

时间序列分析课程期末报告  
-----基于ARIMA模型的股票价格分析及预测

作者：邵云卿

学号：1807403021

## 摘要

股票一直是人们十分热衷的获取经济利益的方式，股票价格的波动反映了一个国家在政治、经济等领域的发展状况和动向。因此如何准确分析股票市场的行情以及股价的波动趋势成为了重要的课题，准确的分析股市的行情，对于国家在经济上采取怎样的宏观调控以及投资者如何做出最优的决策，都有着十分重要的意义。

股票的价格序列是一个十分复杂的非线性动态系统，而传统式时间序列预测方法都是通过时间序列的统计关系来体现线性动态系统的特征以及变化，因此很难准确预测股票价格预测。本文对时间序列的相关理论进行了简单的介绍，并利用ARIMA模型对恒瑞医药（600276）股票从2015年1月5日到2021年6月28日的每日收盘价格进行建模和预测。

关键词：时间序列；股票价格预测；ARIMA模型

## 第一章 研究背景和意义

### 1.1 研究背景

股票市场是金融市场的重要组成部分，与国民经济发展密切相关，股票作为资本证券化的工具，一种虚拟资本的代表，可以方便地在股票市场融资。从表面上看，股票市场是通常意义下的规律性的。随着数学理论研究的深入和各种数据分析工具开发的迅速发展，人们用不同的方法和工具来分析金融时间序列，做出各种金融时间序列预测的模型，尤其是股票价格的预测模型。

### 1.2 研究意义

九十年代初，我国先后建立了上海和深圳两大证券交易所，是目前社会参与程度最高的金融市场。通过建立合理的模型对股票价格趋势进行短期预测，同时不断利用新的数据更新模型，做出新的预测，我们就可以实现连续预测，这对投资者如何较好地把握买卖时机将起到重要的指导作用。目前证券投资已经成为社会生活的一个重要部分，证券市场的运行状况对我国实体经济的影响逐渐加深，正在真正地发挥着国民经济“晴雨表”和“报警器”的作用。深刻理解股票市场的运行规律，预测分析股票价格的未来走势，无论是对广大投资者的降低投资风险还是宏观经济管理部门的宏观调控，保障我国证券市场的健康持续发展，都有着十分重要的意义。

## 第二章 理论介绍

### 2.1 时间序列的概念及发展

时间序列分析即通过对时间序列数据的统计和分析,对其发展和变化的规律进行研究,并对时间序列进行进一步的预测。我们常用按时间顺序排列的一组随机变量  $\dots X_1, X_2, \dots, X_t \dots$  来表示一个随机事件的时间序列,简记为  $\{X_t, t \in T\}$  或用  $\{X_t\}$ 。用  $x_1, x_2, \dots, x_n$  或  $\{x_n, n = 1, 2, \dots, n\}$  式表示该随机序列的  $n$  个有序观察值,称之为序列长度为  $n$  的观察值序列。我们研究的目的就是揭示随机时序  $\{X_t\}$  的性质,而要实现这个目标就是通过分析它的观察值序列  $\{x_t\}$  的性质,由观察值序列的性质来推断随机时序  $\{X_t\}$  的性质。

时间序列分析始于二十世纪二十年代,是随着树立学统计的发展而产生的,其研究重心将数据比较转移到了时间序列内在的关系上。随机变量具有很强的随机性,因此通过描述性时间序列分析来研究这些数据已经很难探寻出其中的规律了。而统计时间序列分析能通过对数据内部相互关系的分析,找出随机变量的发展趋势。统计时间序列分析的这一特点使其广泛的应用于金融、保险、心理学等社会科学研究领域中。

### 2.2 平稳时间序列

平稳性时时间序列分析的基础,平稳性的基本思想为:决定时间序列过程特性的统计规律不会随着时间的变化而变化。平稳性根据限制条件的严格程度可分为严平稳和宽平稳。

严平稳是一种条件比较苛刻的平稳性定义,它认为只有当序列所有的统计性质都不会随着时间的推移而发生变化时,该序列才能被认为平稳。因为随机变量族的统计性质完全由它们的联合概率分布族决定。所以严平稳的定义如下设  $\{X_t\}$  为一时间序列,对任意正整数  $m$ ,任取  $t_1, t_2, \dots, t_m \in T$ ,对任意整数  $\tau$ ,有

$F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m) = F_{t_1 + \tau, t_2 + \tau, \dots, t_m + \tau}(x_1, x_2, \dots, x_m)$ , 则称时间序列  $\{X_t\}$  严平稳。

宽平稳是使用时间序列的特征统计量来定义的一种平稳性。判断一个随机过程  $\{X_t\}$  是否是宽平稳的时间序列(或者二阶矩平稳)主要是判断其是否满足以下两个条件:

(1)均值函数在所有时间上恒为常数;

(2) $\gamma_X(s, t) = \gamma_X(0, t - s)$  对所有的时间  $t, s$  都成立。

相对于严平稳严苛的判断条件,宽平稳更加容易判断,应用起来更加方便,在时间序列分析中得到了更广泛的应用,我们进行统计分析的时候所说的时间序列的平稳性通常指的就是宽平稳。

### 2.3 平稳性的检验

对序列的平稳性有两种检验方法,一种是根据时序图和自相关图显示的特征做出判断的图检验方法;一种是构造检验统计量进行假设检验的方法。图检验方法是一种操作简便、运用广泛的平稳性判别方法,它的缺点是判别结论带有很强的主观色彩,所以最好能用统计检验方法加以辅助判断。

#### 2.3.1 时序图检验

根据平稳时间序列均值、方差为常数的性质, 平稳序列的时序图应该显示出该序列始终在一个常数值附近随机波动, 而且波动的范围有界的特点. 如果序列的时序图显示出该序列有明显的趋势性或周期性, 那么它通常不是平稳序列. 根据这个性质, 对于很多非平稳序列, 可以通过查看它的时序图将其识别出来。

### 2.3.2 自相关图检验

平稳序列通常具有短期相关性. 该性质用自相关系数来描述就是随着延迟期数 $k$ 的增加, 平稳序列的自相关系数 $\rho_k$ 会很快地衰减向零. 反之, 非平稳序列的自相关系数 $\rho_k$ 衰减向零的速度通常比较慢, 这就是我们利用自相关图进行平稳性判断的标准。

