

券商研报对公司股票走势的影响和投资策略

摘 要

本文运用事件研究法研究券商研报对公司股票走势的影响和投资策略。券商研报是指证券公司的研究人员对证券及相关产品的价值，或者影响其市场价格的因素进行分析，所作出的研究报告。通过研究券商研报和外界因素对证券公司股票走势的影响，提出明确的投资策略。

对于问题一，我们通过主成分分析对从研报中提取指标，并用 PCA 算法模型在 SPSS 进行数据降维并选用 KMO 和 Bartlett 球形度检和进行因子分析，提取出 2-3 个主要因子，确定为研报的特征指标

对于问题二：我们建立在第一题的基础上，建立了 AR 模型、MA 模型、ARMA 模型和 ARIMA 模型，借助 ACF 图和 PACF 图、拖尾和截尾，最终确定建立了 ARIMA 模型，然后对数据进行差分，选用 BIC 贝叶斯信息准则平衡预测误差和参数个数，最终确定得出股票预测趋势图并制定明确的投资策略。

对于问题三：我们根据股票预测分析的假设条件，建立了 ARIMA 模型，并对模型进行了合理的理论证明和推导，分析得出股价短期内变化的规律。我们借助同花顺软件，选取了 2020 年 6 月至九月间十只股票的收盘价并选用新冠疫情爆发初期的数据进行数据，进行对比，最后得出各股票的股价走势图并分析短期内股价变化，总结概况对股价的影响。

对于问题四：我们建立在第二题的基础上，综合第三题的影响分析，通过分析总结研报、外界环境因素对股票走势的影响，提出相关新的投资策略。

最后，针对所建立的模型进行更为普遍的推广，并就已建立模型进行评价与改进。

关键词

特征因子、ARIMA 模型、股票走势、投资策略

目录

| | |
|--------------------------------|----|
| 一、 问题重述 | 1 |
| 二、 问题分析 | 1 |
| 2.1. 问题一的分析 | 1 |
| 2.2. 问题二的分析 | 1 |
| 2.3. 问题三的分析 | 1 |
| 2.4. 问题四的分析 | 1 |
| 三、 模型假设 | 1 |
| 四、 模型的建立与求解 | 2 |
| 4.1. 问题一模型的建立与求解 | 2 |
| 4.1.1. 券商研报提取指标 | 2 |
| 4.1.2. 数据降维——PCA 模型的建立 | 2 |
| 4.2. 模型的求解 | 4 |
| 4.2. 问题二模型的建立与求解 | 7 |
| 4.2.1. 数据处理 | 7 |
| 4.2.2. 建立 ARIMA 模型 | 8 |
| 4.2.3. 模型的求解 | 10 |
| 4.2.4. 问题二结果——投资策略 | 12 |
| 4.3. 问题三模型的建立与求解 | 12 |
| 4.3.1. 数据筛选——股票预测分析的假设条件 | 12 |
| 4.3.2. ARIMA 模型的建立 | 13 |
| 4.3.3. 模型的求解 | 13 |
| 4.3.4. 问题总结 | 19 |
| 4.4. 问题四模型的建立与求解 | 19 |
| 五、 模型的评价、改进 | 20 |
| 5.1. 模型的优点 | 20 |
| 5.2. 模型的缺点 | 20 |
| 5.3. 模型的改进 | 20 |
| 六、 参考文献 | 21 |
| 附录 | 22 |

一、问题重述

券商研报是指证券公司的专业研究人员对其证券及其相关产品的价值，或者影响其市场价格的因素进行分析所作出的研究报告。基于券商研报，综合分析券商研报和外界环境因素对证券公司股票走势的影响，试建立相应的数学模型，根据相关资料，解决以下问题：

- 1、在大湾区指数的 30 支股票中选取 10 支股票的券商研报并提取特征指标。
- 2、对选取的 10 支股票，分析研报特征指标对股票走势的影响并提出明确的投资策略。
- 3、研究突发事件闪现、舆情等因素对选取的 10 支股票走势有什么影响。
- 4、分析证券机构的券商研报和外界因素对股票走势的影响，并修改第二题中的策略，提出新的投资策略。

二、问题分析

2.1. 问题一的分析

问题一是我们通过分析证券公司的券商研报，提取出其中的特征指标，这就需要懂得如何看懂券商研报，目前，市面上流通的研报都需要具有专业的金融知识才能看懂，此题的一大难点就是看懂研报，再有就是能够理解特征指标，特征指标可以简单的理解成对股票走势影响最大的指标，针对此问题，面对这种多变量的课题时，是否可以将特征指标等价于主成分，也就是通过主成分分析对从研报中选取出来的指标进行数据降维，提取出 2-3 个主成分，一个成分包含了二种及以上的指标的线性组合。

2.2. 问题二的分析

问题二是让我们通过提取出来的特征指标对股票未来趋势进行预测，并给出明确的投资策略，针对此问题，我们采取建立 ARIMA 模型，对股价未来走势进行建模预测。

2.3. 问题三的分析

问题三是让我们建立模型对突发事件的闪现、舆情、自然灾害等外界环境因素造成的股价影响进行分析，并预测股票走势，提出相应的投资策略。

2.4. 问题四的分析

问题四是结合问题三，并考虑外界研报的影响，对问题二的投资策略进行改进。

三、模型假设

- 1.假设经过一阶差分后股票行情实际数据趋于平稳，数据的平稳性高；
- 2.假设股票价格不会随投资者对各种因素的心理预期的变化而波动，或者波动不大；
- 3.假设只有疫情会对股票行情产生影响影响；

四、模型的建立与求解

4.1. 问题一模型的建立与求解

4.1.1. 券商研报提取指标

从券商研报中获取了特征指标，以塔牌集团的股市行情为例，我们从其券商研报中挑选了十个特征指标，以及数据（见下表）

| 年份 | 营业收入(百万元) | 增长率() | EBITDA(百万元) | 净利润(百万元) | 增长率() |
|----------|-----------|----------|-------------|-----------|--------|
| 2019 | 6890.7 | 3.93 | 2680.94 | 1733.48 | 0.6 |
| 2020 | 7046.66 | 2.26 | 2762.75 | 1782.15 | 2.81 |
| 2021E | 8570.34 | 21.62 | 3767.6 | 2698.79 | 51.43 |
| 2022E | 10303.3 | 20.22 | 3878.22 | 3152.22 | 16.8 |
| 2023E | 10988.78 | 6.65 | 4215.59 | 3463.18 | 9.87 |
| EPS(元/股) | 市盈率(P/E) | 市净率(P/B) | 市销率(P/S) | EV/EBITDA | |
| 1.45 | 7.28 | 1.29 | 1.83 | 3.61 | |
| 1.49 | 7.08 | 1.21 | 1.79 | 3.34 | |
| 2.26 | 4.68 | 1.08 | 1.47 | 1.55 | |
| 2.64 | 4.01 | 0.96 | 1.23 | 1 | |
| 2.9 | 3.65 | 0.85 | 1.15 | 0.33 | |

4.1.2. 数据降维——PCA 模型的建立

(1) 将原始数据按列组成 n 行 m 列矩阵 X

$$X = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{pmatrix} \quad (1)$$

(2) 数据标准化

对所有特征指标的数值进行中心化，也即去均值：这一步也称数据的标准化，将所有数据变成在 $(0,1)$ 区间内。

$$\overline{x_1} = \frac{1}{M} \sum_{i=1}^M x_1^i : \quad (2)$$

平均值公式：

利用上述公式 (2)，将样本数据的每一个值减去其对应的平均值，得到了如下的数据：

| 年份 | 营业收入(百万元) | 增长率() | EBITDA(百万元) | 净利润(百万元) | 增长率() |
|------|-----------|---------|-------------|----------|----------|
| 2019 | -1.00583 | -0.7565 | -1.12200 | -1.05859 | -0.76072 |
| 2020 | -0.9219 | -0.9368 | -1.00433 | -0.9967 | -0.65365 |

| | | | | | |
|--------------|----------------|----------------|----------------|---------------|----------|
| | 1 | 3 | | 1 | |
| 2021 E | -0.1020 3 | 1.1536 5 | .44096 | .16890 | 1.70187 |
| 2022 E | .83046 | 1.0024 8 | .60007 | .74549 | .02413 |
| 2023 E | 1.1993 1 | -0.4628 0 | 1.08531 | 1.1409 1 | -0.31161 |
| EPS(元/ 股) | 市 盈 率 (P/E) | 市 净 率 (P/B) | 市 销 率 (P/S) | EV/EBITD A | |
| - 1.05822 | 1.12704 | 1.18382 | 1.07728 | 1.13627 | |
| -0.99758 | 1.01085 | .73710 | .94903 | .94966 | |
| .16980 | -.38343 | .01117 | -.07695 | -.28752 | |
| .74591 | -.77266 | -.65892 | -.84643 | -.66766 | |
| 1.14009 | -.98180 | -1.27317 | -1.10293 | -1.13074 | |

(3) 求协方差矩阵 C

[1]设我们有 m 个 n 维数据记录，将其排列成矩阵 $X_{n,m}$ ，设 $C = \frac{1}{m}XX^T$ ，则 C 是一个对称矩阵，其对角线分别对应各个变量的方差，而第 i 行 j 列和 j 行 i 列元素相同，表示 i 和 j 两个变量的协方差。

