

报名序号： 5766

赛题题目：光伏建筑一体化板块指数发展趋势分析及预测

# 基于多模型的 BIPV 一体化指数序列分析方法

## 摘 要

光伏一体化板块是目前一个炙手可热的领域，有着重大的商业价值。为了对一体化板块进行时间序列意义上的建模，参考多模型结合的模式解决实际问题。

对于问题一，利用板块指数求解出**一体化板块**，将 37 支股票归一化处理。然后利用时间序列分析中常用的移动平均模型进行求解，可以绘制出相应的曲线。

对于问题二，需要在问题一的基础上进行改进。这里采用三种改进模型：广义线性移动平均模型、移动平均-指数平滑模型和移动平均-ARIMA 模型进行对比分析。其中，移动平均-指数平滑模型在 20 日线上表现较好，而 ARIMA 模型更能反映大体趋势。提出的**融合式移动平均方法**具有一定可推广性。

对于问题三，在获取上证指数以后可以利用一些常用统计手段进行分析。使用皮尔逊相关系数和方差分析是我们的常见手段。**相关性分析模型**在统计学中比较常用，可以形成基本流程。

对于问题四，风险评价问题使用**层次分析法-蒙特卡洛模拟综合风险评估模型**得出了相关的风险系数等级并排序，按照排序结果再将股票分配抽象为一个线性规划问题，便于求解。

最后，问题五中对整个光伏建筑一体化市场进行了分析，结合前面的结论给出了一些思考，有一定参考价值。针对每个问题的多模型对比，使得讨论结果更可靠。

**关键词：**一体化板块，融合式移动平均方法，相关性分析模型，层次分析-蒙特卡洛模拟综合风险评估

## 1. 问题重述

### 1.1 问题背景

国家《第十四个五年规划和 2035 年远景目标纲要》中提出，将 2030 年实现“碳达峰”与 2060 年实现“碳中和”作为我国应对全球气候变暖的一个重要远景目标。光伏建筑一体化(BIPV)，是应用太阳能发电的一种新概念，简单地讲就是将太阳能光伏发电方阵安装在建筑的围护结构外表面来提供电力。它是充分利用工业建筑、公共建筑屋顶等资源实施分布式光伏发电工程，它对我国实现“碳达峰”和“碳中和”起到重要作用。目前已有北京、天津、上海等 31 地发布光伏建筑一体化未来三至五年相关政策，这将对光伏建筑一体化相关上市企业的发展带来一定影响。

### 1.2 问题的重述

问题一：利用附件一数据，给出 5 日、10 日、20 日等的 BIPV 板块指数的移动平均线模型，并绘制 2019 年 4 月 1 日至 2021 年 4 月 30 日该板块指数的移动平均线。

问题二：利用 2021 年 5 月 6 日至 2021 年 5 月 28 日数据对该模型进行修正，利用修正后的模型对 5 月 28 日后的 20 个交易日的日移动平均线，3 周的周移动平均线、2 个月的月移动平均线。

问题三：利用 2019 年 4 月 1 日至 2021 年 5 月 28 日数据，以两个月为一个时间段，对上证指数和 BIPV 板块指数进行相关性分析。

问题四：建立合理模型给出 BIPV 板块 37 支股票 2021 年 6 月投资风险由低到高的排序，并假定投资者持有 100 万人民币，全部用于该板块的投资，请给出 2021 年 6 月份的最优投资方案。

问题五：给相关部门写一份不少于 1500 字关于我国 BIPV 行业未来发展趋势的报告。

## 2. 问题分析

1. 对于问题一，这个问题很明显使用时间序列中的移动平均模型。我们分别取步长为 5 日、10 日、20 日，可以很快绘制出对应平滑曲线。难点在于板块一体化指数的构建，需要使用怎样的权重模型。这一点可以查一查资料，或者使用主成分分析将其降维。
2. 问题二是在问题一基础上的改进，对移动平均的改进可以引入权重因素来凸显历史数据的爆发时间对现实的影响。另外，还可以结合指数平滑和 ARIMA 等时间序列常用分析方法进行综合。
3. 问题三是一个相关性分析问题，相关性分析在统计学中主要以三种方法呈现，第一种是最为基本的相关系数法例如皮尔逊相关系数，第二种更加常用也更加精确，是方差分析法。这里只涉及两个量，所以用单变量方差分析。第三种比较少见，是信息熵法，这里主要使用前两种。
4. 问题四是一个风险评价和投资规划问题，首先风险不一定指亏损，我们将风险分为盈利的积极风险和亏损的消极风险分别进行排序，再进行综合判断。最后，我们利用综合判断的结果构建投资规划的线性规划模型。对于风险的评价我们同样有两种策略，对积极风险我们使用层次分析法从准则层逐渐到方案，而对消极风险我们试图用蒙特卡洛方法构建损失的经验分布并求解期望。
5. 问题五是对前面四个问题的综合，参考目前的一些文献即可。

## 3. 模型假设

- (1) 宏观政策是平稳的。

- (2) 股市公司决策者决策正确, 没有违规操作行为。  
 (3) BIPV 市场供求比较稳定。

## 4. 符号说明

表 1. 部分符号及解释

符号	释义
$N$	股票总天数
$m$	股票总数
$n$	n 天移动平均值
$i$	第 i 天( $1 \leq i \leq N-n+1$ )
$S_i$	第 i 天的市值
$A_i$	第 i 天的板块指数
$F_{n+i}$	第 i 个移动平均值

若未列入表 1 则以文中具体释义为准。

## 5. 模型的建立与求解

### 5.1 问题一

#### 5.1.1 板块指数的计算

- (1) 以基准日 2019 年 4 月 1 日作为基准日, 计算其总市值, 将其等价为 1000 点。  
 (2) 市值计算方法:

$$(1). S_i = \frac{P_1 W_1 + P_2 W_2 + \dots + P_m W_m}{m}$$

其中  $m$  为股票总数,  $P_x$  为第  $x$  只股票的价格,  $W_x$  为第  $x$  只股票的权重, 此处我们设定为 1 (相等权重)。

- (3) 板块指数=当前市值/基准日市值\*1000.