### 2.3.3 ADF检验

对任一AR(p)过程:  $x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t$ , 它的特征方程为:  $\lambda^p - \varphi_1 \lambda^{p-1} - \cdots - \varphi_p = 0$ . 如果该方程所有的特征根都在单位圆内, 则序列 $\{X_t\}$ 平稳. 如果有一个单位根存在, 不妨设 $\lambda_1 = 1$ , 则序列 $\{X_t\}$ 非平稳, 且自回归系数之和恰好等于1:  $\varphi_1 + \varphi_2 + \cdots + \varphi_p = 0$

因而, 对于AR(p)过程我们可以通过检验自回归系数之和是否等于1来考察该序列的平稳性. 记 $\rho = \varphi_1 + \varphi_2 + \cdots + \varphi_p - 1$ , 若序列 $\{X_t\}$ 非平稳, 则至少存在一个单位根, 等价于 $\rho = 0$ , 则AR(p)过程单位根检验的假设条件可以确定为:

$$H_0: \rho = 0 (\{X_t\} \text{非平稳})$$

$$H_1: \rho < 0 (\{X_t\} \text{平稳})$$

构造ADF检验统计量:  $\tau = \frac{\hat{p}}{S(\hat{p})}$

通过蒙特卡洛方法, 可以得到检验统计量的临界值表. 显然DF检验是ADF检验在自相关阶数为1时的一个特例, 所以统称为ADF检验.

## 2.4 白噪声检验

### 2.4.1 白噪声序列的定义

拿到一个观察值序列之后, 首先是判断它的平稳性. 通过平稳性检验, 序列可以分

为平稳序列和非平稳序列两大类. 如果序列经过平稳性判断之后被认为是平稳的, 那么我们有一套比较成熟的平稳序列分析方法对其建模, 然而并不是所有的平稳序列都值得建模. 只有那些序列值之间具有密切相关关系、历史数据对未来发展有一定影响的序列, 才值得我们花时间去挖掘历史数据中的有效信息, 用来预测序列未来的发展. 为了确定平稳时间序列是否值得分析下去, 我们需要对平稳序列进行白噪声检验. 如果序列值彼此之间没有任何相关性, 那就意味着该序列是一个没有记忆的序列, 过去的行为对将来的发展没有丝毫影响, 这种序列称为白噪声序列, 它满足如下两条性质: (1)  $EX_t = \mu, \forall t \in T$ ;

$$(2) \gamma(s, t) = \begin{cases} \sigma^2, & s = t \\ 0, & s \neq t \end{cases}, \forall s, t \in T$$

### 2.4.2 白噪声检验

我们知道如果一个序列是纯随机序列，那么它的序列值之间应该没有任何相关关系，即满足  $\gamma(k) = 0, \forall k \neq 0$ 。这是一种理论上才会出现的理想状况。实际上，由于观察值序列的有限性，导致纯随机序列的样本自相关系数不会绝对为零。

Barlett证明，如果一个时间序列是纯随机的，得到一个观察期数为  $n$  的观察序列，那么该序列的延迟非零期的样本自相关系数将近似服从均值为零，方差为序列观察期数倒数的正态分布，即  $\hat{\rho}_k \sim N\left(0, \frac{1}{n}\right), \forall k \neq 0$ 。

根据Barlett定理，我们可以构造检验统计量来检验序列是否为白噪声序列。原假设：延迟期数小于或等于  $m$  期的序列值之间相互独立

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, \forall m \geq 0$$

备择假设：延迟期数小于或等于  $m$  期的序列值之间有相关性

$$H_1: \text{至少存在某个 } \rho_k \neq 0, \forall m \geq 0, k \leq m$$

$$Q \text{ 统计量: } Q = n \sum_{k=1}^m \hat{\rho}_k^2 \sim \chi^2(m)$$

$$LB \text{ 统计量: } LB = n(n+2) \sum_{k=1}^m \left( \frac{\hat{\rho}_k^2}{(n-k)} \right) \sim \chi^2(m)$$

当统计量的P值小于  $\alpha$  时，则可以以  $1-\alpha$  的置信水平拒绝原假设，认为该序列为非白噪声序列；否则，接受原假设，认为该序列为白噪声序列。

## 2.5 ARIMA (p, d, q) 模型

### 2.5.1 ARIMA (p, d, q) 模型结构

ARIMA (p, d, q) 模型称为差分自回归移动平均模型(Autoregressive Integrated Moving Average Model, 简记ARIMA)，ARIMA (p, d, q) 模型是由博克思(Box)和詹金斯(Jenkins)提出的著名时间序列预测方法，所以它又称为Box-Jenkins模型或博克思-詹金斯法。其中ARIMA (p, d, q) 模型称为差分自回归移动平均模型，它的自回归是AR，且自回归项数为  $p$ ；移动平均是MA，且移动平均项数为  $q$ ；并且把时间序列变为平稳时所做的差分次数是  $d$ ，那么所谓ARIMA模型，就是先把非平稳的数据转化为平稳的，进而再建立模型。ARIMA模型根据原序列是否平稳及回归中所包含部分的不同，所以它包括移动平均过程(MA)、自回归过程(AR)、自回归移动平均过程(ARMA)及ARIMA过程。其一般表达式为：

$$\begin{cases} W_t = \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \dots + \varphi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\ \theta_q \neq 0, \varphi_p \neq 0 \\ E(W_t \varepsilon_t) = 0 \\ \varepsilon_t \sim WN(0, \sigma^2) \end{cases}$$

其中， $W_t = \nabla^d Y_t = \nabla^{d-1} Y_t - \nabla^{d-1} Y_{t-1}$ ， $\varphi_1, \varphi_2, \dots$  为自回归系数， $\theta_1, \theta_2, \dots$  为移动平均系数，其中  $p$  表示  $p$  阶自回归， $q$  表示  $q$  阶滑动平均。综合考虑可知，自回归模型AR( $p$ )和滑动平均模型MA( $q$ )都是ARIMA (p, d, q) 模型的特殊形式。