用于衡量两个变量的总体误差，期望值分别为 $E[X]$ 与 $E[Y]$ 的两个实随机变量 X 与 Y 之间的协方差 $Cov(X,Y)$ 定义为：

$$Cov(a, b) = \frac{1}{m-1} \sum_{i=1}^m (a_i - \mu_a)(b_i - \mu_b) \quad (3)$$

假设我们只有两个变量分别是 a 和 b ，我们将他们组成一个矩阵 X ，例如（1）所示，然后由上述公式（3）得到一个协方差的矩阵，即：

$$\frac{1}{m}XX^T = \begin{pmatrix} \frac{1}{m} \sum_{i=1}^m a_i^2 & \frac{1}{m} \sum_{i=1}^m a_i b_i \\ \frac{1}{m} \sum_{i=1}^m a_i b_i & \frac{1}{m} \sum_{i=1}^m b_i^2 \end{pmatrix} = \begin{pmatrix} Cov(a, a) & Cov(a, b) \\ Cov(b, a) & Cov(b, b) \end{pmatrix} \quad (4)$$

由上述公式（4），我们很容易可以知道假如我们有 m 个 n 维数据记录，将其排列成矩阵 $X_{n,m}$ ，设 $C = \frac{1}{m}XX^T$ ，则 C 是一个对称矩阵，其对角线分别对应各个变量的方差，而第 i 行 j 列和 j 行 i 列元素相同，表示 i 和 j 两个变量的协方差。

（4）求协方差矩阵 c 相对应的特征向量和特征值

利用线性代数中矩阵的知识，求协方差矩阵 c 相对应的特征向量 u 和特征值 λ （每一个特征值只对应一个特征向量）：

$$Cu = \lambda u \quad (5)$$

特征值 λ 会有 N 个，每一个 λ_i 对应一个特征向量 u_i ，将特征值 λ 按照从大到小的顺序排序，选择最大的前 k 个，并将其相对应的 k 个特征向量拿出来，我们会得到一组 $\{(\lambda_1, u_1), (\lambda_2, u_2), \dots, (\lambda_k, u_k)\}$ 。

（5）将原始特征投影到选取的特征向量上，得到降维后的新 K 维特征

这个选取最大的前 k 个特征值和相对应的特征向量，并进行投影的过程，就是降维的过程。对于每一个样本，对于每一个样本 X_i ，原来的特征是 $(x_{i1}, x_{i2}, \dots, x_{in})^T$ ，投影之后的新特征是 $(y_{i1}, y_{i2}, \dots, y_{ik})^T$ ，新特征的计算公式如下：

$$\begin{bmatrix} y_1^i \\ y_2^i \\ \vdots \\ y_k^i \end{bmatrix} = \begin{bmatrix} u_1^T \cdot (x_1^i, x_2^i, \dots, x_n^i)^T \\ u_2^T \cdot (x_1^i, x_2^i, \dots, x_n^i)^T \\ \vdots \\ u_k^T \cdot (x_1^i, x_2^i, \dots, x_n^i)^T \end{bmatrix} \quad (6)$$

对于我们的样本数据来说，每一个样本 X_i ，由原来的 $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ，变成了现在 $X_i = y_i$ ，达到了降维的目的。

4.2. 模型的求解

考虑到数据量较大，以及 PCA 算法模型在 SPSS 里有完善的处理模式，测出的数据更为准确，所以我们采取了 SPSS 进行数据的降维。

这里以塔牌集团为样本，进行模型的求解。

(1) 数据标准化

将数据导入 SPSS 中，我们得到一个样本数据集的标准数据集形式（图 1）

| Z营业 收入百 万元 | Z增长 率 | ZEBIT DA百 万元 | Z净 利润 百 | Z增 长率 _A | ZEP S元 股 | Z市 盈率 PE | Z市 净率 PB | Z市 销率 PS | ZEV EBIT DA | FAC 1_1 | FAC 2_1 |
|------------------|----------|--------------------|---------------|----------------|----------------|----------------|----------------|----------------|-------------------|------------|------------|
| -1.00583 | -.75650 | -1.12200 | -1.05859 | -.76072 | -1.05822 | 1.12704 | 1.18382 | 1.07728 | 1.13627 | -1.12894 | -.19277 |
| -.92191 | -.93683 | -1.00433 | -.99671 | -.65365 | -.99758 | 1.01085 | .73710 | .94903 | .94966 | -.99586 | -.33460 |
| -.10203 | 1.15365 | .44096 | .16890 | 1.70187 | .16980 | -.38343 | .01117 | -.07695 | -.28752 | .34487 | 1.59690 |
| .83046 | 1.00248 | .60007 | .74549 | .02413 | .74591 | -.77266 | -.65892 | -.84643 | -.66766 | .76597 | .06891 |
| 1.19931 | -.46280 | 1.08531 | 1.14091 | -.31161 | 1.14009 | -.98180 | -1.27317 | -1.10293 | -1.13074 | 1.01396 | -1.13844 |

图 1、样布数据的标准化结果 1

(2) 因子分析

在 SPSS 中没有主成分分析，需先对样本数据进行因子分析，在这里我们选用 KMO 和 Bartlett 球形度检验（Bartlett 球形检验用于检验相关阵中各变量间的相关性，是否为单位阵，即检验各个变量是否各自独立。）(图 2)

| 总方差解释 | | | | | | | |
|-------|----|--------------------|------------|---------|---------|--------|--------|
| | 成分 | 初始特征值 ^a | | | 提取载荷平方和 | | |
| | | 总计 | 方差百分比 | 累积 % | 总计 | 方差百分比 | 累积 % |
| 原始 | 1 | 8.389 | 83.895 | 83.895 | 8.389 | 83.895 | 83.895 |
| | 2 | 1.426 | 14.264 | 98.159 | 1.426 | 14.264 | 98.159 |
| | 3 | .173 | 1.730 | 99.889 | | | |
| | 4 | .011 | .111 | 100.000 | | | |
| | 5 | 8.760E-16 | 8.760E-15 | 100.000 | | | |
| | 6 | 1.395E-16 | 1.395E-15 | 100.000 | | | |
| | 7 | 2.236E-18 | 2.236E-17 | 100.000 | | | |
| | 8 | -1.488E-16 | -1.488E-15 | 100.000 | | | |
| | 9 | -2.453E-16 | -2.453E-15 | 100.000 | | | |
| | 10 | -7.013E-16 | -7.013E-15 | 100.000 | | | |
| 重新标度 | 1 | 8.389 | 83.895 | 83.895 | 8.389 | 83.895 | 83.895 |
| | 2 | 1.426 | 14.264 | 98.159 | 1.426 | 14.264 | 98.159 |
| | 3 | .173 | 1.730 | 99.889 | | | |
| | 4 | .011 | .111 | 100.000 | | | |
| | 5 | 8.760E-16 | 8.760E-15 | 100.000 | | | |
| | 6 | 1.395E-16 | 1.395E-15 | 100.000 | | | |
| | 7 | 2.236E-18 | 2.236E-17 | 100.000 | | | |
| | 8 | -1.488E-16 | -1.488E-15 | 100.000 | | | |
| | 9 | -2.453E-16 | -2.453E-15 | 100.000 | | | |
| | 10 | -7.013E-16 | -7.013E-15 | 100.000 | | | |

提取方法：主成分分析法。

a. 在分析协方差矩阵时，原始解与重新标度的解的初始特征值相同。

图 2、因子分析结果 1

通过因子我们得出十个变量的特征值，从图 2 可以看到特征值 ($\lambda_1=8.389$, $\lambda_2=1.426$, $\lambda_3=0.173$, ..., $\lambda_{10}=-7.013E-16$)，由上图 2 可以看到 λ_1 的累积贡献值是 83.895%， λ_1 和 λ_2 的累积贡献值是 98.159%，在主成分分析中，我们规定累积贡献值超过 85% 的时候，我们就选取与其相对应的特征值，而一个特征值对应着一个主成分，所以这个样本集对应了两个主成分，从图 3 我们可以进一步的验证这一点。

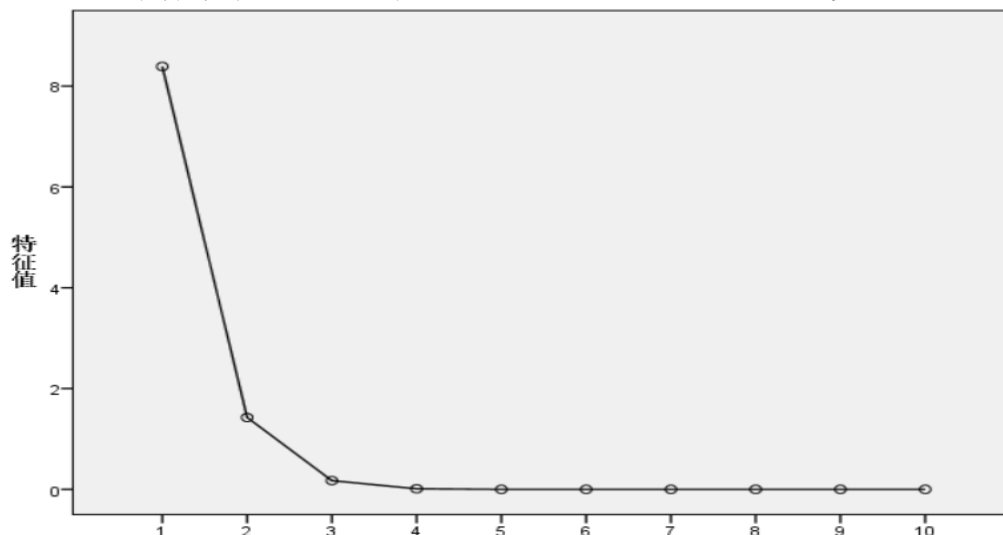


图 3、碎石图 1

从如图 3 的碎石图中，可以看到前两个特征值代表的斜线的斜率远远大于后面特征值，所以我们就说前 2 个特征值对样本数据的影响远远大于后面 8 个特征值，后 8 个特征值可忽略不计，得出结论此数据集的主成分数有两个。当知道有几个主成分后，我们需要知道一个主成分对应了哪几个指标，SPSS 给出了成分矩阵图和载荷图供我们查看（图 4，图 5）

| | 原始成分 | | 重新标度成分 | |
|---------------------|--------|-------|--------|-------|
| | 1 | 2 | 1 | 2 |
| Zscore: 营业收入(百万元) | .968 | -.242 | .968 | -.242 |
| Zscore: 增长率() | .621 | .724 | .621 | .724 |
| Zscore: EBITDA(百万元) | .995 | .016 | .995 | .016 |
| Zscore: 净利润(百万元) | .993 | -.110 | .993 | -.110 |
| Zscore: 增长率() | .450 | .860 | .450 | .860 |
| Zscore: EPS(元/股) | .994 | -.109 | .994 | -.109 |
| Zscore: 市盈率(P/E) | -1.000 | -.026 | -1.000 | -.026 |
| Zscore: 市净率(P/B) | -.966 | .237 | -.966 | .237 |
| Zscore: 市销率(P/S) | -.989 | .137 | -.989 | .137 |
| Zscore: EV/EBITDA | -.996 | .061 | -.996 | .061 |