#### 5.1.2 移动平均线

根据上述 2019 年 4 月 1 日至 2021 年 4 月 30 日所求出的板块指数  $A_i$  (第  $i$  天), 我们可以计算出  $N-n+1$  个移动平均值  $F_{n+i}$  ( $N$  为总天数,  $n$  代表是  $n$  日的移动平均线模型,  $1 \leq i \leq N-n+1$ )

$$(2). F_{n+i} = \frac{A_i + A_{i+1} + \dots + A_{n+i-1}}{n}$$

分别取  $n=5, 10, 20$  日时我们可以做出图 1 所示的平滑曲线图:

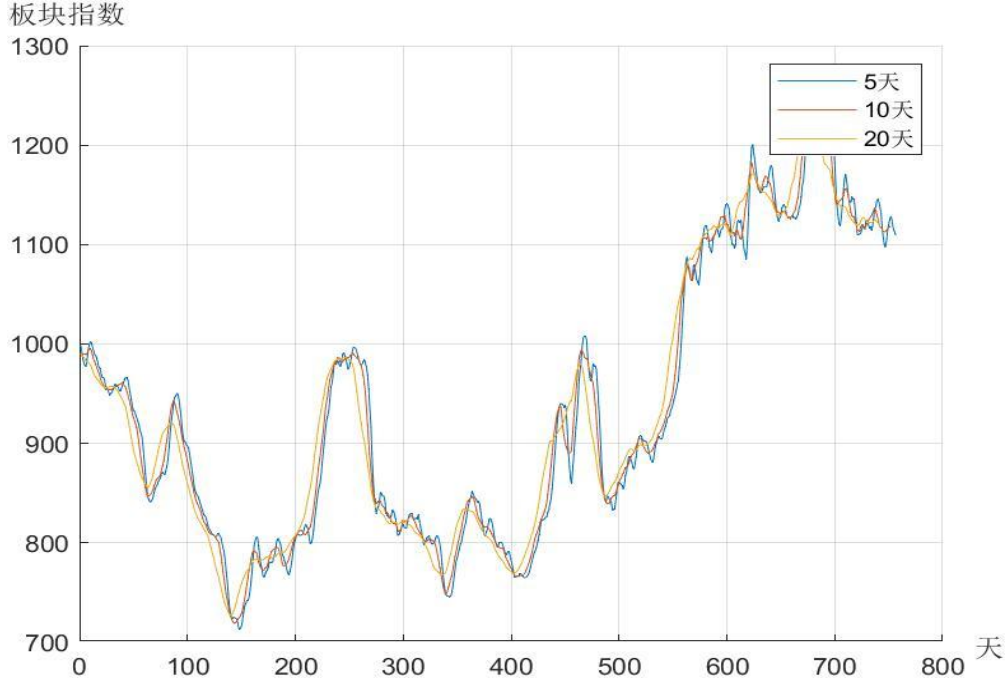


图 1. 移动平均曲线图

可以看到，三条曲线趋势相近，当时间步长逐渐拉长时曲线越平滑。

## 5.2 问题二

对于问题二，我们提出三种改进模型：

### 5.2.1 广义线性移动平均模型

在(2)式中可以看到，对于移动平均法而言此前  $i$  步内每一项权重相等（因为是普通平均），没有考虑到时间因素。事实上，类似于 LSTM 在 RNN 上所做的改进，当时间距离当前时间点越远则对当前影响越小，故我们考虑为每一项加上权重：

$$(3). F_{n+i} = \frac{w_1 A_i + w_2 A_{i+1} + \dots + w_n A_{n+i-1}}{w_1 + w_2 + \dots + w_n}$$