### 2.5.2 ARIMA (p, d, q) 模型建模

ARIMA (p, d, q) 模型的建模步骤如下图：

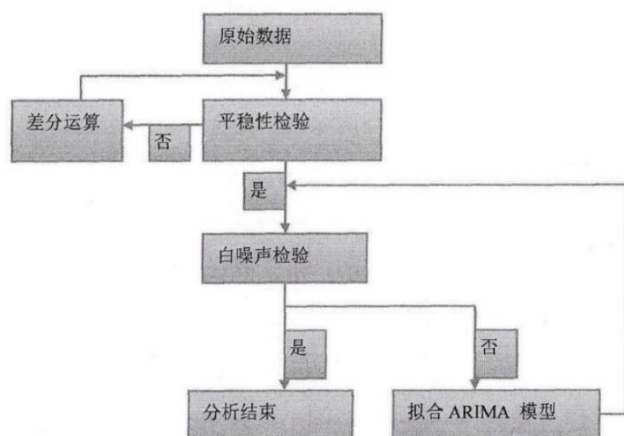


图1 ARIMA(p, d, q)模型建模步骤

首先检验原始数据的平稳性，如果原始数据非平稳，则将原始数据进行一阶、二阶、...、d阶差分直至差分后序列满足平稳性要求。接下来观察差分后序列的自相关图和偏自相关图。如果自相关图拖尾，偏自相关图p阶截尾则拟合模型ARI (p, d)。如果自相关图q阶结尾偏自相关图拖尾则拟合模型IMA(d, q)。如果自相关图和偏自相关图均拖尾，则拟合模型ARIMA (p, d, q)。接下来对差分后序列进行白噪声检验，若检验通过，则分析结束。

## 第三章 实证分析

### 3.1 恒瑞医药（600276）股票每日收盘价的描述性分析

为了对恒瑞医药（600276）股票价格进行分析，本文通过同花顺软件收集了该股票自2015年1月5日至2021年6月28日的每日股票价格的收盘价，合计1577个数据。观察股票收盘价数据,可知每日收盘价之间的差距并不大，在给出的2000个数据中，收盘价最高的为115.01元，收盘价最低的34.47元，极差为77.54元。没有明显的上升或者下降的趋势,而是在不断的波动。

```

      收盘
Min.    : 37.47
1st Qu.: 48.00
Median  : 63.81
Mean    : 66.07
3rd Qu.: 83.62
Max.    :115.01
    
```

图2 收盘价描述性统计分析

### 3.2 ARIMA模型的建立

#### 3.2.1 平稳性检验

为了方便计算,我们将2015年1月5日的数据记为第1个，1月6日的数据记为第2个，以此类推，则2021年6月28日的数据为第1577个。绘制恒瑞医药（600276）股票每日收盘价的时间序列图，如图3所示。

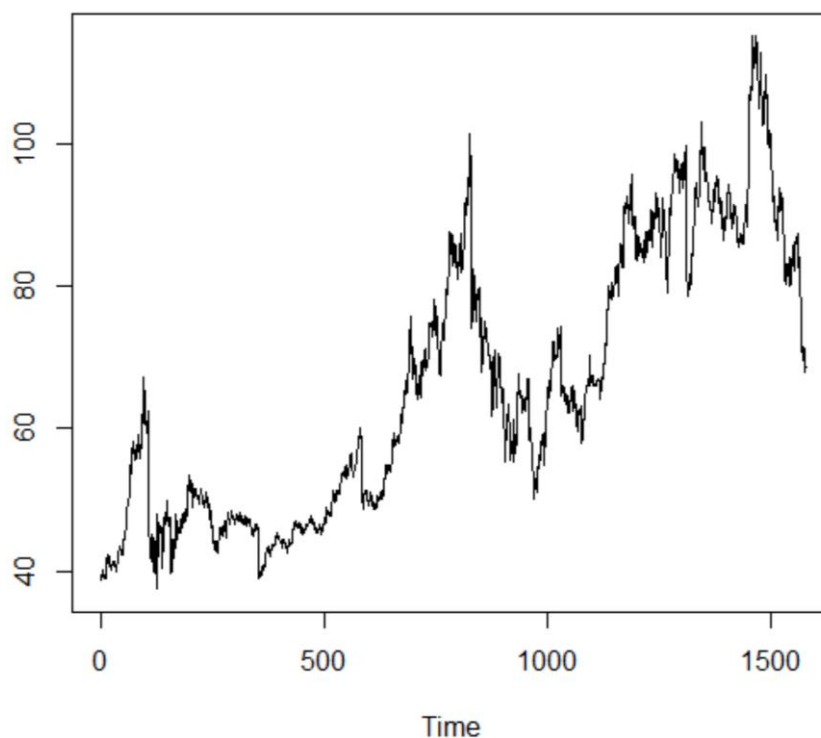


图3 恒瑞医药每日收盘价的时间序列图

观察时间序列图可知，从第600个数据到第750个数据总体呈现上升的趋



势，而第750个数据到第1000个数据，呈现明显的下降趋势，且波动幅度比较大，这之后的序列也总是在不断的波动。总之该时间序列图的总体趋势在不断的波动，波动幅度较为明显，所以初步判断该时间序列是非平稳的。

时间序列图判断序列的平稳性比较方便直观,是判断时间序列平稳性的第一步。但是，时间序列图对于稳定性的判断结果不是十分准确。只通过时间序列图判断平稳性会造成偏差，为了进一步判断时间序列的稳定性,绘制了股票价格的自相关函数图，图4为恒瑞医药（600276）收盘价的自相关函数图。

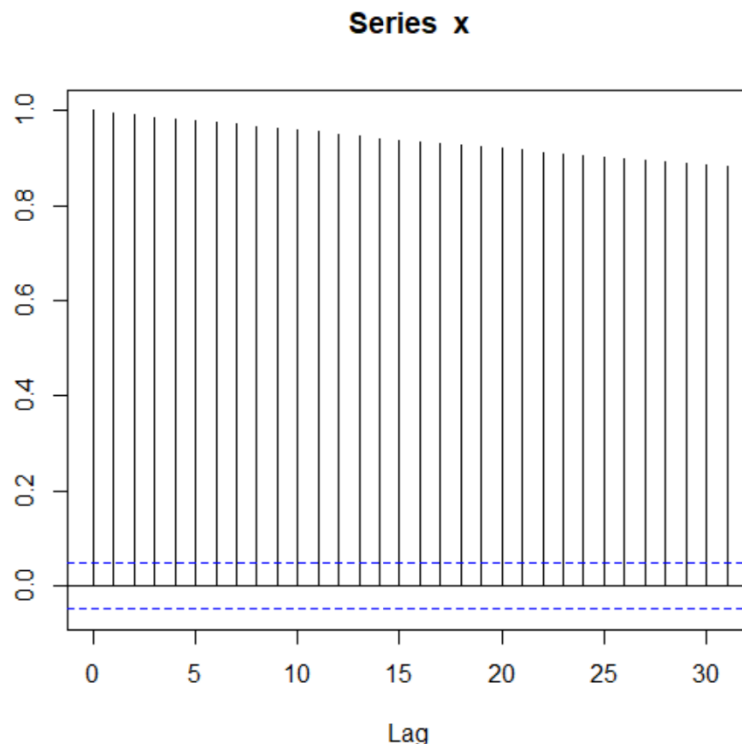


图4 恒瑞医药收盘价的自相关函数图

可以看出，延迟33阶期自相关函数均超过两倍标准差范围，没有增加或者下降的趋势，所以可以判定恒瑞医药（600276）股价的时间序列非平稳。

### 3.2.2非平稳序列的平稳化处理

由上述分析可知恒瑞医药（600276）每日股票的收盘价是非平稳的,为了做进一步的分析,我们需要将非平稳的时间序列化为平稳的时间序列来处理。在这里我们将恒瑞医药（600276）的收盘价进行一阶差分处理，差分后的时间序列图如图5所示：