图 4、成分矩阵

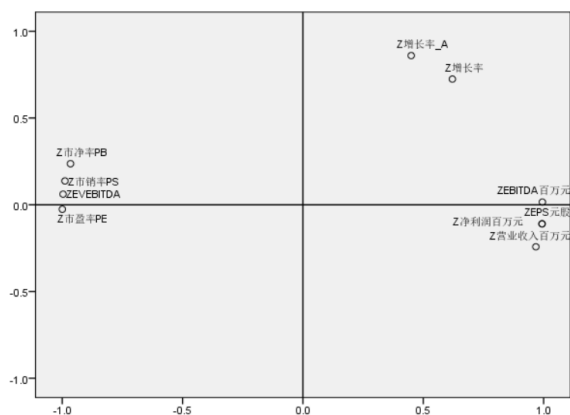


图 5、载荷图

根据上面的图 4 在重新标度成分列的第一主成分中，我们看到大于在区间绝对值大于等于 0.9 的指标有 8 个，故我们就说这 8 个指标在第一主成分上具有代表性，另外两个指标在第二主成分列中占比远远大于其他指标，所以另外两个指标在第二主成分上具有代表性，从右边的图 5、载荷图，我们能更清晰的看出来，越靠近 x 轴的与第一主成分高度相关，越靠近 y 轴的与第二主成分高度相关。

(3) 特征向量的求解

上述得到的成分矩阵是因子分析法所得，现在需要将因子分析法所得的成分矩阵转为主成分分析法的特征向量，在此利用主成分分析法与因子分析法的关系式： $u_{ij} = a_{ij} / \lambda^{1/2}$ 即特征向量等于成分矩阵除以根号下 λ ，通过公式我们得到如下图所示的数据：

| W1 | W2 |
|------|------|
| .33 | -.20 |
| .21 | .61 |
| .34 | .01 |
| .34 | -.09 |
| .16 | .72 |
| .34 | -.09 |
| -.35 | -.02 |
| -.33 | .20 |
| -.34 | .11 |
| -.34 | .05 |

图 6、特征向量 W1、W2

(4) 主成分的求解

知道了样本数据的特征向量，想要求得主成分，只需特征向量乘以十个指标数据所对应向量，即 $y_i = u_i^T (x_{i1}, x_{i2}, \dots, x_{i10})^T$ ，按照此公式求得塔牌集团数据集的两个主成分分别是：

$$y_1 = u_1^T (x_{i1}, x_{i2}, \dots, x_{i10})^T = (-3.25, -2.87, 1.00, 2.20, 2.92)$$

$$y_2 = u_2^T (x_{i1}, x_{i2}, \dots, x_{i10})^T = (-0.24, -0.41, 1.91, 0.09, -1.35)$$

(5) 按此方法求得另外 9 支股票的主成分分别是：

中国平安：

$$y_1 = u_1^T (x_{i1}, x_{i2}, \dots, x_{i10})^T = (-2.71, -2.14, -0.50, 1.99, 3.36)$$

$$y_2 = u_2^T (x_{i1}, x_{i2}, \dots, x_{i10})^T = (2.10, -0.10, -3.28, 0.35, 0.93)$$

$y_3=u_3T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-0.65, 1.24, -0.33, -1.14, 0.88)$

中炬高新:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-1.02, 1.46, -4.21, 1.39, 2.38)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(1.92, 1.17, -1.10, -0.33, -1.66)$

中顺洁柔:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(1.19, 3.85, -2.34, -1.45, -1.25)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(1.65, 5.26, -3.20, -1.99, -1.72)$

珠江啤酒:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-3.06, -2.04, -0.52, 1.87, 3.76)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-6.96, -4.62, -1.19, 4.22, 8.53)$

顺丰控股:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-0.41, 0.64, -3.73, 1.00, 2.50)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(2.44, 1.26, -1.14, -1.50, -1.05)$

$y_3=u_3T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-0.24, 0.62, -0.45, 1.53, -1.46)$

粤水电:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-3.69, -2.10, 0.14, 1.93, 3.71)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(1.92, -1.67, -1.13, 0.09, .99)$

$y_3=u_3T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(0.29, -1.59, 1.76, 0.44, -0.90)$

丽珠集团:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-3.14, -0.83, -0.48, 1.65, 2.81)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-0.43, 1.82, -1.45, 0.49, -0.43)$

国药一致:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(-2.68, -1.50, -1.22, 1.45, 3.94)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(1.45, 0.97, -2.68, -0.40, 0.66)$

格力电气:

$y_1=u_1T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(1.11, 3.33, 0.72, -1.62, -3.54)$

$y_2=u_2T(x_{i1},x_{i2},\cdots,x_{i10})$ $T=(2.47, -0.49, -1.58, -0.73, 0.34)$

4.2. 问题二模型的建立与求解

4.2.1. 数据处理

[2]ARIMA 模型的全称叫做自回归移动平均模型, 全称是(ARIMA, Autoregressive Integrated Moving Average Model)。也记作 ARIMA(p,d,q), 是统计模型(statistic model)中最常见的一种用来进行时间序列预测的模型。

(1) 检查数据集平稳性

平稳性: 平稳性就是要求经由样本时间序列所得到的拟合曲线, 在未来的一段期间内仍能顺着现有的形态“惯性”地延续下去。

这里只需用 python 画出股票的收盘价的趋势图, 便可以看出数据是否具有平稳性。

(2) 数据的准备——差分

观察我们所要测试的数据是否稳定, 稳定的数据是没有趋势(trend), 没有周期性(seasonality)的; 即它的均值, 在时间轴上拥有常量的振幅, 并且它的方差, 在时间轴上是趋于同一个稳定的值的。

ARIMA 模型要求所给数据的时序数据是稳定的(stationary), 或者是通过差分化(differencing)后是稳定的。

差分是一种处理数据的巧妙且简单的方法, 差分的作用是减轻数据之间的不规律

波动，使其波动曲线更平稳。

在差分法中，用到了两个数组：原数组 $a[]$ 、差分数组 $D[]$ 。其中差分数组的定义如下：

$$D[k]=a[k]-a[k-1] \quad (7)$$

即原数组 $a[]$ 相邻元素之间的差,从(7)中可以推出 $a[k]=D[1]+D[2]+\cdots+D[k]$

经过十个数据的测试，发现股票数据的收盘价经过一阶差分后数据已经具有平稳性，故我们无需进行二阶差分。

4.2.2. 建立 ARIMA 模型

(1) 自回归模型 AR

自回归模型描述当前值与历史值之间的关系，用变量自身的历史时间数据对自身进行预测。自回归模型必须满足平稳性的要求。

自回归模型首先需要确定一个阶数 p ，表示用几期的历史值来预测当前值。 p 阶

自回归模型的公式定义为：
$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t \quad (9)$$

上式中 y_t 是当前值, μ 是常数项, p 是阶数, γ_i 是自相关系数, ϵ_t 是误差。

(2) 移动平均模型 MA

移动平均模型关注的是自回归模型中的误差项的累加， q 阶自回归过程的公式定义如下：

$$y_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (10)$$

移动平均法能有效地消除预测中的随机波动。

(3) 自回归移动平均模型 ARMA

自回归模型 AR 和移动平均模型 MA 模型相结合，我们就得到了自回归移动平均模型 ARMA(p,q)，计算公式如下：

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (11)$$

(4) 差分自回归移动平均模型 ARIMA

将自回归模型、移动平均模型和差分法结合，我们就得到了差分自回归移动平均模型 ARIMA(p,d,q)，其中 d 是需要对数据进行差分的阶数。

(5) 通过 ACF 图 and PACF 图确立模型以及定阶

模型的识别问题和定阶问题，主要是确定 p ， d ， q 三个参数，差分的阶数 d 一般通过观察图示，1 阶或 2 阶即可。这里我们主要介绍 p 和 q 的确定。我们首先介绍两个函数。

自相关函数即 ACF(Autocorrelation Function)，指任意时间 t ($t=1,2,3\dots n$) 的序列值 X_t 与其自身的滞后（这里取滞后一阶，即 $\text{lag}=1$ ）值 X_{t-1} 之间的线性关系。



图 7、 X_t 与 X_{t-1} 的关系 1

$$ACF(k) = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \quad (8)$$

自相关函数公式：

其中 k 代表滞后期数，如果 $k=2$ ，则代表 y_t 和 y_{t-2}

了解了自相关函数后，需要将数据带入上述（8）公式进行计算，绘出自相关函数图，即 ACF 图，ACF 图就是指以滞后值（上图 7 就是滞后值为 1 阶）为 x 轴，自相关系数（ X_t 与 X_{t-1} 的相关系数值）为 y 轴画出的图