其中  $w$  的定义为与当前时间间隔的倒数。

那么，我们重新定义权重，并为原始模型加上白噪声作为随机扰动以增强模型的泛化能力，模型可以重写为：

$$(4). F_{n+i} = \sum_{j=1}^n \beta_j A_{j+i} + \varepsilon_t$$

我们使用的白噪声是满足正态分布的一个样本，即：

$$(5). \varepsilon_t \sim N(\mu, \sigma^2)$$

样本期望与方差为原来模型的预测误差的均值和方差。对 20 日线进行改进模型的测试实验，结果如图 2 所示：

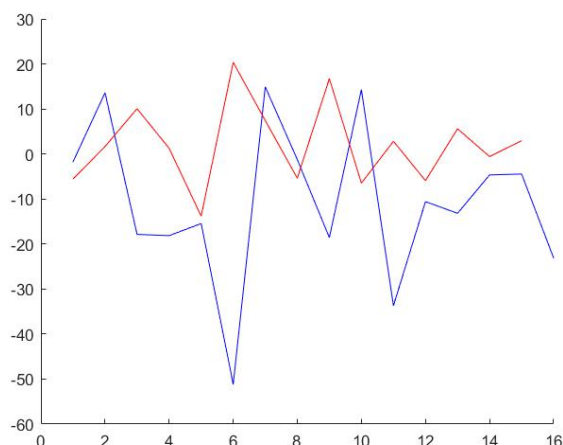


图 2. 广义移动平均模型在 20 日平均线上的测试

可以看到，尽管对模型进行了一些改进，原始数据和预测数据的偏差仍然较大。这可能是由于我们加入了过强的噪声导致。为了削弱不确定因素造成的影响，我们改进了另一个模型：

### 5.2.2 移动平均-指数平滑模型：

我们在移动平均法的基础上加入指数平滑法。指数平滑法同样是一个递推模型，基本形式如下：

$$(6). M_t = \alpha M_{t-1} + (1 - \alpha) y_t$$

以第一问的移动平均值作为输入，在此基础上使用指数平滑法，分别取 $\alpha$ 为 0.2, 0.5, 0.8，可以得到针对 20 日的图像如图所示：

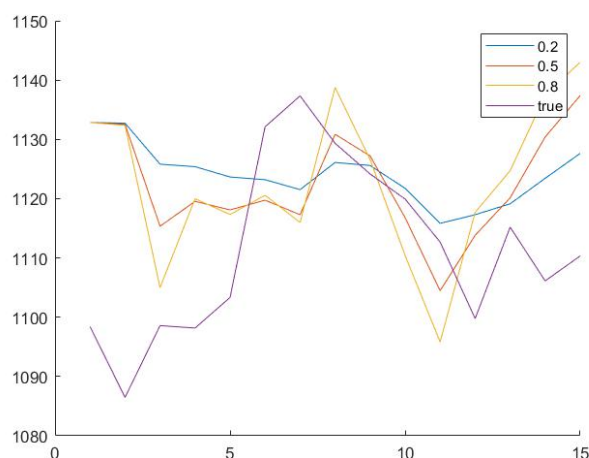


图 2. 20 日平滑曲线的移动平均-指数平滑结果

周平均和月平均并非简单的把步长拉大，因为即使是拉大了步长，60 日的日平均和两个月的月平均仍然是有差别的。在这里，周平均指的是只需要使用每周最后一日的数据，然后步长为 3；月平均即使用每月最后一日的数据，步长为 2 即可。

经对比分析，当 $\alpha=0.8$  时曲线会比较贴合真实情况。我们使用 $\alpha=0.8$  进行分析计算，得到的三条平均线如下：

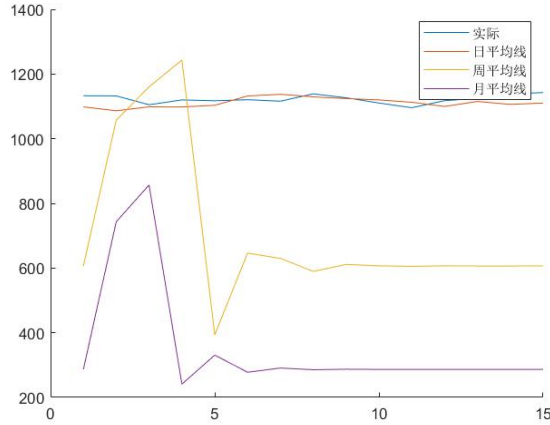


图 3. 基于移动平均-指数平滑模型求解的平均线

很显然，对于日平均线模型结果有显著提升，但月平均和周平均偏差仍然较大。所以我们进行了进一步改进：

### 5.2.3 移动平均-ARIMA 综合模型

基于此，我们对移动平均法进行了进一步的改进。考虑到 ARIMA 模型也可以对时间序列进行建模，我们使用 ARIMA 模型进行模型改良。若使用时间序列 ARIMA 模型来进行预测，我们可以预测接下来 20 日内的一体化指数[1]。

ARIMA 模型包含 3 个部分，即自回归（AR）、差分（I）和移动平均（MA）三个部分。对其每一个部分，都有其递推公式定义。

（1）自回归模型：

对于  $p$  阶自回归模型（AR），其递推公式形如：

$$(7). y_t = \mu + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t$$

（2）移动平均模型：

对于  $q$  阶移动平均模型（MA），其递推公式形如：

$$(8). u_t = \varepsilon_t + \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \nu$$

（3）差分模型：

$$(9). \nabla^{(1)} y_t = y_t - y_{t-1}$$

$$\nabla^{(d)} y_t = \nabla^{(d-1)} y_t - \nabla^{(d-1)} y_{t-1}$$

即由自回归模型阶数  $p$ 、差分阶数  $d$  和移动平均阶数  $q$  就可以确定 ARIMA 模型的基本形式，我们将其简记为 ARIMA( $p, d, q$ )

$$(10). y_t = \sum_{i=1}^p \alpha_i \nabla^{(d)} y_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \varepsilon_t$$

这一模型具有一些特有性质，例如自相关性、白噪声和平稳性。为了探索模型的最优参数(p,d,q)，我们通过 AIC 准则寻优，对 2 个月平滑曲线的数据进行了建模。我们最终选取的模型为 ARIMA(6,1,5)。

图 4 为 ARIMA(6,1,5)的测试结果。第一幅图是每个点的残差图，第二幅图是残差的统计分布直方图，可以看到误差近似于分布在 0 左右，说明噪声确实是白噪声。第三幅图是自相关图，第四幅图是偏自相关图，它们常用于判断时间序列的平稳性。经观察，可以推知其满足平稳性。第五幅图是 QQ 图，用于描述理论分布和实际值的偏差，可以看到大部分点都分布在理论值附近（红线），说明它满足基本的统计规律（正态性）。第 6 幅图为预测结果，为了充分考虑到结果的可信度问题，我们还纳入了 95%置信区间进行分析。