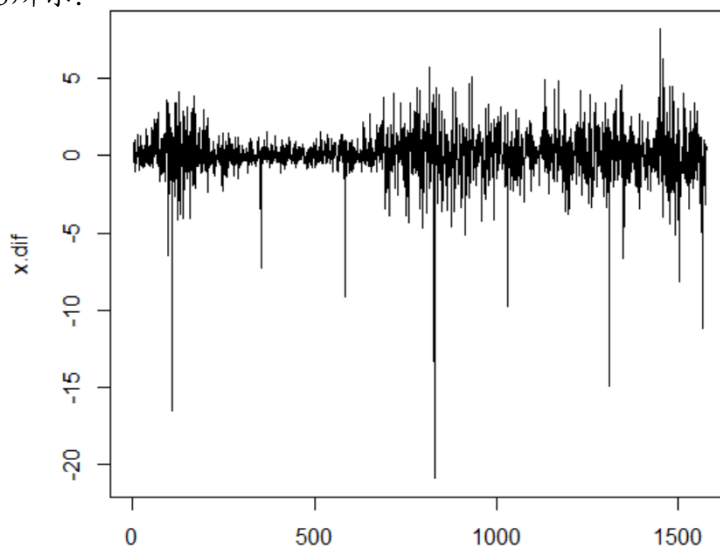


图 5 一阶差分后的每日收盘价时序图

观察差分后的时间序列图可知，一阶差分之后的时间序列数据围绕着数值 0 上下波动，但波动幅度落差较大，由此可以初步判定，一阶差分后的时间序列是不平稳的。由第六章的理论介绍可知，ADF 单位根检验法可以有效的判断差分后序列的稳定性，所以本文选用 ADF 单位根检验法对差分后的序列平稳性进行判断。

Box-Pierce test

```
data: x.dif
X-squared = 0.15202, df = 1, p-value = 0.6966
```

图 6 一阶差分后时间序列的 ADF 单位根检验结果

图6为一阶差分后时间序列的ADF单位根检验结果，可知p值大于0.01,说明在0.05的显著性水平下，一阶差分后的时间序列不平稳，故进行二阶差分。

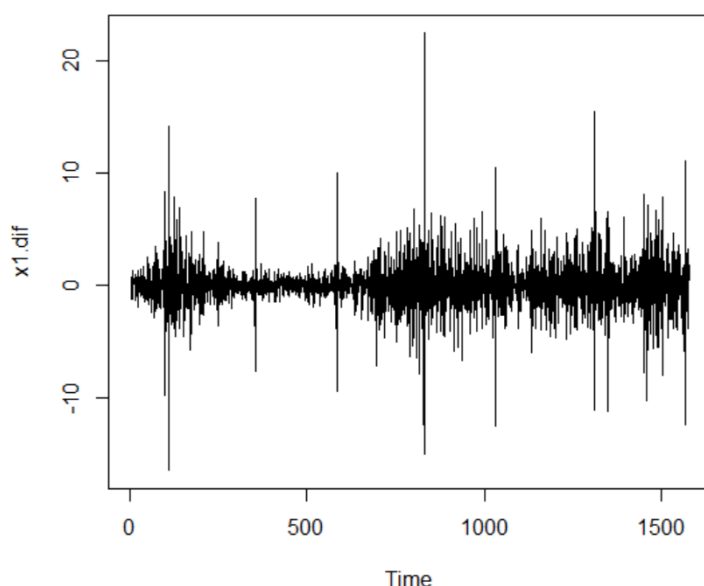


图 7 二阶差分后的时序图

由图7可以看出二阶差分后波动幅度有所减小，故继续进行ADF单位根检验，如图8所示。

Box-Pierce test

```
data: x1.dif
X-squared = 357.77, df = 1, p-value < 2.2e-16
```

图 8 二阶差分后时间序列的 ADF 单位根检验

由图8可以看出，二阶差分后的时间序列的p值远小于0.01，说明在0.05的显著水平下，拒绝存在单位根的原假设，即二阶差分后的时间序列是平稳的。

### 3.2.3模型识别

根据时间序列的性质及时间序列数据所提供的信息，提出一个合适的模型，对时间序列数据进行分析,这一过程称为模型的识别<sup>25</sup>。对于平稳的时间序列数

据,一般通过时间序列的自相关函数(ACF函数)和偏自相关函数(PACF函数)的拖尾性和截尾性可以对模型的形式进行判断。若其样本的自相关函数,当时,自相关函数值近似为0,可以初步判断该时间序列是MA(q)序列,可以建立q阶滑动平均模型。若样本的偏自相关函数,在 $k > p$ 时,偏自相关函数值近似为0,则可以初步判断该时间序列是AR(p)序列,可以建立p阶自回归模型。然而当样本的自相关系数和偏自相关系数没有明显的截尾性,该时间序列很可能是ARMA过程,这时的滞后阶数p和q比较难以判断,可以通过依次从较小的阶数到较大的阶数逐个取值的方法来进行尝试。

对于时间序列的分析,模型的定阶是十分重要的一步。观察一阶差分序列的自相关函数图,如图6,可以看出滞后期为2期的自相关函数第一次落在了两倍标准差范围之外。由一阶差分序列的偏自相关函数图如图7,可知,在滞后期为2期、5期、14期、20期、21期、25期、27期、28期、29期、31期的时候,样本的偏自相关函数落在了两倍标准差范围之外,其余滞后期的系数都落在两倍标准差范围内,所以通过差分序列的自相关函数与偏自相关函数不能直接确定模型的参数。为了能够准确的选取模型参数,本文通过将参数从低阶到高阶逐个取值的方法,然后根据AIC准则和修正的AIC系数选取最合适的模型。