[3]ACF 是一个完整的自相关函数，可为我们提供具有滞后值的任何序列的自相关值。它描述了该序列的当前值与其过去的值之间的相关程度。而 PACF 是部分自相关函数或者偏自相关函数。直观上来说，PACF 只描述观测值 y_t 和其滞后项 y_{t-k} 之间的直接关系，调整了其他较短滞后项（ $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$ ）的影响。



图 8、PACF 解释

由图 8 可以更直观的知道 PACF 的部分性，即 PACF 图是部分相关图。

[4]建立完 ACF 图和 PACF 图后，需通过这两图分别确定 ARMA 模型的 p, q 值，方法如下：

| 模型（序列） | AR (p) | MA (q) | ARMA (p, q) |
|--------|------------|------------|-------------|
| 自相关函数 | 拖尾 | 第 q 个后截尾 | 拖尾 |
| 偏自相关函数 | 第 p 个后截尾 | 拖尾 | 拖尾 |

图 9、模型识别

简单介绍一下拖尾和截尾：拖尾指序列以指数率单调递减或震荡衰减，而截尾指序列从某个时点变得非常小：

[4]出现以下情况，通常视为(偏)自相关系数 d 阶截尾：

- 1) 在最初的 d 阶明显大于 2 倍标准差范围
- 2) 之后几乎 95% 的(偏)自相关系数都落在 2 倍标准差范围以内
- 3) 且由非零自相关系数衰减为在零附近小值波动的过程非常突然

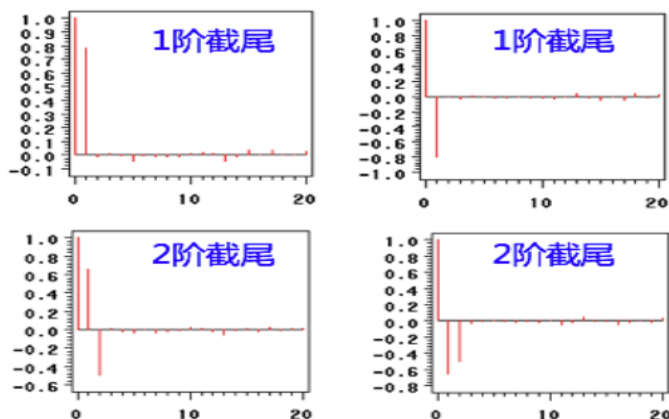


图 12、截尾

[4]出现以下情况，通常视为(偏)自相关系数拖尾：

- 1) 如果有超过 5%的样本(偏)自相关系数都落入 2 倍标准差范围之外
- 2) 或者是由显著非 0 的(偏)自相关系数衰减为小值波动的过程比较缓慢或非常连续。

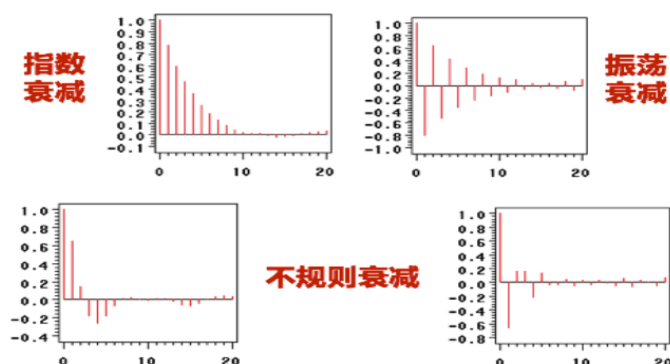


图 13、拖尾 1

通过画出 ACF 图和 PACF 图观察是 AR 模型还是 MA 模型或者是 ARMA 模型，如下所示

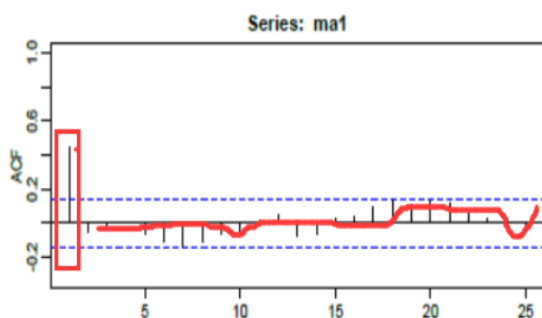


图 10、ACF 图

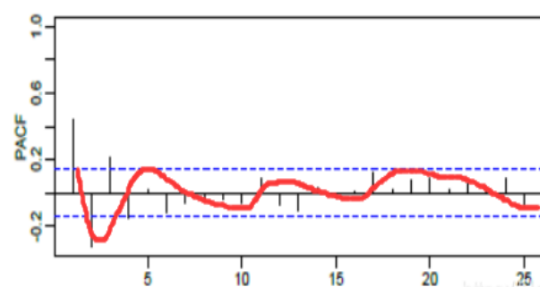


图 11、PACF 图

可以看到 ACF 自 1 阶过后都落在 2 倍标准差范围内，PACF 呈现拖尾（拖尾指序列以指数率单调递减或震荡衰减，而截尾指序列从某个时点变得非常小），所以通过图 9 确定了模型是 MA(1)，即 ARMA(0,1)模型

(6) 参数估计

通过拖尾和截尾确认了模型以及 p, q 的阶数后，也就是对模型定阶，这样出来的 p, q 往往具有很强的主观性，存在一定的误差，想要平衡预测误差和参数个数，我们可以根据信息准则函数法，在这里我们选用了 BIC (Bayesian Information Criterion) 贝叶斯信息准则准则，计算公式如下：

$$\text{BIC} = \ln(n) * (\text{模型中参数的个数}) - 2\ln(\text{模型的极大似然函数值}), n \text{ 是样本容量}$$

(9)

4.2.3. 模型的求解

(1) 导入数据

这里以塔牌集团（股票代码：002233）为例利用 python 的 tushare 包从 tushare 官网下载塔牌集团的数据集，选取收盘价作为评判标准，用 resample 函数按周统计收盘价，于是得到一个有时序性的数据，我们将其画出来，趋势图如下：



图 14、股票趋势图

(2) 对数据进行差分，得到该数据集的一阶差分图

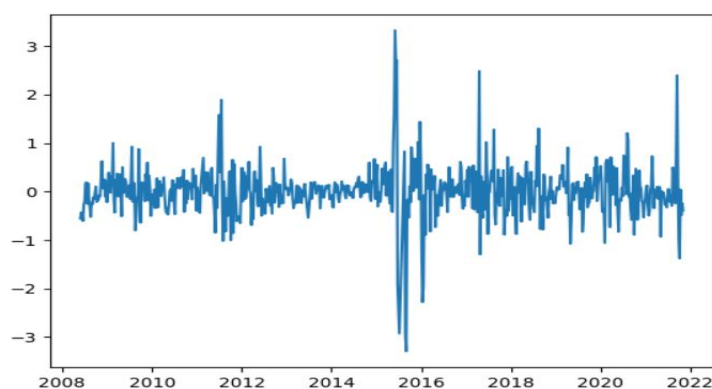


图 15、一阶差分

通过差分图可以知道，该数据集经过一阶差分已经基本平稳。

(3) 画出 ACF 图和 PACF 图，确定模型以及 p , q 值，如下图所示：

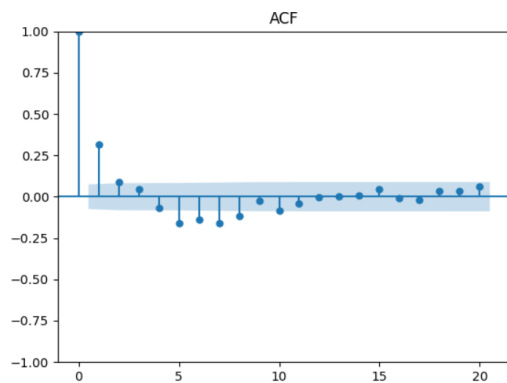


图 16、ACF 图

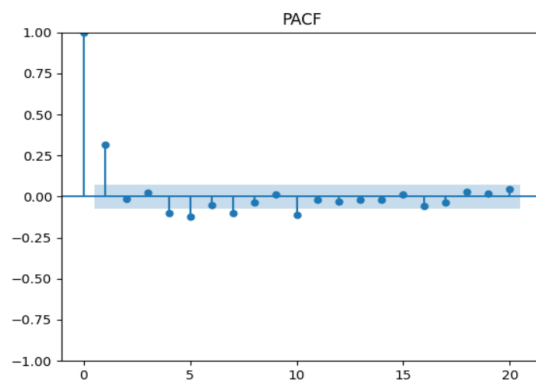


图 17、PACF 图

从图 16 可以看出 ACF 图是拖尾，PACF 图是 1 阶截尾，所以我们选用 AR 模型。

(4) 用 python 的 `ARIMA()` 函数训练模型，再用 `predict()` 函数预测股票未来五天的走向，我们得到一个如下的图片：

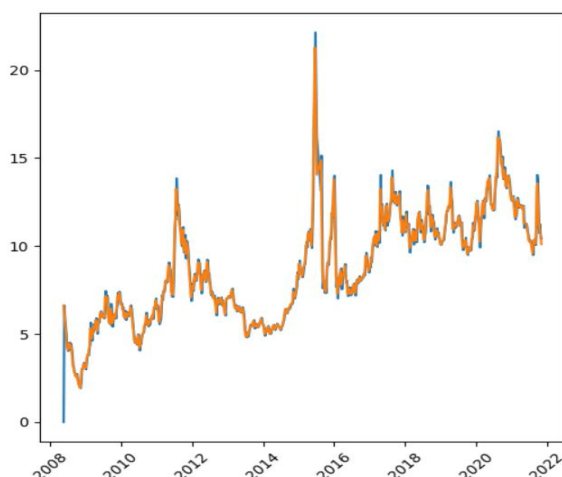


图 18、预测趋势图

4.2.4. 问题二结果——投资策略

以塔牌集团为例：

由图 17 可以看到股票近期会持续下跌，从 08 年-21 年的整体趋势，我们发现该股整体上是涨了，但跌宕起伏，峰值与低谷差距颇大，建议在股票下跌期买入股票，一有上涨趋势应时刻观望，一般上涨 1-2 天后可选择抛出，采用趋势投资法，并且是短期的，买入抛出历时不能太长。即短线玩家入市，跌破 5 日均线出局。持股的过程中可能会出现局部的高点，会有做差价高抛低吸的机会可以采用股价对趋势线的乖离度叠加成交量、历史振幅等判断是否应该高抛。