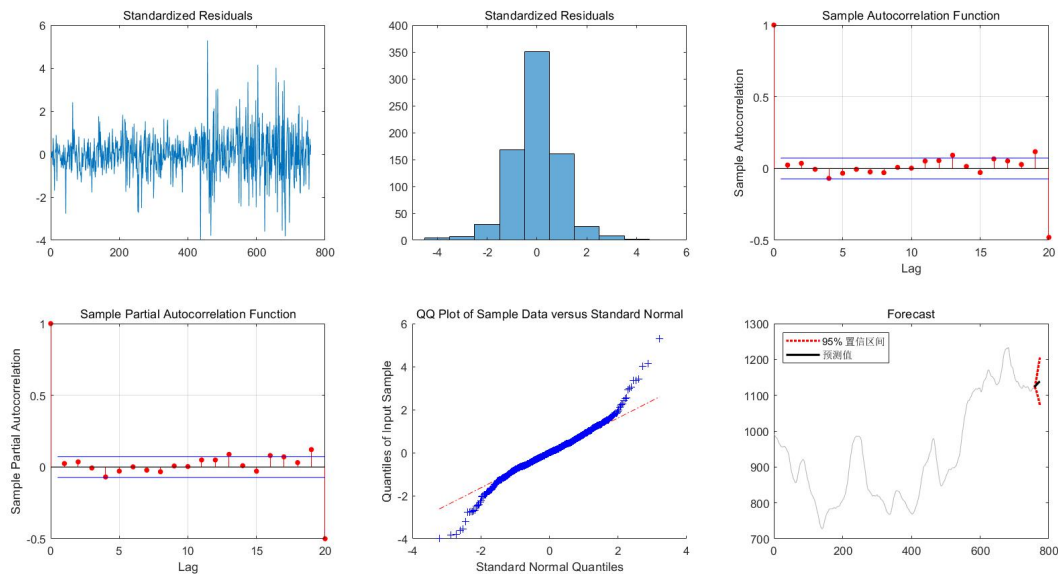


图 4. ARIMA(6,1,5)测试结果

可以说这一模型效果还比较好，我们利用 AIC 准则确定三条曲线中每条曲线的最合适参数并进行建模，将曲线图绘制如图 5 所示：

可以看到，ARIMA 的结果并没有表现出前两个模型的波动性而是趋势性。我们的结果能够从大概念上反映股市的走向与趋势，可以看到，3 周的周平均线和 2 月的月平均线都在上升，日平均可能是受到了扰动干扰，系统表现出一定的初始条件敏感特征；而周平均步长合适，跨度不大，在一定程度上更能反映整体的趋势。

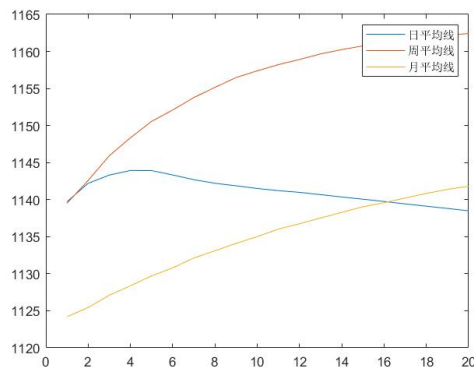




图 5. 基于移动平均-ARIMA 的测试结果

#### 5.2.4 对比分析

我们的模型改进方案是逐层递进式的。第一种方案是对移动平均的加权求和，但引入的随机成分影响太大导致结果不准确。而第二种改进方案更为合理，但在 20 日平均线上的测试误差高于 ARIMA 模型。ARIMA 模型是时间序列中的常用模型，主要是参数寻优的过程比较繁琐。

#### 5.3 问题三

我们从网络上收集得到上证指数，随后对其进行相关性分析。由于上证指数中两个月为一期的话通常是取两个月的最后一天，那么我们采取同样的方法：

表 2. 上证指数与光伏一体化板块指数

日期	2019-5	2019-7	2019-9	2019-11	2020-1	2020-3	2020-5	2020-7	2020-9	2020-11	2021-1	2021-3	2021-5
上证指数	1059.2	783.449	1018.11	1707.39	1477.44	797.396	473.954	1206.56	1055.07	2102.29	1464.15	1055.78	965.382
一体化指数	1358.80	853.203	782.399	1127.506	1134.875	717.626	870.371	964.296	912.063	1086.73	1125.725	1131.867	1141.906

##### 5.3.1 皮尔逊相关系数

皮尔逊相关系数是检验两个数据列之间统计关系最直接的方法。其定义为

$$(11). r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \sqrt{\sum_{i=1}^n y_i^2 - n \bar{y}^2}}$$

这一统计结果也是检验线性回归方程拟合程度的检验统计量。经求解，上证指数和一体化板块指数的皮尔逊相关系数为 0.4759。这一结果并不高，说明二者有一定相关性但并不那么强烈。我们来进行进一步分析：

##### 5.3.2 方差分析

方差分析的基本思想是：用于两个及以上的样本均数差别的显著性检验。通过分析研究不同变量的差异对总变异的贡献大小，确定控制变量对结果的影响力大小。它的使用条件是样本服从正态分布，这一点经过 K-S 检验是通过的。