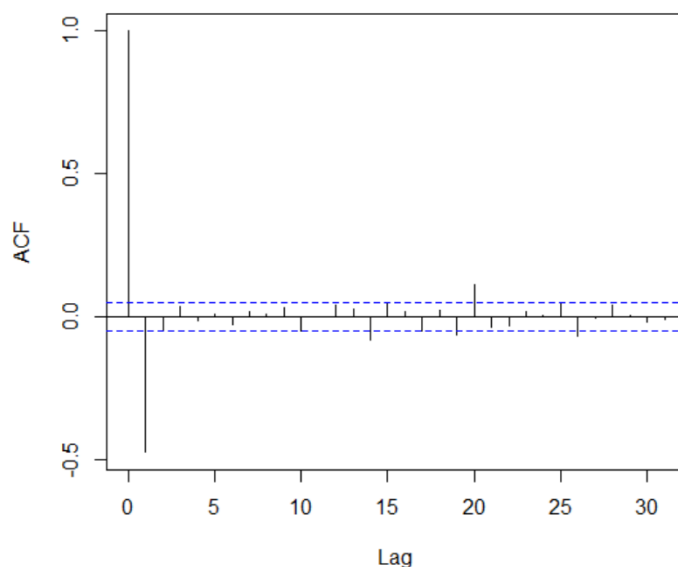


图9 一阶差分后的自相关函数图

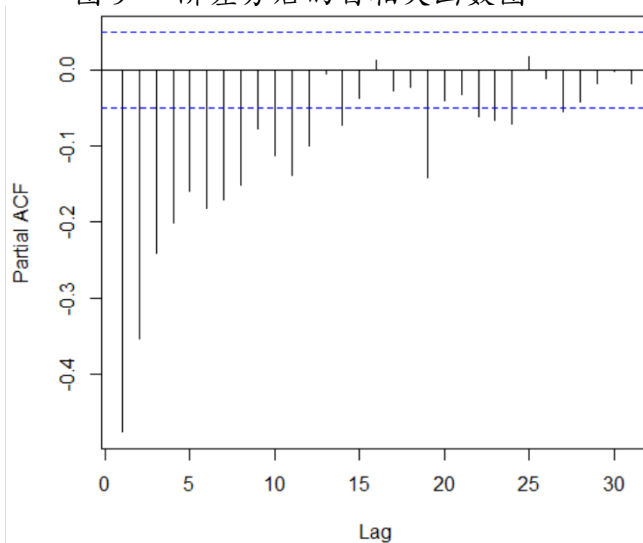


图 10 一阶差分后的偏自相关函数图

### 3.2.4 AIC 准则

AIC 准则是由日本统计学家 Akaike 于 1973 年提出的,它的全称是最小信息量准则(Akaike information criterion)。

该准则的指导思想是一个拟合模型的优劣可以从两方面去考察:-方面是大家非常熟悉的常用来衡量拟合程度的似然函数值;另一方面是模型中未知参数的个数.通常似然函数值越大,说明模型拟合的效果越好.模型中未知参数个数越多,说明模型中包含的自变量越多,自变量越多,模型变化越灵活,模型拟合的准确度就会越高.模型拟合程度高是我们所希望的,但是我们又不能单纯地以拟合精度来衡量模型的优劣,因为这样势必会导致未知参数的个数越多越好.未知参数越多,说明模型中自变量越多,未知的风险越多.而且参数越多,参数估计的难度就越大,估计的精度也就越差.所以一个好的拟合模型应该是拟合精度和未知参数个数的综合最优配置.

AIC 准则就是在这种考虑下提出的,它是拟合精度和参数个数的加权函数:

$$AIC = -2\ln(\text{模型的极大似然函数值}) + 2(\text{模型中未知参数个数})$$

### 3.2.5 参数估计与模型的建立

由于序列的自相关函数图和偏自相关函数图无法直接判断模型的阶数,我们直接调用 R 语言中 forecast 包中的 auto.arima() 函数,对序列进行模型的自动拟合,对不同的 p 和 q 可得到 AIC 函数的值和修正的 AIC 系数,比较修正的 AIC 系数,取使得修正的 AIC 值最小 p 和 q 为模型的参数。

由图 11 可知,最后得出,在  $p=5, q=0$  时,我们得到的修正的 AIC 函数值最小。且由图 11 中 ARIMA(5, 2, 0) 模型的参数可知恒瑞医药(600276)股票的价格主要受其滞后 1 期的价格影响和滞后 1 期的随机波动影响较大,所得到的模型所有参数都显著不为 0,所以可以初步认为该模型是合适的。

```
Series: x1.dif
ARIMA(5,0,0) with zero mean

Coefficients:
          ar1          ar2          ar3          ar4          ar5
      -0.8122   -0.6756   -0.4867   -0.3256   -0.1599
s.e.      0.0249    0.0312    0.0333    0.0311    0.0249

sigma^2 estimated as 3.681:  log likelihood=-3259.11
AIC=6530.23   AICc=6530.28   BIC=6562.4
```

图 11 建立 ARIMA 模型的参数

此外,由于 ARIMA 模型经常会出现一个问题:并不是所有的系数都是显著,也就是回归的系数可能等于 0。当存在不显著的系数的时候就需要进行剔除。图 12 是对 ARIMA (5, 2, 0) 置信区间的计算结果,发现没有置信区间包含 0 值,说明此模型各个系数都很显著,不需要进行剔除。

	2.5 %	97.5 %
ar1	-0.86092270	-0.76338263
ar2	-0.73663238	-0.61450524
ar3	-0.55195390	-0.42140453
ar4	-0.38664616	-0.26456566
ar5	-0.20864601	-0.11114402
intercept	-0.02775467	0.02698888

图 12 置信区间对显著性进行检验

### 3.3 模型诊断与残差分析

模型建立好以后，有必要对模型进行诊断，判断所建立的模型是否合适。残差序列真实值减去估计值之后而得到的序列，如果残差序列中还有信息没有被完全提取出来，则说明我们所建立的模型还没有充分提取样本数据的信息，需要继续对模型进行修正与补充。如何判断样本序列中的信息已经完全提取出来的？通常情况下，我们主要是对残差序列进行随机性检验。如果所得到的残差序列彼此之间相互独立，则表明残差序列之间没有影响，样本中的信息已经提取完毕，所建立的模型是合适的。模型诊断常用的有标准残差图检验法、残差的自相关函数图检验法和 Ljung-Box 随机性检验法。下图我们调用 `tsdiag()` 函数进行以上三个检验，检验结果如图 13 所示。