4.3. 问题三模型的建立与求解

4.3.1. 数据筛选——股票预测分析的假设条件

股票价格的技术分析具有三个假设条件：第一个是假设市场行为涵盖了股票相关的一切数据信息，它是我们进行技术分析的基础，影响股票价格的所有因素都会在市场行为中或多或少不同程度地反映出来，比如问题三给出的突发事件的闪现、舆情、自然灾害等事件对市场的影响都会在价格的波动中直接体现出来；第二是股票价格通常可以用图表表示并沿着某种趋势发展，即认为股票市场作为一种客观经济现象，而它必然遵循一定的规律，股票价格是交易双方供求关系的平衡点，因此股票价格的变动总是趋向于使供需双方达到平衡的方向发展，达到平衡后，新的因素产生打破平衡，导致股票价格发生变化，从而再次达到新的平衡，因此股票价格的变化过程是某种动态平衡的发展；三是经济事件会重复，人的天性是固定的，股民一般不会在一夜之间改变他们的动机，因此人们的反应具有相似性，研究过去历史市场所呈现的现象可以帮助我们判断与预测未来。

因此，本文在同花顺金融软件上查询并提取某一时间区段内股票开盘价、收盘价、涨幅等关键数据，并打包为规范数据集。这里以广州白云山医药集团（600332）为代表。

| 时间 | 开盘 | 最高 | 最低 | 收盘 | 涨幅 | 振幅 | 总手 | 金额 | 换手% | 成交次数 |
|---------------|-------|-------|-------|-------|--------|-------|------------|-------------|-------|-------|
| 2021-06-01, 二 | 32.43 | 32.5 | 31.82 | 31.96 | -0.99% | 2.11% | 11,631,349 | 373,226,220 | 0.827 | 23287 |
| 2021-06-02, 三 | 31.83 | 31.99 | 31.3 | 31.48 | -1.50% | 2.16% | 9,037,782 | 285,659,520 | 0.643 | 17558 |
| 2021-06-03, 四 | 31.56 | 32.05 | 31.41 | 31.69 | 0.67% | 2.03% | 8,218,640 | 261,021,650 | 0.585 | 18371 |
| 2021-06-04, 五 | 31.74 | 32.22 | 31.54 | 31.85 | 0.50% | 2.15% | 9,736,540 | 310,245,690 | 0.693 | 16999 |
| 2021-06-07, 一 | 31.7 | 31.73 | 31.14 | 31.2 | -2.04% | 1.85% | 9,614,266 | 301,772,230 | 0.684 | 15764 |
| 2021-06-08, 二 | 31.2 | 31.66 | 31 | 31.15 | -0.16% | 2.12% | 9,031,447 | 282,874,280 | 0.642 | 16153 |
| 2021-06-09, 三 | 31.19 | 31.25 | 30.59 | 30.87 | -0.90% | 2.12% | 7,310,150 | 225,453,360 | 0.52 | 14536 |
| 2021-06-10, 四 | 30.95 | 31.03 | 30.28 | 30.36 | -1.65% | 2.43% | 9,930,729 | 303,173,600 | 0.706 | 16235 |
| 2021-06-11, 五 | 30.29 | 30.46 | 29.77 | 29.9 | -1.52% | 2.27% | 10,481,508 | 314,594,610 | 0.746 | 20329 |
| 2021-06-15, 二 | 30.13 | 30.17 | 29.67 | 29.94 | 0.13% | 1.67% | 7,538,339 | 225,553,120 | 0.536 | 17604 |
| 2021-06-16, 三 | 29.9 | 30.27 | 29.6 | 29.6 | -1.14% | 2.24% | 7,251,516 | 216,526,650 | 0.516 | 17907 |
| 2021-06-17, 四 | 29.64 | 29.92 | 29.5 | 29.76 | 0.54% | 1.42% | 6,309,636 | 187,241,210 | 0.449 | 13882 |
| 2021-06-18, 五 | 29.8 | 29.8 | 29.38 | 29.5 | -0.87% | 1.41% | 6,676,314 | 197,138,930 | 0.475 | 14546 |
| 2021-06-21, 一 | 29 | 29.62 | 28.88 | 29.49 | -0.03% | 2.51% | 7,094,645 | 208,136,240 | 0.505 | 15152 |
| 2021-06-22, 二 | 29.5 | 30.07 | 29.25 | 29.91 | 1.42% | 2.78% | 9,323,994 | 278,008,810 | 0.663 | 18158 |
| 2021-06-23, 三 | 30 | 30.09 | 29.7 | 29.93 | 0.07% | 1.30% | 6,384,467 | 190,885,430 | 0.454 | 12433 |
| 2021-06-24, 四 | 29.95 | 30.08 | 29.61 | 29.96 | 0.10% | 1.57% | 5,400,201 | 161,074,130 | 0.384 | 11244 |
| 2021-06-25, 五 | 29.9 | 30 | 29.64 | 29.97 | 0.03% | 1.20% | 5,985,422 | 178,933,760 | 0.426 | 12494 |

图 19、白云山集团股票数据

针对问题仅选取收盘价作为指标，进行相关预测分析。相对于总利润、最高价、最低价而言，以调整后的收盘价作为分析指标可以较准确地反映出该股票的价值。本题使用 ARIMA 模型对未来股票价格进行预测并与先前的股价走势进行比较，并对突发事件对股票价格的影响总结和提出相关策略。

4.3.2. ARIMA 模型的建立

ARIMA 模型使用历史已有的数据作为估计基础，同时也可以股票价格遭受突发事件的冲击的时候进行快速调整，从而来预测平稳时间序列的一个时间段内的数据。Ruppert 和 Matteson 指出，任何非随机白噪声的“非季节性”时间序列都可以使用 ARIMA 模型进行建模。ARIMA 模型是自回归模型（AR）、移动平均模型（MA）和差分的组合，可以表示为 ARIMA(p,d,q)，其中 p 和 q 分别是 AR 和 MA 的滞后阶数，d 是需要差分时间序列以确保平稳性的阶数。AR 可以帮助人们观察过去值和现在值之间的相似性，MA 可以通过考虑过去残差来显示过去误差和现在值之间的相似性。本文所采用的 ARIMA(p,d,q) 的一般公式可以表示为：

$$\Delta P_t = \beta_0 + \beta_1 \Delta P_{t-1} + \beta_2 \Delta P_{t-2} + \dots + \beta_q \Delta P_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

4.3.3. 模型的求解

(1) 数据标准化

对选取的十支股票疫情前后时间段的股票收盘价进行整合，打包为规范数据集如下：

| A | A | A | A | A | A |
|----------|----------|----------|----------|----------|----------|
| VarNa... | VarNa... | VarNa... | VarNa... | VarNa... | VarNa... |
| 数值 | 数值 | 数值 | 数值 | 数值 | 数值 |
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 |
| 10 | 10 | 10 | 10 | 10 | 10 |
| 11 | 11 | 11 | 11 | 11 | 11 |
| 12 | 12 | 12 | 12 | 12 | 12 |
| 13 | 13 | 13 | 13 | 13 | 13 |
| 14 | 14 | 14 | 14 | 14 | 14 |
| 15 | 15 | 15 | 15 | 15 | 15 |
| 16 | 16 | 16 | 16 | 16 | 16 |
| 17 | 17 | 17 | 17 | 17 | 17 |
| 18 | 18 | 18 | 18 | 18 | 18 |
| 19 | 19 | 19 | 19 | 19 | 19 |
| 20 | 20 | 20 | 20 | 20 | 20 |
| 21 | 21 | 21 | 21 | 21 | 21 |
| 22 | 22 | 22 | 22 | 22 | 22 |
| 23 | 23 | 23 | 23 | 23 | 23 |
| 24 | 24 | 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | 28 | 28 |
| 29 | 29 | 29 | 29 | 29 | 29 |
| Sheet1 | Sheet1 | Sheet1 | Sheet1 | Sheet1 | Sheet1 |

| A | A | A | A | A | A | A |
|----------|----------|----------|----------|----------|----------|----------|
| VarNa... | VarNa... | VarNa... | VarNa... | VarNa... | VarNa... | VarNa... |
| 数值 | 数值 | 数值 | 数值 | 数值 | 数值 | 数值 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| 17 | 17 | 17 | 17 | 17 | 17 | 17 |
| 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| 19 | 19 | 19 | 19 | 19 | 19 | 19 |
| 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 21 | 21 | 21 | 21 | 21 | 21 | 21 |
| 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| 23 | 23 | 23 | 23 | 23 | 23 | 23 |
| 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| Sheet1 | Sheet1 | Sheet1 | Sheet1 | Sheet1 | Sheet1 | Sheet1 |

图 20、21、十支股票疫情前后时间段的股票收盘价

(2) 数据处理

在 Matlab 软件中编码,对数据进行时间序列平稳化,去白噪音,一阶差分的操作,旨在得到完全随机性的数据进行预测分析。

(3) 问题求解

数据选用 2020 年 6 月份至 9 月份工作日内十支股票的收盘价数据作为样本,并选用 2020 年 1 月份至 4 月份疫情初爆发期间数据为对比,运用 ARIMA 模型进行分析预测,比较分析股价短期内变化的规律。