方差分析的检验统计量 F 服从自由度为(s-1,n-s)的 F 分布，我们取显著度为 0.05，若概率值大于 0.05 则接受原假设“H0:两样本无显著差异”。实验结果表明，我们的样本统计量计算得到 F=1.3，概率为 0.2648>0.05，故二者确实差异不明显。

$$(12). F = \frac{(n-s)S_A}{(s-1)S_E} \sim F(s-1, n-s)$$

图 6 为两样本的箱线图：

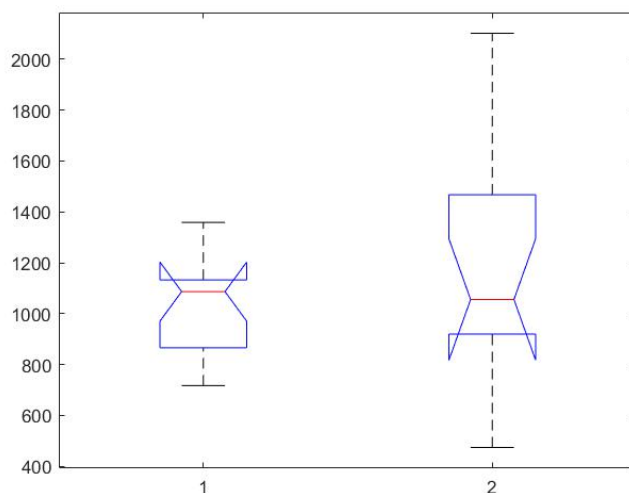


图 6. 板块一体化指数和上证指数的箱线图

图 6 中我们可以看到，二者的均值大致相当，但是上证指数的上四分位点和最大点远大于一体化指数，而最小点则小于一体化指数。这说明上证指数具有更强的不确定性因素在其中。

经检验，结果通过事后多重比较，结果可接受。

#### 5.4 问题四

为了量化描述购买股票时的风险因素，我们采用以下两种模型进行对比求解：

##### 5.4.1 基于层次分析法的风险评价

首先，我们的目标层非常确定，就是投资风险。而与投资风险有关的准则层，分为收益与亏损风险。进一步进行细化，我们会考虑收盘价与开盘价的差  $I_1$  作为历史收益数据，这是最可考的。其次，最高价与开盘价的差额  $I_2$  意味着收益的最大可能，最低价与开盘价的差额  $I_3$  是亏损不超过多少，这二者次要重要。而开盘价  $I_4$  和目前的底金有关，他决定你能买多少股票，所以放在第四位。

构造好变量以后，我们考虑通过层次分析法构建评价准则模型。层次分析法是一种解决多目标的复杂问题的定性与定量相结合的决策分析方法。该方法将定量分析与定性分析结合起来，用决策者的经验判断各衡量目标之间能否实现的标准之间的相对重要程度，并合理地给出每个决策方案的每个标准的权数，利用权数求出各方案的优劣次序，比较有效地应用于那些难以用定量方法解决的课题[2,3]。

层次分析法的第一步是考虑这些指标的重要程度并进行比较排序，这一步带有一定主观性，我们的想法是构造的四个变量  $I_1 > I_2 > I_3 > I_4$ 。这样一来就可以构建矩阵：

$$(13). \begin{cases} a_{ij} = \frac{1}{a_{ji}} \\ a_{ij} = a_{ik} a_{kj} \\ a_{:1} = [1, 3, 4, 5] \end{cases}$$

其中，元素值越大表示两两相比第一项更重要的程度。定义一致性指标：

$$(14). CI = \frac{\lambda - n}{n - 1}$$

一致性指标越接近 0 则一致性越好。那么为了衡量 CI 的相对大小，削弱主观因素的影响，引入随机一致性指标 RI（常数表格）。然后计算：

$$(15). CR = \frac{CI}{RI}$$

若 CR 小于 0.1，则通过检验。最终我们的权重矩阵结果为 CR=8.9e-16，通过检验，各指标的权重系数分别为 0.385, 0.308, 0.231, 0.077。

那么这 37 支股票的风险系数分别如表 3 所示：

表 3. 层次分析法中的股票风险系数

代号	1	2	3	4	5	6	7	8	9	10
系数	0.302 05283 4	0.263 25354 2	0.470 84789 4	0.335 44141 7	0.594 46971 6	0.419 63681 5	0.453 00651 3	0.412 51082 2	0.386 72640 4	1.009 01710 6
代号	11	12	13	14	15	16	17	18	19	20
系数	0.828 74962 2	0.537 30986 5	0.677 52554 8	2.071 02154 8		0.670 36889 9		1.008 48098 9		0.918 64987 8
代号	21	22	23	24	25	26	27	28	29	30
系数		0.440 39253 6		0.850 03054 4	1.144 16049 5				0.330 34772 2	1.405 59151 8
代号	31	32	33	34	35	36	37			
系数	0.478 46832 6	1.091 28572 2	0.600 95157 8	1.108 15478 5	1.036 97096 1	0.847 59426 1	0.939 38183 1			