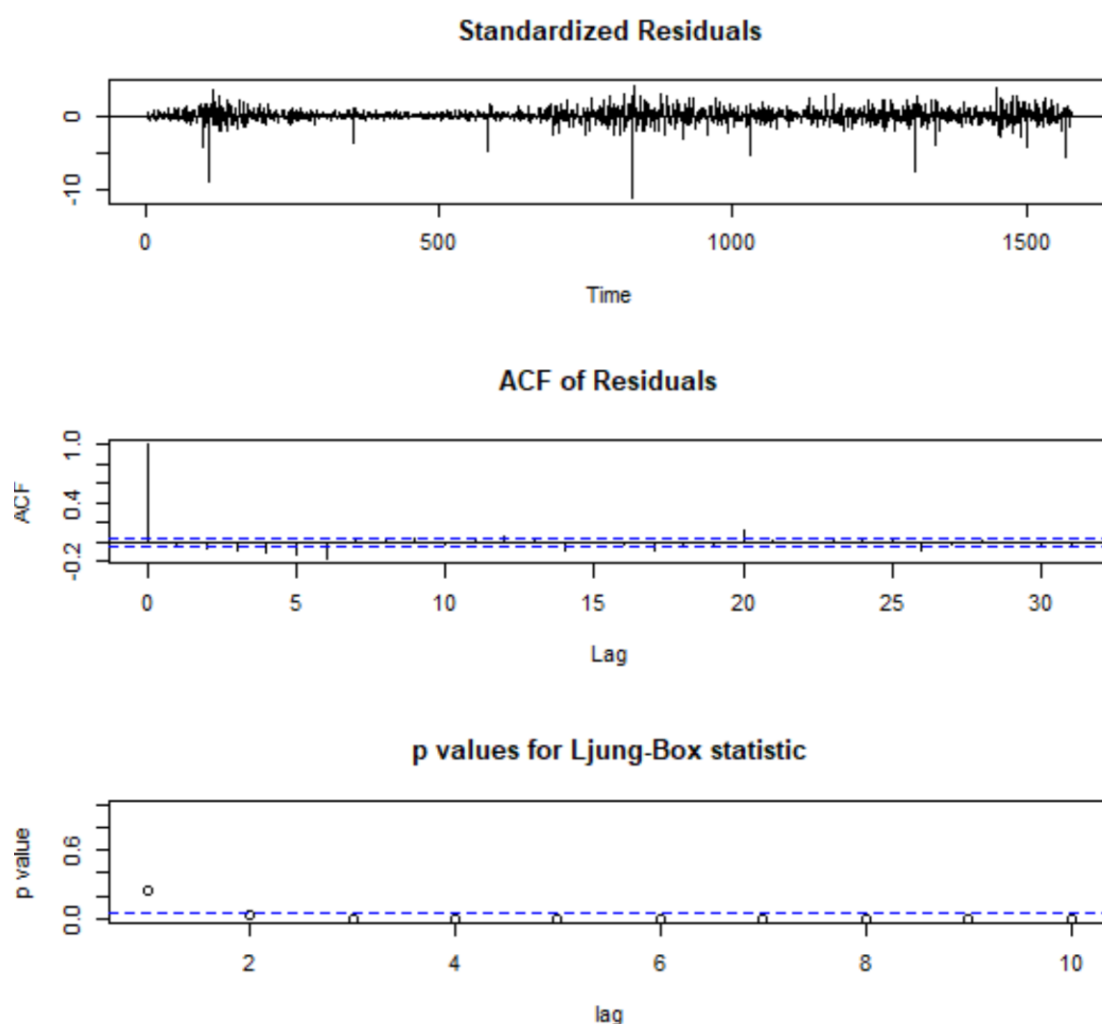


图 13 标准残差、残差自相关、L-B 检验结果

通过对以上这三个图形进行分析可对 ARIMA 模型进行诊断：

- (1) 标准化残差图。由图可以看出 ARIMA(5,2,0) 的标准化残差虽然有一定的聚集性但仍可用于预测分析。
- (2) 残差自相关图。由 ACF 图可以看出 ARIMA(5,2,0) 没有显著的自相关性。
- (3) Ljung-Box 统计量 p 值图。由图可以看出如果 Ljung-Box 统计量的 p 值几乎

都为 0，说明残差中的信息并未提取完全。故只能对数据按照以上的建模方法重新建模，进行三阶差分。三阶差分后的模型为  $ARIMA(3, 3, 3)$ ，时序图、自相关图、偏自相关图及残差检验图如下。