白云山

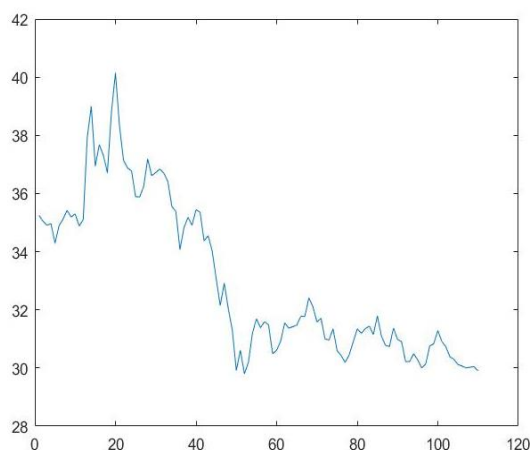


图 1

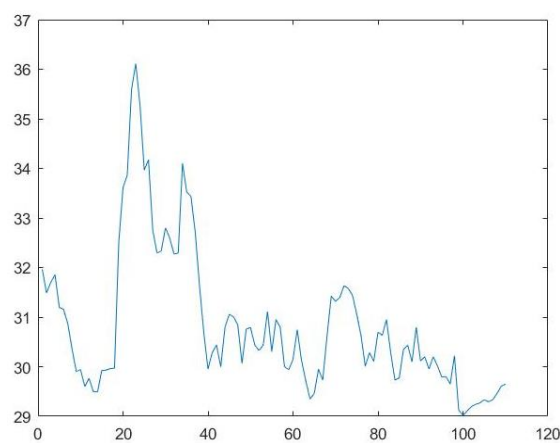


图 2

图 1 为根据疫情初期白云山医药集团股票收盘价分析所得股价走势,如图可见总体呈下跌式,但初期仍有较高峰值。

图 2 为疫情彻底爆发后白云山医药集团股票收盘价分析所得股票走势,大体围绕在较低价格,并跌破疫情爆发前最低点,总体较为颓势。

美盈森

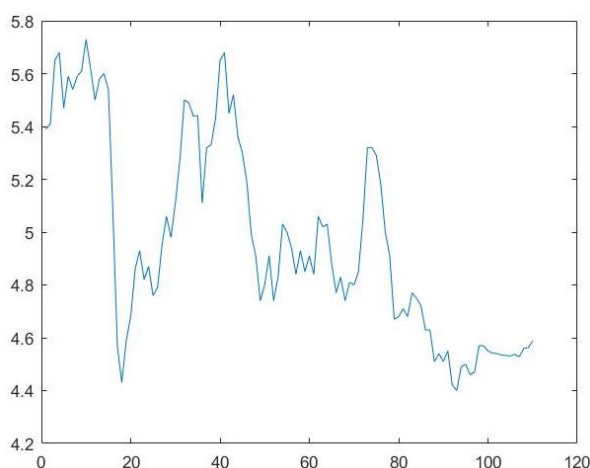


图 3

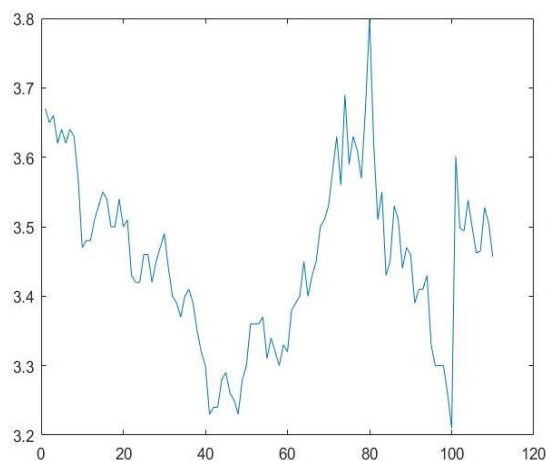


图 4

图 3 为据疫情初期美盈森集团股票收盘价分析所得股价走势,上下浮动较大,但整体处于较高水平。

图 4 为疫情彻底爆发后美盈森集团股票收盘价分析所得股票走势,总体股价已大幅跌落,且在较低水平反复,股价不太乐观,质量较低。

粤水电

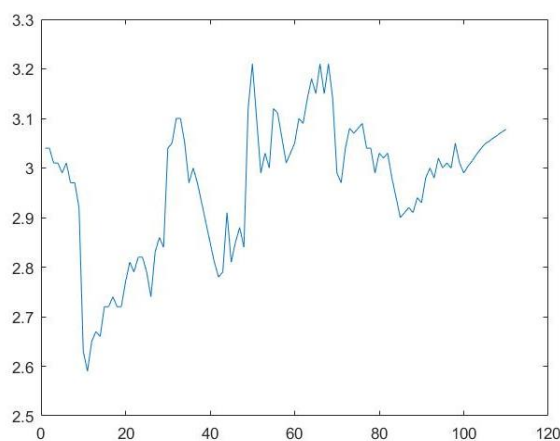


图 5

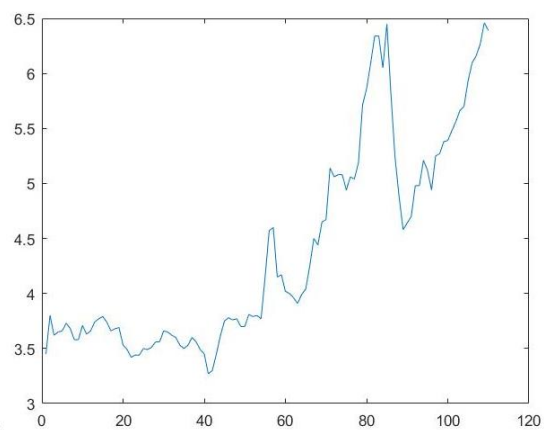


图 6

图 5 为疫情初期粤水电集团股票收盘价分析所得股价走势，整体水平较为稳定。图 6 为疫情彻底爆发后粤水电集团股票收盘价分析所得股价走势，整体为上升态势，股票质量较高，股价持续上升。