#### 5.4.2 蒙特卡洛模型的风险评估

另外，为了衡量其风险，我们考虑使用蒙特卡洛模拟的方法做出一个损失率的概率分布，再来求损失的期望。我们设计的算法流程如下[4]：

算法：蒙特卡洛模型应用于风险评估

- (1) . 对股价进行差分以后求差分与原始值的商。
- (2) . 对偏差进行排序。
- (3) . 由于只能持有 5 支股票，所以

$$(16). d = \frac{\max \delta - \min \delta}{n} k$$

- (4) . 对每个日期进行统计，得到经验分布
- (5) . 获得损失期望

最终我们求解的期望分数如表 4：

表 4. 蒙特卡洛模型求解的风险系数

代号	1	2	3	4	5	6	7	8	9	10
系数	0.001 54194 3	0.001 90261 6	0.001 20794	0.000 28121 9	0.000 46626 5	0.000 44894 7	0.000 47072 6	0.001 35271 4	0.000 92274 4	0.000 76603 4
代号	11	12	13	14	15	16	17	18	19	20
系数	0.001 00931 4	0.000 58179 9	0.000 33975 6	0.000 70460 9	0.000 45342 3	0.000 58391 5	0.003 33228 2	0.001 36442 9	0.001 49501 8	0.003 06812
代号	21	22	23	24	25	26	27	28	29	30
系数	0.000 56822 3	0.000 42749	0.000 83806 5	0.000 94369 7	0.001 36364 3	0.000 68146 7	0.000 62731 8	0.000 84399 7	0.001 01193 3	0.002 16147 1
代号	31	32	33	34	35	36	37			
系数	0.001 37904 8	0.001 79753 8	0.000 62926 3	0.000 17397 5	0.001 64548 8	0.000 57695 5	0.000 54466 7			

#### 5.4.3 投资方案的设计

从蒙特卡洛模拟中我们可以看到，风险最小的前 5 支股票分别为 34、4、13、22、6 五支股票，而在层次分析法的结果中，结果既有风险又有收益，我们将其理解为收益的系数，因为风险不一定只能指损失风险，这里可以指盈利风险。层次分析法中收益率最大的前 5 支股票为 26、30、27、34、35。

综合二者的关系来看，我们可以选择的前五支股票为 34、22、30、37、13。我们选择的原因在于，同时考虑盈利风险和亏损风险，表 3 中盈利风险较高者优先购买，而表 4 中盈利风险较低者优先购买，为了综合二者，我们考虑引入新的量并称其盈亏风险：

$$(17). d = \frac{d_1 - \min(d_1)}{\max(d_1) - \min(d_1)} - \frac{d_2 - \min(d_2)}{\max(d_2) - \min(d_2)}$$

购买的数量可以抽象为一个线性规划问题：

$$(18). \begin{cases} \min J = \sum_{i=1}^5 d_i n_i \\ \sum_{i=1}^5 w_i n_i \leq 100 \end{cases}$$

其中  $w$  代表每只股票在六月份的开盘价，需要给定数据或从根据历史数据进行预测。以第二问中提供的方法进行预测，可以得到一个较为合理的方案。

## 5.5 问题五

### 光伏建筑一体化的行业现状和发展趋势

光伏建筑一体化（即 BIPV Building Integrated PV, PV 即 Photovoltaic）是一种将太阳能发电（光伏）产品集成到建筑上的技术,不同于光伏系统附着在建筑上(BAPV: Building Attached PV)的形式[5]。光伏建筑一体化可分为两大类：一类是光伏方阵与建筑的结合。另一类是光伏方阵与建筑的集成。如光电瓦屋顶、光电幕墙和光电采光顶等。说的通俗一点 BIPV 就是将光伏组件融入到建筑物的设计当中,使建筑物具有光伏发电的附加功能。BIPV 建筑物的照明、中央空调、内部用电等都可以由光伏设施产生的电能所满足。

光伏建筑一体化有很多优点。随着新能源的不断发展和城市节能减排、绿色环保需求的日益增加,太阳能光伏建筑一体化越来越成为太阳能应用发电的新潮流。传统的建筑室温调节依靠空调机和燃煤,虽然能够暂时提供舒适的环境,但是其耗能大,并且污染环境。太阳能光伏建筑一体化产生的是绿色能源,是应用太阳能发电,不会污染环境。太阳能是最清洁并且是免费的,开发利用过程中不会产生任何生态方面的副作用。它又是一种再生能源,取之不尽,用之不竭。并且不占用土地,光伏阵列一般安装在闲置屋顶,这对于昂贵的城市土地来说是十分友好的[6]。

当然其也存在一些缺点,制约着其发展。光伏建筑一体化的成本比较高,而且受天气影响大,波动性大,太阳不是一天 24 小时都有,其在粗存电量等方面还有待解决。在顶层设计方面,“十四五”规划及碳中和目标利于 BIPV 的发展。我国“十四五”规划中,强调碳中和及绿色生态的建立,倡导更合理的能源配置,未来将进一步推动绿色能源及节能减排的发展。另外,我国提出了“2030 年碳达峰”和“2060 年碳中和”的目标,BIPV 可以有效利用太阳能代替其他高碳排放能源,前瞻认为,在未来将得到长足的发展。

近两年,我国光伏建筑一体化在波动中上升,最近一段时间处于波动下降后维持一个比较平稳的状态。观察股市情况可以发现,我国光伏一体化虽然处于不稳定的状态但是总体而言呈现上升的趋势。当然,偶有下降也是市场的一些突发情况或者偶然因素,是在所难免的。

BIPV 项目的建设有着诸多优势。首先,国家关于「碳中和」已经有了明确的顶层设计,所以 BIPV 项目的第一个优势就是符合国家产业政策。其实关于绿色建筑,国家就有了相应的产业政策。2020 年 7 月,住建部、发改委等 7 部委局联合发布《绿色建筑创建行动方案》,要求到 2022 年,当年城镇新建建筑中绿色建筑面积占比要达到 70%。既然属于国家提倡的绿色产业,那么在土地审批和资金来源方面,BIPV 项目就更容易获得一些政策方面的鼓励和倾斜。