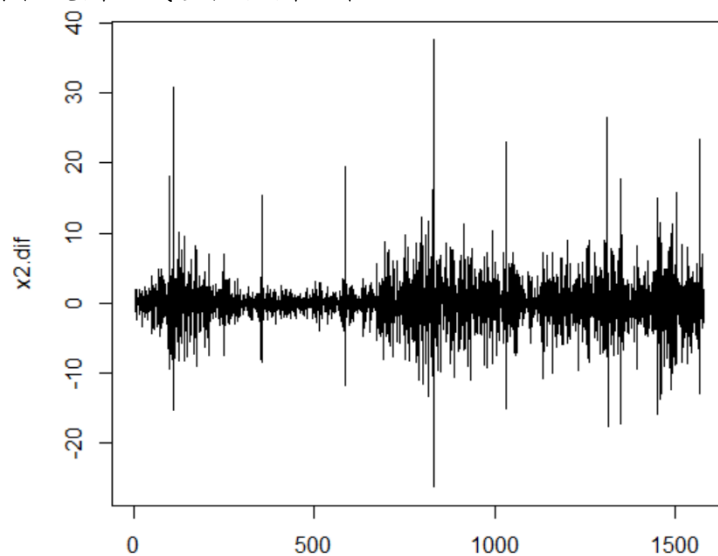


图 14 三阶差分后的时序图

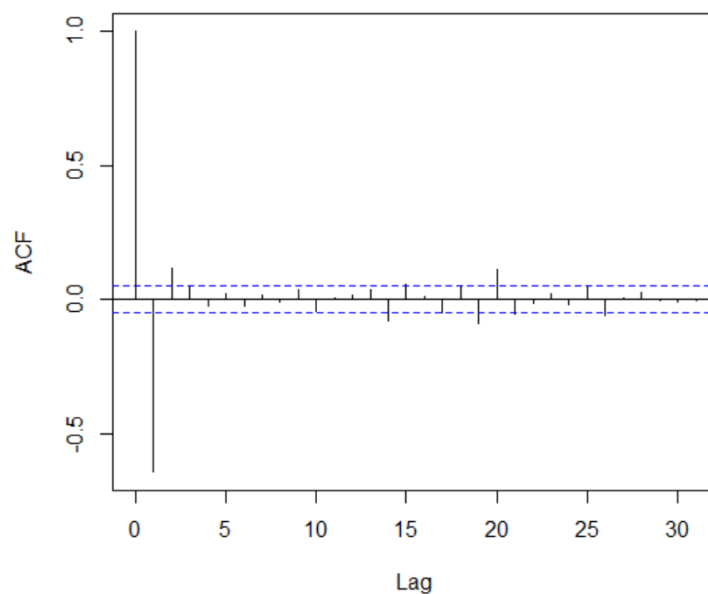


图 15 三阶差分后的自相关图

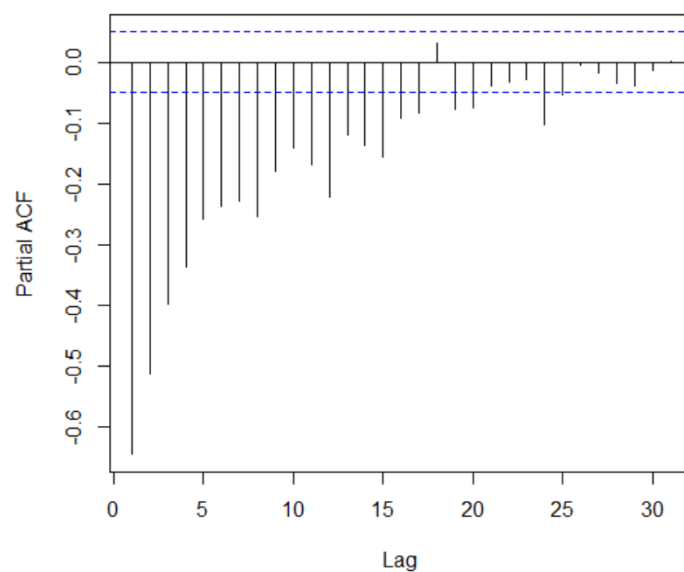


图 16 三阶差分后的偏自相关图

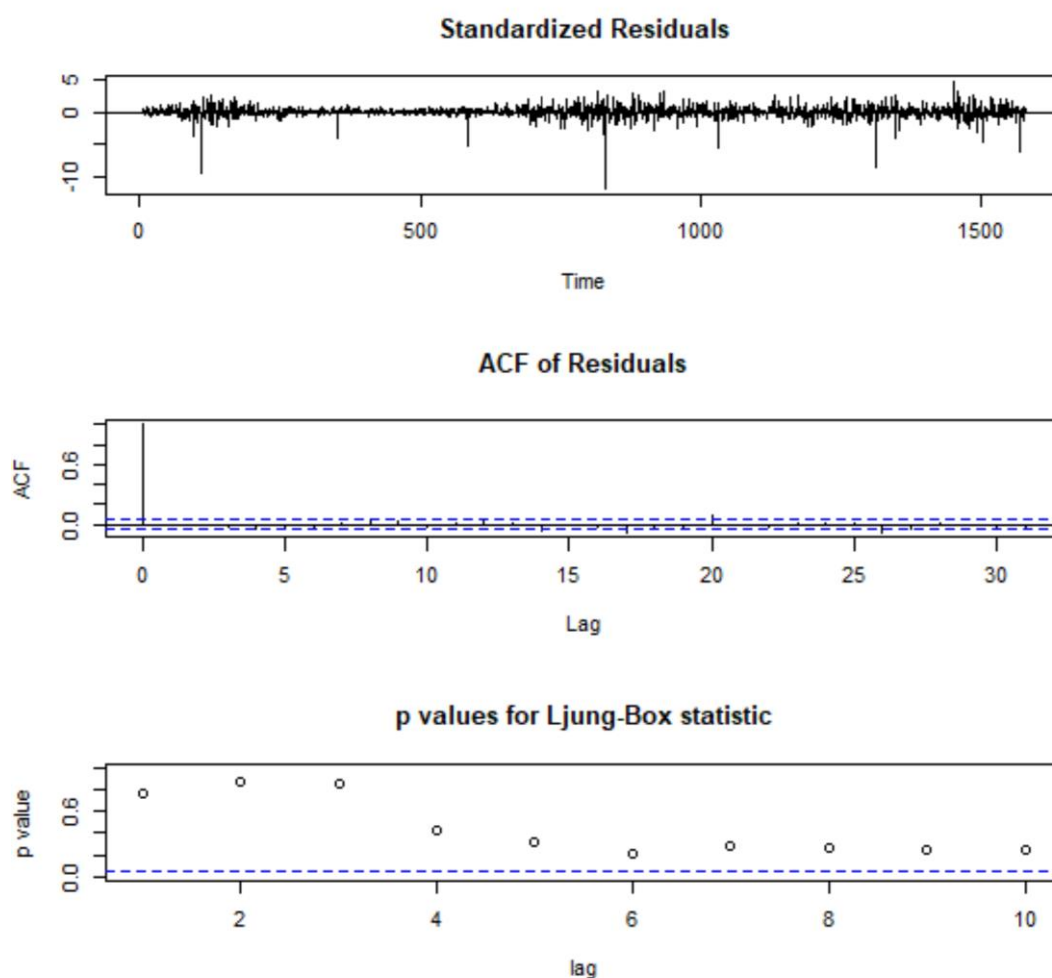


图 17 三阶差分后的残差检验图

由图 17 中 L-B 统计量的残差检验图可以看出，P 值均较大，说明原  $ARIMA(3, 3, 3)$  模型的有用信息已被提取完全，残差序列是白噪声序列。我们所建立的时间序列模型能够很好的描述恒瑞医药（600276）股票价格的变化情况。