中炬高新

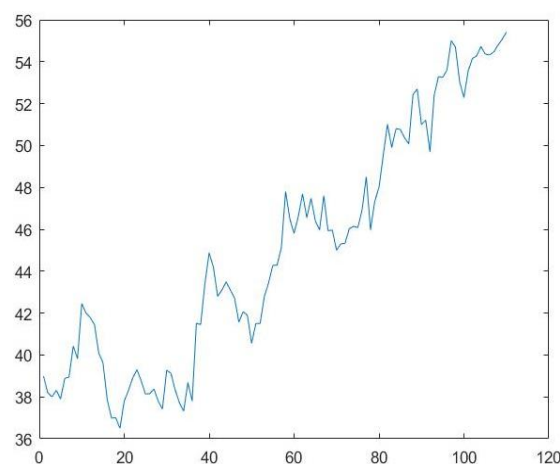


图 7

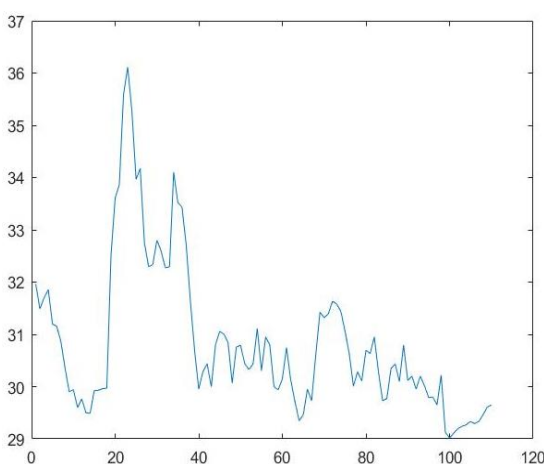


图 8

图 7 为疫情初期中炬高新集团股票收盘价分析所得股价走势，整体为上升态势，股票价格持续升高，股票质量较好。

图 8 为疫情彻底爆发后中炬高新集团股票收盘价分析所得股价走势，相比图 7 整体价格跌落维持在较低水平，股价持续低迷，股票质量较差。

中顺洁柔

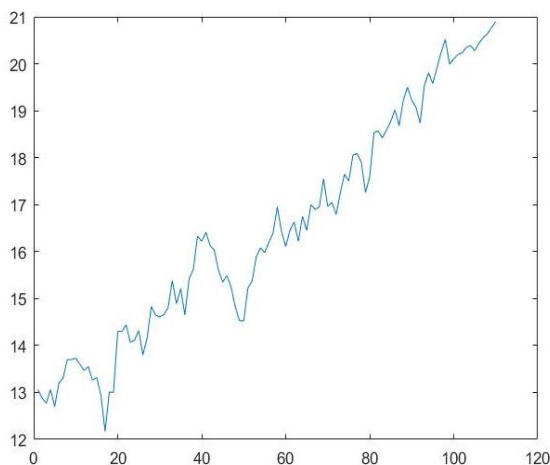


图 9

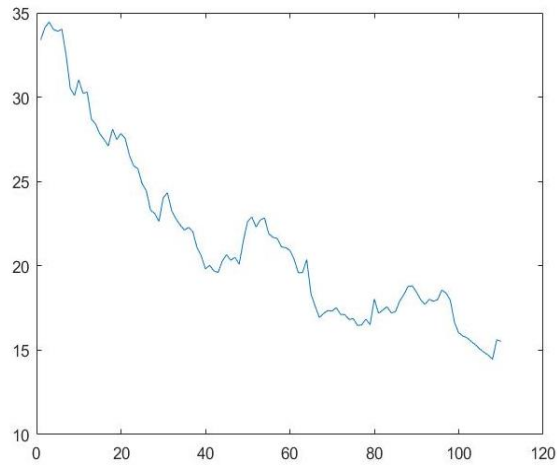


图 10

图 9 为疫情初期中顺洁柔集团股票收盘价分析所得股价走势，股票价格持续升高，股票质量较好。

图 10 为疫情彻底爆发后中顺洁柔集团股票收盘价分析所得股价走势，与图 9 相比，股票价格大跌，并持续走低，股票质量较差。

顺丰控股

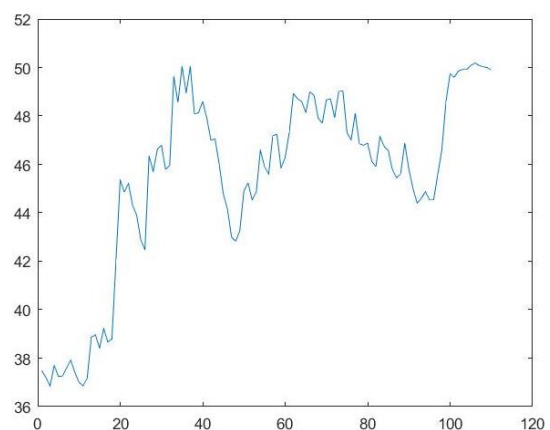


图 11

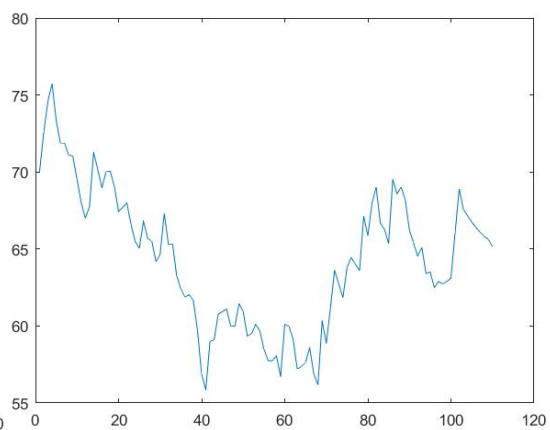


图 12

图 11 为疫情初期顺丰控股股票收盘价分析所得股价走势，股票价格先增加后比较稳定，在一定的价格区间内增减，且价格都比较高，股票质量较好。

图 12 为疫情彻底爆发后顺丰控股股票收盘价分析所得股价走势，与图 11 相比，股票价格偏高，股票价格跌至最低时价格仍较高，且后又逐渐回升，股票质量相对比较稳定。

国药一致

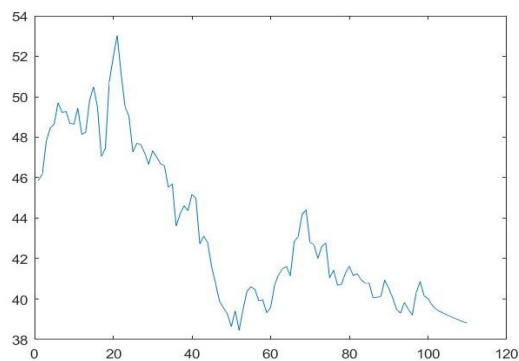


图 13

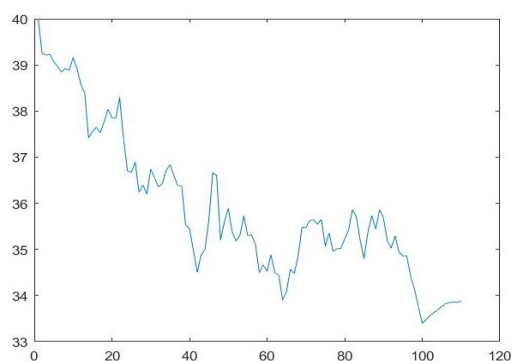


图 14

图 13 为疫情初期国药一致股票收盘价分析所得股价走势，股票价格先增后一路下降，中途有较小回升，股票质量较差。

图 14 为疫情彻底爆发后顺丰控股股票收盘价分析所得股价走势，与图 13 相比，股价较低且不断下跌，股票质量仍然较差。

塔牌集团

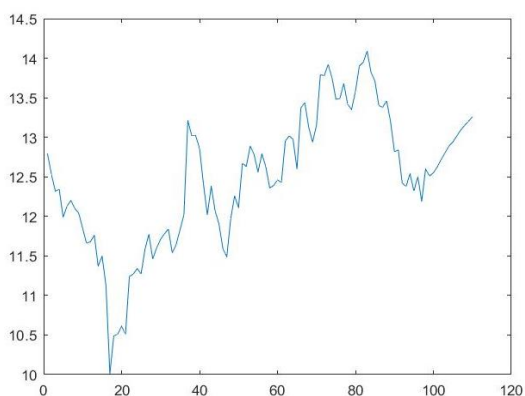


图 15

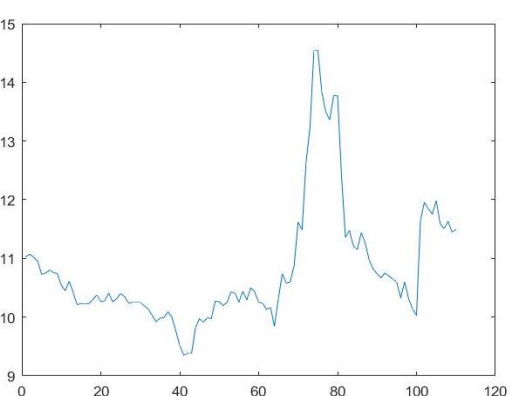


图 16

图 15 为疫情初期塔牌集团股票收盘价分析所得股价走势，股价整体偏低，先大幅度下跌后回升并稳定在一定价格区间，总体较稳定。

图 16 为疫情彻底爆发后顺丰控股股票收盘价分析所得股价走势，与图 15 相比，股价较为稳定且后有大幅度提升但很快回落，整体较为稳定。

珠江啤酒

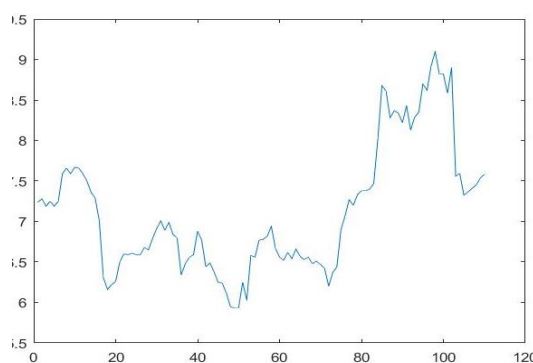


图 17

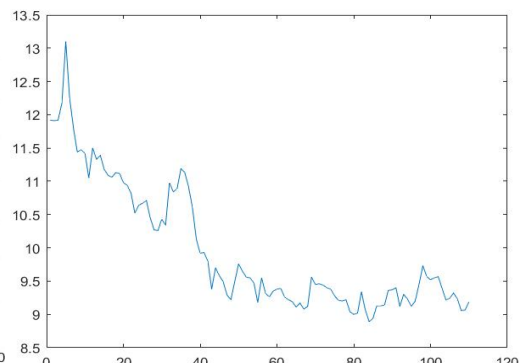


图 18

图 17 为疫情初期珠江啤酒股票收盘价分析所得股价走势，股价较低，但整体在一定的价格区间内起伏，股票价格较为稳定。

图 18 为疫情彻底爆发后顺丰控股股票收盘价分析所得股价走势，与图 17 相比，

股票价格从较高价持续下跌后稳定在较低股价水平，股票质量较差。

丽珠集团

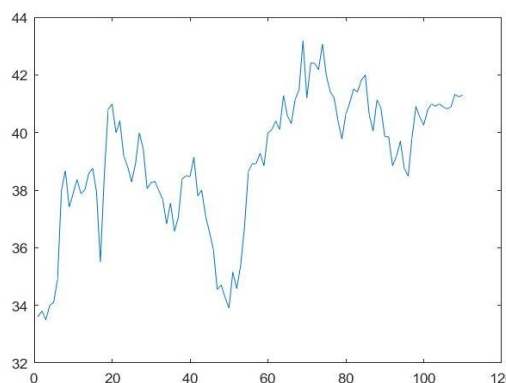


图 19

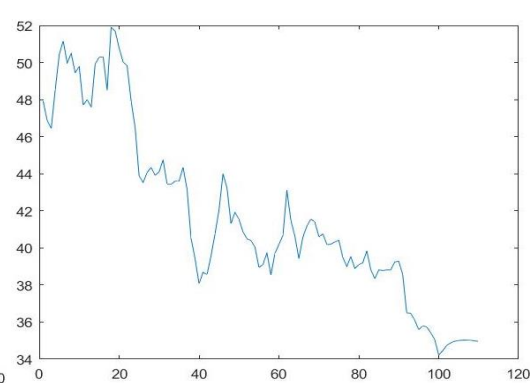


图 20

图 19 为疫情初期丽珠集团股票收盘价分析所得股价走势，股价在一定价格区间内起伏，整体较为稳定。

图 20 为疫情彻底爆发后顺丰控股股票收盘价分析所得股价走势，与图 20 相比，股价从高价持续下跌，股票质量较差。

4.3.4. 问题总结

综上所述，经过建立 ARIMA 模型，选取疫情作为突发事件对 10 支股票的影响，通过对比疫情爆发初期及其彻底爆发后的股价走势，可以看出疫情对于股票价格的影响有利也有弊，不同的产业股票所受到的影响也会不同。例如：在疫情爆发初期粤水电和顺丰控股股价整体较为稳定，但是在疫情彻底爆发之后，粤水电的股价走势继续上升，而顺丰控股股价下跌后稳定。但是其他行业可见，疫情爆发后股价大幅度下跌或是稳定但也有下跌的趋势。可见：