除了符合产业政策外,BIPV 项目还具有众多优势。比如通过光伏发电,可以有效降低建筑碳排放,满足环保指标,降低建筑供热和室内空调冷负荷、减少能耗。其次 BIPV 建筑项目无需安装蓄电池,所以开发商无需大幅增加建造成本。并且在建筑物投入使用后 BIPV 项目也不受蓄电池储能的限制,可以实现发自用、余量上网。上网的多余电能和获得的碳排放指标都可以增加开发商后续的收入,可以加快开发商回收成本。

第三、BIPV 地产项目的开发具有一定的专业性,涉及到整个项目电力系统与光伏系统的链接,以及发电系统与电网的对接。相对于传统的地产项目,在设计、施工的时候对相关方都提出了更高的要求。所以一些小型的地产开发商或没有相关技术优势的企业将被排除在 BIPV 项目之外。而这对于一些大型的地产开发商和专业的施工机构是一个难得的发展机遇,可以享有更多的市场份额。

我国新增建筑面积不断扩大,BIPV 市场潜力巨大。其中,中衡设计、广田集团等股

份近年来异军突起，占据了光伏建筑一体化市场中很大一块份额。我们定义，以板块指数=当前市值/基准日市值\*1000 作为权值，将市场上现存较强的 37 支股票综合分析可以定义光伏建筑一体化指数来描述这一市场整体的趋势。

从 5 日-10 日-20 日的日移动平均线来看，这一市场仍充满一定的不稳定性，但从整体趋势上大体向好。我们根据移动平均法、指数平滑法、ARIMA 模型等多种时间序列模型分析预测，得出相关预测结果，说明结果可预见。

而从这一板块和上证指数的关系上来看，我们通过统计学方法例如皮尔逊相关系数和方差分析，分析了这一板块和上证指数之间的统计学关系。我们发现，二者之间的关联性并不明显，所以这一板块相对来说具有更强的独立性。

从投资风险的角度来看，我们将风险分为盈利的积极风险和亏损的负面风险，并基于层次分析法和蒙特卡洛模拟方法获得了相应的风险系数。综合评估，我们可以得到最终的盈亏风险，对现有的 37 支股票进行了排序。最终结果对投资者购买股票具有一定参考意义。

在这些股票中，除了隆基股份、南玻这样光伏上游产业公司外，还有启迪设计这样的装配式建筑设计企业。当然，更广泛的说具有 BIPV 项目开发能力的大中型地产商和施工单位也应该被算在光伏建筑一体化概念的范围内。

投资 BIPV 项目的时候我们也要注意到这个产业所具有的一些风险。说到底，光伏建筑一体化的载体还是建筑，是加入光伏元素的地产项目。整个项目的成败仍然取决于出租率和销售情况，目前地产业的大环境并不像前些年那样火爆。所以虽然“碳中和”会在行政审批和现金流回笼方面带来一定的优势，但是大规模建设期过后，想要通过 BIPV 项目获得超额收益是比较困难的。对于这一点，投资者应该给予充分的考虑。

## 6. 模型的评价与推广

### 6.1 模型的优点

1. 综合考虑多方面因素，结合现实中较复杂的各类情况来进行建模，其中对模型进行了多次修正，使之较符合实际情况。对数据也进行了充分使用与深加工处理。
2. 使用多种模型结合，互相对比补充，参考了大量工程实际的相关资料，真实可靠，有一定参考价值。
3. 结合生产实际，对光伏建筑一体化产业进行了较为深入的探究。

### 6.2 模型的缺点

1. 平均线还存在一定的误差，另外存在趋势与细节的不能双重保证。
2. 未能使用更高精度的模型对问题进行建模。

### 6.3 模型的推广

本次我们建立的模型还可以使用于其他时间序列问题中，并且在模型中时间序列 ARIMA 是股票预测问题中常用的一类模型，对于“准则+方案”类问题层次分析配合规划问题也能够对问题进行合理的建模求解。

**参考文献:**

- [1]许立平, 罗明志. 基于 ARIMA 模型的黄金价格短期分析预测[J]. 财经科学, 2011, 2011(001):26-34.
- [2]郭金玉, 张忠彬, 孙庆云. 层次分析法的研究与应用[J]. 中国安全科学学报, 2008, 18(005):148-153.
- [3]金菊良, 魏一鸣, 丁晶. 基于改进层次分析法的模糊综合评价模型[J]. 水利学报, 2004, 000(003):65-70.
- [4]Sufianti. 期望损失分析:基于回归方法[J]. 2012.
- [5]李芳, 沈辉, 许家瑞,等. 光伏建筑一体化的现状与发展[J]. 电源技术, 2007(8):659-662.
- [6]段青. 太阳能光伏建筑一体化应用现状及发展趋势[J]. 房地产导刊, 2017, 000(014):204.



