为达到平仓线，要么参考 MCAD 与 DEA 得交点，设置平仓线也是一项重要的决策，现实中也很难保证能在最高点抛出。当然也可以按理想状态来，但最终得给出结果并说明，金融行业检验职业操盘手的一项参考为总资产的变化曲线。总资产=手上剩余资金+买入股票的价值，一般总资产的变化曲线呈梯度向上形，如果最后出来的曲线分布类似，那么可以直接吹自己的模型很优秀了。

如果有懂炒股的小伙伴可以设置跟复杂的决策，每到一个决策点，如果是涨到平仓线可以选择继续加仓，而不是卖了之后再买入，这样可以赚到更多，如果跌至平仓线，如果股票存在潜力，可以选择性平掉 50%仓位。甚至可以增设平仓线的调整决策，例如股票一直涨，达到一定程序后，重新规划平仓线。

为什么说建议绘制 K 线，下面有两篇文章介绍了 K 线几组合形态的看涨看跌信号，这也是众多炒股高手们参考的重要信号。如果还想模型跟复杂化，可以参考下。

<https://www.163.com/dy/article/EH3QUSOH0519BVN5.html>

<https://zhuanlan.zhihu.com/p/58824781>

其实第四问参考的角度很多，思路举了几种参考方式，也可以参考我群里 322273643 发的三本电子书：

《精通 MATLAB 金融计算》

《金融数量分析：基于 MATLAB 编程》

《量化投资——以 MATLAB 为工具》

最后一问就是写建议，关于行业的发展，投资只是其中一部分，更多的还是政府的扶持。