一、新冠疫情对股票市场的影响以负面影响为主，整体趋势均为下跌和稳定，通过数据可见，短期内的影响更加显著。

二、从行业进行分析，在人们日常所需如水电、纸业等在疫情初期需求量增加，而物流、娱乐暂停导致初期时股价不断下跌，其他行业股价也有不同程度下跌或升高然后在一定价格区间里稳定波动。而在疫情彻底爆发后，对各行业需求会发生一定的变化，而后股价也呈稳定趋势。

三、新冠疫情也会影响投资者的判断和情绪等，投资者会根据疫情带来的影响而进行选择，可能不会对之前所选的股票进行投资下注而选择其他的股票。而新冠疫情也给当时整个社会带来恐慌和低迷，投资者也可能会有错误性预判或者冲动投资等行为，负面情绪会导致股票市场整体下跌。而在彻底爆发后，各行业有了一定的措施进行应对，股价会逐渐趋于稳定。

4.4 问题四模型的建立与求解

引用问题二三所建模型，ARIMA 模型可以有效分析预测未来股票价格趋势，其中券商研报会影响股票市场的波动，在一定程度上左右投资者的投资，形成利好或利空。如果研报给出的是利好，则会吸引投资者对该股大量的买入，从而会推动股价的上涨；如果研报是利空，则会引起投资者恐慌和躁动，大量的抛出手中股票，导致股价下跌。

个股的走势也受市场上的众多因素影响，在研报利好的情况下，股价也可能会下跌，在研报利空的情况下，股价也可能会上涨。证券研报是证券研究部门对上市公司或行业观、行情预测等研究性报告。因为同时面向众多机构投资者发布,如有大资金的

机构相信了其研究结论,就会根据研报信息进行买卖操作,从而带动股票价格波动。

外界环境因素在不同程度上都会影响股票价格的走势,如各类突发事件、自然灾害导致的市场动荡、行业波动都会直接影响股票价格。在买入股票时外界环境因素是一个重要的考量指标。

由此修改问题二所提出策略:当股票出现下跌期时,不应盲目抄底买入,而应观摩几个周期,并充分考证相当一段时期内外界证券机构所给出的研报分析,再收集资料判断近期相关市场是否会遭受外界因素冲击、行业是否出现颓势,预测股票是在小周期内反复波动亦或是收到某种冲击导致的股票价格持续性走低,这样就能大幅度避免盲目抄底带来的严重损失。

五、模型的评价、改进

5.1 模型的优点

1.具备数据筛选、数据降维等数据预处理能力,整体运行逻辑较为合理,有明确的分析逻辑和数学公式,预测结果较为科学。

2.模型十分简单,只需要内生变量而不需要借助其他外生变量。

5.2 模型的缺点

分析时考虑的因素不够全面,采纳的主要指标较为主观。采用的 ARIMA 模型过于简单了,对数据的限制太多,如:

1.要求时序数据是稳定的 (stationary), 或者是通过差分化(differencing)后是稳定的。

2.本质上只能捕捉线性关系,而不能捕捉非线性关系。

3.代码实践时没有运用 BIC 准则降低 p 、 q 的误差

5.3 模型的改进

在原有基础上增加考虑的指标,增强模型可行性。或者可以考虑换一个模型,采用多因子量化选股模型,将第一题提取出来的主成分因子代入此模型,求出股票的趋势图,由于此模型用到了更多特征指标,且都是选取该股最需要的指标,所以预测出来的数据更加准确且稳定。

六、参考文献

- [1] 阿泽,【机器学习】降维——PCA, <https://zhuanlan.zhihu.com/p/77151308>, 2021-11-5
- [2] HHXUN , ARIMA 模型介绍 , <https://blog.csdn.net/HHXUN/article/details/79858672>, 2021-11-6
- [3] 小白掌柜, 如何根据自相关 (ACF) 图和偏自相关 (PACF) 图选择 ARIMA 模型的 p 、 q 值 , https://blog.csdn.net/weixin_41013322/article/details/108801516, 2021-11-6
- [4] lyp_1020k , 自相关系数与偏自相关系数, 拖尾与截尾, https://blog.csdn.net/lyp_1020k/article/details/109420638, 2021-11-7

附录

附录 1

介绍：python，预测股票趋势（ARIMA 模型）

```
import tushare as ts
import pandas as pd
import statsmodels.api as sm
from statsmodels.tsa.arima.model import ARIMA
import pmdarima as pm
from matplotlib import pyplot as plt
import seaborn as sns
from matplotlib import font_manager
pro = ts.pro_api('3ec68402dc7159e31f2437d77f287be3814a432d3b0839b0bd5ccae5')

data = ts.get_k_data('002233',start='2005-08-24', end='2021-11-02',autype='qfq')
data.to_csv('002233_text.csv')
df=pd.read_csv('002233_text.csv',index_col='date',parse_dates=True)
#index_col 表示把第一列设为 index,parse_dates=[0]表示把第一列按 datetime 格式解析

stock_week=df['close'].resample('W-TUE').mean()
#将收盘价作为评判标准，resample 指按周统计平均数据(可以指定哪天为基准日，此处是周二为基准)
stock_train=stock_week['2008':'2021'].dropna()
#选取 2010-2021 的数据
stock_train.to_csv('0022.csv')
stock_train=pd.read_csv('0022.csv',index_col='date',parse_dates=True)
stock_train.plot(figsize=(12,5))
plt.title('Close Price')
sns.despine() #边框控制，默认没有上方和右方的框
plt.show()
stock_diff=stock_train.diff().dropna() #对数据进行差分，目的使数据平缓,满足平稳性的要求
plt.figure()
plt.plot(stock_diff)
font_loc="
font_set = font_manager.FontProperties(fname='C:\Windows\Fonts\HGB2_CNKI.ttf,size=20)
plt.title('一阶差分',fontproperties=font_set) #注意要设置字体，否则会乱码
plt.show()
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
acf=plot_acf(stock_diff,lags=20)
plt.title('ACF')
acf.show()
plt.show()
```



```

pacf=plot_pacf(stock_diff,lags=20)
plt.title('PACF')
pacf.show()
plt.show()
model=pm.auto_arima(stock_train,start_p=1,start_q=1,max_p=3,max_q=3,m=12,start_P=0,seasonal=True,d=1,D=1,trace = True,error_action='ignore',suppress_warnings = True,stepwise=True,freq='W-TUE')
model=ARIMA(stock_train,order=(1,1,0))#训练模型,order 表示 (p,d,q)
result=model.fit()
pred=result.predict(n_periods=(len(stock_train)))
plt.figure(figsize=(6,6))
plt.xticks(rotation=45)
plt.plot(pred)
plt.plot(stock_train)
plt.show()

```

附录 2

介绍：Python 语言编写，MATLAB 软件运行，ARIMA 模型，预测股票价格

%是通过一批现有数据（代码中 batch=100），向后预测后 n 个数据

```
[data] = xlsread('xx 股票收盘价');
```

% 设置数据块大小

```
batch = 100;
```

%预测数量

```
predict_num = 10;
```

% Box_Cox 转换

```
[transdata,lambda]=boxcox(data(:,1));
```

% ARIMA 模型处理 利用前 100 真实数据 滑动窗口预测

```
Predict=transdata(1:batch,1);
```

```
for iii=1:predict_num
```

```
Data=Predict(iii:iii+batch-1,1);
```

%滑动窗口大小为 100

```
SourceData=Data(1:batch,1);
```

%前 100 个训练集

```
step=1;
```

%向后预测一个

（效果比较好）

```
TempData=SourceData;
```

```
TempData=detrend(TempData);
```

%去趋势线

```
TrendData=SourceData-TempData;
```

%趋势函数

% -----差分，平稳化时间序列-----

```
H=adftest(TempData);
```

```
difftime=0;
```

```
SaveDiffData=[];
```

```
while ~H
```

```
SaveDiffData=[SaveDiffData,TempData(1,1)];
```

```
TempData=diff(TempData);
```

%差分，平稳化时间序列

```
difftime=difftime+1;
```

%差分次数

```

H=adftest(TempData);          %adf 检验，判断时间序列是否平稳化
end
% -----模型定阶或识别-----
u = iddata(TempData);
test = [];
for p = 1:5                    %自回归对应 PACF,给定滞后长度上
    限 p 和 q，一般取为 T/10、ln(T)或  $T^{(1/2)}$ ,这里取 T/10=12
    for q = 1:5                %移动平均对应 ACF
        m = armax(u,[p q]);
        AIC = aic(m);          %armax(p,q),计算 AIC
        test = [test;p q AIC];
    end
end
for k = 1:size(test,1)
    if test(k,3) == min(test(:,3))    %选择 AIC 值最小的模型
        p_test = test(k,1);
        q_test = test(k,2);
        break;
    end
end
% -----1 阶预测-----
TempData=[TempData;zeros(step,1)];
n=iddata(TempData);
m = armax(u,[p_test q_test]);        %armax(p,q),[p_test q_test]对应 AIC 值最
小，自动回归滑动平均模型
P1=predict(m,n,1);
PreR=P1.OutputData;
PreR=PreR';
% -----还原差分-----
if size(SaveDiffData,2)~=0
    for index=size(SaveDiffData,2):-1:1
        PreR=cumsum([SaveDiffData(index),PreR]);
    end
end
% -----预测趋势并返回结果-----
mp1=polyfit([1:size(TrendData',2)],TrendData',1);
xt=[];
for j=1:step
    xt=[xt,size(TrendData',2)+j];
end
TrendResult=polyval(mp1,xt);
PreData=TrendResult+PreR(size(SourceData',2)+1:size(PreR,2));
tempx=[TrendData',TrendResult]+PreR;    % tempx 为预测结果
Predict(100+iii,1)=tempx(1,101);

```

```
end  
% 求逆变换 % 最后的预测值在 Inverse_transdata 中  
Inverse_transdata=Inverse_BoxCox(Predict,lambda);
```