## 附录

### 附录

#### 介绍：部分 Matlab 程序

```
clear
load('data.mat');
P=zeros(777,37);
for i=1:37
    [L,~]=size(data{i});
    if L>776
        P(:,i)=data{i}(end-776:end,4);
    else
        P(:,i)=[zeros(777-L,1);data{i}(1:end,4)];
    end
end

S=sum(P,2);
n=37-sum(~P,2);
S=S./n;
S=S./(0.001*S(1));
S_=S(end-15:end);
S=S(1:end-16);
save('S','S','S_')
```

```
clear
load('S.mat')
S=[S;S_];
n=[5,10,20];
r=zeros(3,4);
hold on
f1=ma(S,n(1));
f2=ma(S,n(2));
f3=ma(S,n(3));
plot((1:length(f1))+n(1),f1,'r-',(1:length(f2))+n(2),f2,'b-','g-');
% plot(S,'LineWidth',1)
xlabel('天');

ylabel('一体化指数')

legend('5 日','10 日','20 日');
hold off
```

```

load('S.mat')
w=(1:20)./210;
f=zeros(16,1);
S=diff(S);
S=[S(end-19:end);zeros(16,1)];
f(1)=sum(S)./20;
S(21)=f(1);
for i=2:16
    f(i)=w*S(i:i+19)+normrnd(0.1452,16.1463);
    S(i+20)=f(i);
end
hold on
plot(1:16,f,'b');
plot(1:15,diff(S_),'r');

```

```

load('S.mat')
n=2;d=1;
S=syue;
S=short(S,d);
S_=S(end-14:end);
S=S(1:end-15);
w=(1:n)./((n+1)*n/2);
f=zeros(length(S_),1);
S=[S(end-n+1:end);zeros(length(S_),1)];
f(1)=sum(S)/20;
S(1+n)=f(1);
for i=2:length(S_)
    f(i)=w*S(i:i+n-1);%+normrnd(0.1452,16.1463)
    S(i+n)=f(i);
end
% f0=S_(1);
% S(1+n)=f(1);
for i=2:length(S_)
    f(i)=w*S(i:i+n-1);%+normrnd(0.1452,16.1463)
    S(i+n)=f(i);
end

yt=S;
%alpha=[0.2,0.5,0.8];
alpha=[0.8];
n=length(yt);m=length(alpha);
yhat(1,1:m)=yt(end);
for i=2:length(S_)

```

```

        yhat(i,:)=alpha*yt(i-1)+(1-alpha).*yhat(i-1,:);
    end
    hold on
    %plot(f)
    plot(yhat)
    function S1=short(S,d)
    n=length(S);
    n=floor(n/d);
    S1=zeros(n,1);
    for i=1:n
        S1(i)=S(i*d);
    end
end
end

```

```

function forData=ARIMA_gushi(f,order)
x=1:length(f);
Y=f;

% 通过 AIC, BIC 等准则暴力选定阶数

%[AR_Order,MA_Order] =
ARMA_Order_Select(Y,max_ar,max_ma,0);
AR_Order=order(1);
d=order(2);
MA_Order=order(3);
Mdl = arima(AR_Order,d,MA_Order);
EstMdl = estimate(Mdl,Y);

[res,~,logL] = infer(EstMdl,Y); %res 即残差

stdr = res/sqrt(EstMdl.Variance);
figure('Name','残差检验')
subplot(2,3,1)
plot(stdr)
title('Standardized Residuals')
subplot(2,3,2)
histogram(stdr,10)
title('Standardized Residuals')
subplot(2,3,3)
autocorr(stdr)
subplot(2,3,4)
parcorr(stdr)
subplot(2,3,5)
qqplot(stdr)

```

```
% 5.预测
```

```
step = 20; %预测步数为 24
```

```
[forData,YMSE] = forecast(EstMdl,step,'Y',Y); %matlab2018
```

```
及以下版本写为 Predict_Y(i+1) =
```

```
forecast(EstMdl,1,'Y0',Y(1:i));
```

```
lower = forData - 1.96*sqrt(YMSE); %95 置信区间下限
```

```
upper = forData + 1.96*sqrt(YMSE); %95 置信区间上限
```

```
subplot(2,3,6);
```

```
plot(Y,'Color',[.7,.7,.7]);hold on;
```

```
h1 =
```

```
plot(length(Y):length(Y)+step,[Y(end);lower],'r:','LineWidth',2);subplot(2,3,6);
```

```
plot(length(Y):length(Y)+step,[Y(end);upper],'r:','LineWidth',2),hold on;
```

```
h2 =
```

```
plot(length(Y):length(Y)+step,[Y(end);forData],'k','LineWidth',2);subplot(2,3,6);
```

```
legend([h1 h2],'95% 置信区间','预测值
```

```
','Location','NorthWest')
```

```
title('Forecast')
```

```
end
```

```
function loss=mtkl(price,k,m1)
```

```
delta=diff(price)./price(2:end);
```

```
delta=-delta(delta<0);
```

```
delta=sort(delta);
```

```
n=length(delta);
```

```
d1=max(delta);
```

```
d2=min(delta);
```

```
d=(d1-d2)/n*k;
```

```
m=ceil(n/k);
```

```
P=zeros(m,1);
```

```
x=1;
```

```
for i=1:n
```

```
    if delta(i)<=d2+d*x
```

```
        P(x)=P(x)+1;
```

```
    else
```

```
        x=x+1;
```

```

        end
    end
    P=P./n;
    x=(d2+d/2):d:(d2+m*d-d/2);
    loss=x*P*sum(delta(end-m1+1:end))/m1;
end

load('data.mat')
m=80;m1=5;n=5;
loss=zeros(1,37);
for i=1:37
    a=data{i};
    loss(i)=mtkl(a(end-m+1:end,4),n,m1);
    %loss(i)=mean(0.385*(a(:,4)-a(:,1))+0.308*(a(:,2)-a(:,
1))+0.231*(a(:,3)-a(:,1))+0.077*a(:,1))
end
loss

```