### 3.4模型预测

对时间序列进行统计分析的主要目的之一是预测时间序列的发展状况,通过上述的分析,我们建立了恒瑞医药(600276)股票的时间序列模型。接下来我们用该模型对恒瑞医药(600276)股票价格自2020年6月29日的未来50个交易日收盘价数据进行预测,如图18、19所示。

```
> p
$pred
Time Series:
Start = 1578
End = 1627
Frequency = 1
[1] -1.51836203 1.00298228 0.10211638 -0.08451081 0.02506165 -0.02496279
[7] 0.02764257 -0.02720765 0.02659473 -0.02616219 0.02573942 -0.02532326
[13] 0.02490501 -0.02450226 0.02409791 -0.02370836 0.02331695 -0.02294016
[19] 0.02256130 -0.02219686 0.02183014 -0.02147764 0.02112267 -0.02078173
[25] 0.02043813 -0.02010837 0.01977577 -0.01945682 0.01913487 -0.01882640
[31] 0.01851474 -0.01821640 0.01791471 -0.01762617 0.01733412 -0.01705506
[37] 0.01677234 -0.01650246 0.01622877 -0.01596777 0.01570281 -0.01545040
[43] 0.01519390 -0.01494980 0.01470148 -0.01446543 0.01422501 -0.01399674
[49] 0.01376399 -0.01354325

$se
Time Series:
Start = 1578
End = 1627
Frequency = 1
[1] 1.761097 3.915928 4.252345 4.255903 4.256922 4.256925 4.256925 4.256927
[9] 4.256928 4.256929 4.256930 4.256931 4.256932 4.256933 4.256934 4.256934
[17] 4.256935 4.256936 4.256937 4.256937 4.256938 4.256939 4.256940 4.256940
[25] 4.256941 4.256941 4.256942 4.256942 4.256943 4.256944 4.256944 4.256945
[33] 4.256945 4.256945 4.256946 4.256946 4.256947 4.256947 4.256948 4.256948
[41] 4.256948 4.256949 4.256949 4.256949 4.256950 4.256950 4.256950 4.256951
[49] 4.256951 4.256951
```

图 18 未来 50 日的收盘价格预测

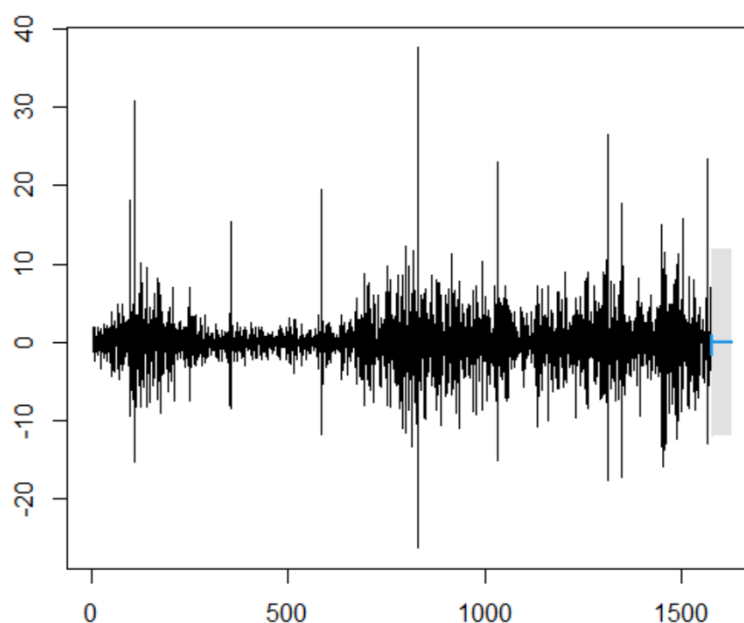


图 19 未来 50 日收盘价预测时序图



## 第四章 研究结论与展望

本文详细地介绍了时间序列模型的相关理论和建模步骤,并根据恒瑞医药(600276)股票价格的收盘价建立了相应的ARIMA(3,3,3)模型,并对2021年6月29日往后的50个交易日的股票价格进行了预测。对于恒瑞医药(600276)股票价格的预测分析可知现阶段恒瑞医药(600276)股票价格的变化比较平稳,波动比较小,发展状况良好。

由于股票序列是一个非线性的系统,因此用线性的ARIMA模型对其进行预测只能保证对于恒瑞医药(600276)每日收盘价的短期预测误差比较小,能够较好的预测股票每日收盘价的发展趋势,而对于长期的预测,效果一定不太理想。因此必须不断更新数据,然后再进行短期预测。

股票是金融市场的重要组成部分,随着中国股票行业的不断发展,股票市场逐渐走向了成熟,股票的发行和交易反过来也会促进和推动国家经济更快更好的发展。股票一方面成为了衡量企业发展状况的一个标签,另一方面也是人们投资理财的一种方式方法。因此对于股票的研究是投资者和企业管理者都十分感兴趣和关注的问题。投资者希望能够对股票的价格进行预测,以获得更高的收益,管理者希望能够通过对股票价格的准确预测,制定相关的决策,使自身得到更好的发展。

如今随着数学理论和计算机技术研究的进一步深入,用于预测时间序列的方法越来越多,很多机器学习、神经网络算法也被运用进入金融时间序列的分析研究当中。本文所采用的ARIMA模型,在经济领域相对来说是效果较佳的一种预测方法。另外,股票市场的实际波动比较大,受很多的影响因干扰,有市场风险、政治风险、技术风险等。单纯仅靠一个模型去做预测是不够的,甚至有时候用同一指标的不同时间段的数据进行预测都会得出不同预测结果。因此,需要结合其他方法与经济事实进行综合判断。

## 参考文献

- [1]姜乐.基于时间序列的股票价格分析研究与应用[D].大连理工大学, 2015.
- [2]赵国顺.基于时间序列分析的股票价格趋势预测研究[D].厦门大学, 2009.
- [3]周永道,王会琦, 吕王勇.时间序列分析及应用.高等教育出版社, 2015

## 附录

代码:

```
data<-read.csv('data.csv')
summary(data)
x<-ts(data$收盘)
plot(x)
acf(x)
pacf(x)
x.dif<-diff(x)
acf(x.dif)
pacf(x.dif)
Box.test(x.dif)
x1.dif<-diff(x.dif)
plot(x1.dif)
Box.test(x1.dif)
acf(x1.dif)
pacf(x1.dif)
install.packages('forecast')
library(forecast)
auto.arima(x1.dif)
m<-arima(x1.dif,order=c(5,0,0))
confint(m)
tsdiag(m)
x2.dif<-diff(x1.dif)
Box.test(x2.dif)
plot(x2.dif)
acf(x2.dif)
pacf(x2.dif)
library(forecast)
auto.arima(x2.dif)
m1<-arima(x2.dif,order=c(3,0,3))
confint(m1)
m1<-arima(x2.dif,order=c(3,0,3),fixed=c(NA,0,NA,NA,NA,NA))
tsdiag(m1)
p<-predict(m1,n.ahead=50)
```

p