

# 运用 ARIMA 方法预测上证指数

史杰灵 19121663

摘要：股票价格涉及很多不确定因素,且各个因素之间的相关关系错综复杂,因此要从理论上彻底弄清楚股市的变化机理十分困难.然而股市是一个运动的、特殊的系统,它必然存在着规律。以上证综合指数为例,利用 ARIMA 原理,结合上证指数数据时序数列平稳性校验与白噪声检验来进行建模,从而对上证指数进行短期预测。

## 一. 简介

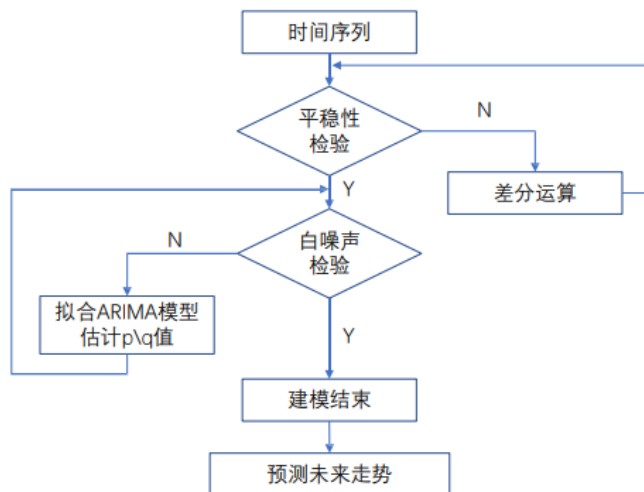
上海证券综合指数简称“上证指数”或“上证综指”，其样本股是在上海证券交易所全部上市股票，包括 A 股和 B 股，反映了上海证券交易所上市股票价格的变动情况，自 1991 年 7 月 15 日起正式发布。从 1990 年至 2010 年，上海证券交易所(以下简称“上交所”)从最初的 8 只股票、22 只债券，发展为拥有 894 家上市公司、938 只股票、18 万亿股票市值的股票市场，拥有 199 只政府债、284 只公司债、25 只基金以及回购、权证等交易品种，初步形成以大型蓝筹企业为主，大中小型企业共同发展的多层次蓝筹股市场，是全球增长最快的新兴证券市场。通过这个学期研讨课的学习，我们学到了一些关于人工智能的知识，其中就包含了人工智能的学习预测。而我们小组的就以之前的上证指数为样本，通过 ARIMA 建模的手段，来预测未来 70 天的走向趋势。

## 二. 数据设置

要基于 ARIMA 的时序信号分析应用来对上证指数进行预测需要非常庞大的过去上证指数的数据，我们使用了老师提供的的数据：从 1990 年 12 月至 2019 年 2 月日上证指数来辅助 ARIMA 的时序信号分析。

## 三. 建模流程

1. 时间序列建模的第一步是考虑现有季节（固定时间段内的重复模式）和/或趋势（数据中的向上或向下移动）。同时应考虑到平稳性的要求，通过差异数



差分平稳时间序列ARIMA模型建模步骤

据或线性回归方法来优化。

2. 时间序列建模的第二步是建立一个基准水平预测。我们还应该注意，大多数基准级预测不需要将数据固定的第一步。为了捕获时间序列模型中自相关的影响，有必要实施自回

归整合移动平均（或 ARIMA）模型。

3. 虽然我们可以看到提供的每个模型的精度都有所提高，但从视觉上确定哪个模型具有最佳精度并不总是可靠的。由此引入平稳性：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，且平稳性要求序列的均值和方差不发生明显变化。

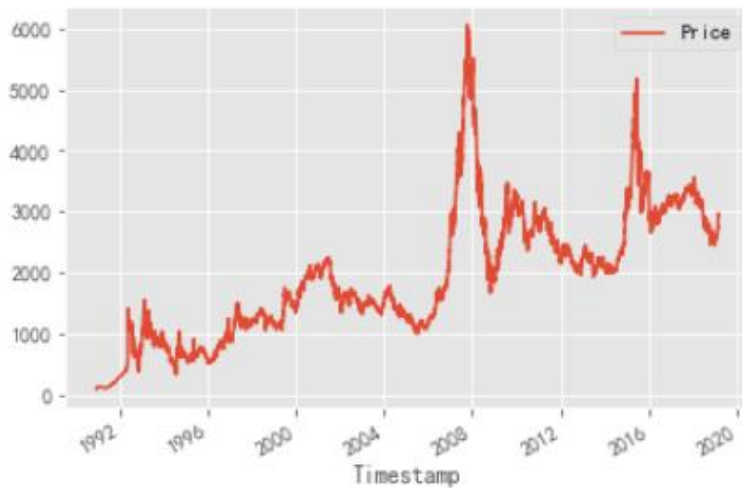
## 四. 分析过程

### 1. 导入库

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

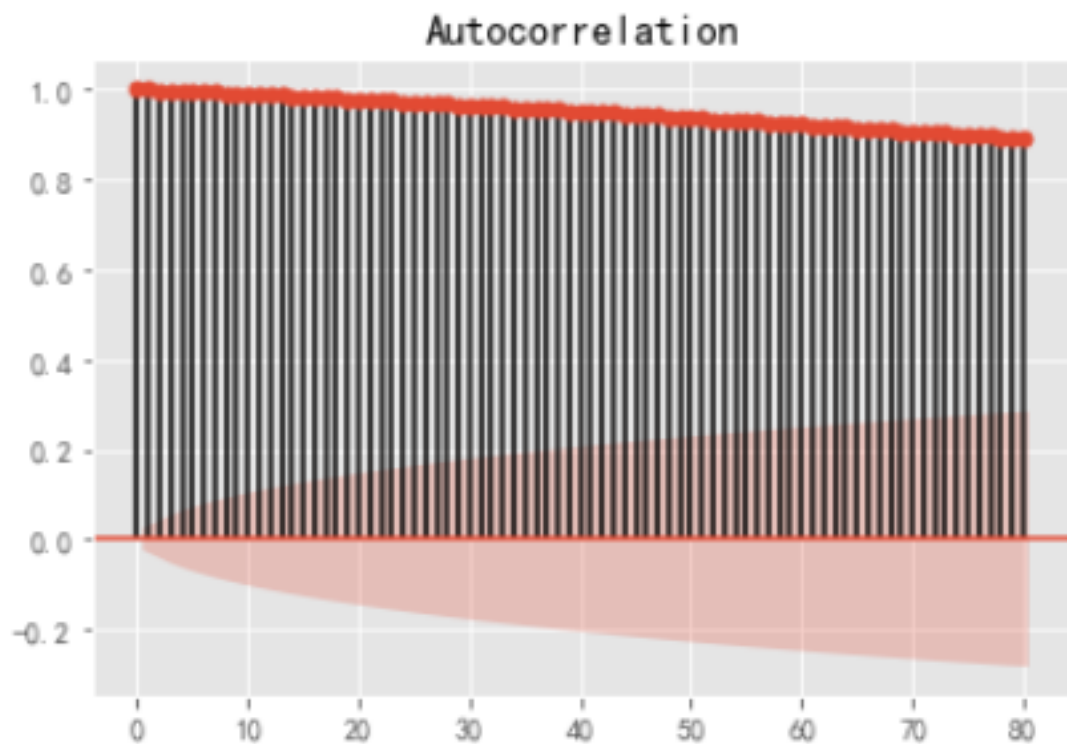
### 2. 检查原始序列（时序图）

```
In [5]: plt.figure(figsize=(10,5))
        sale.plot()
        plt.show()
```



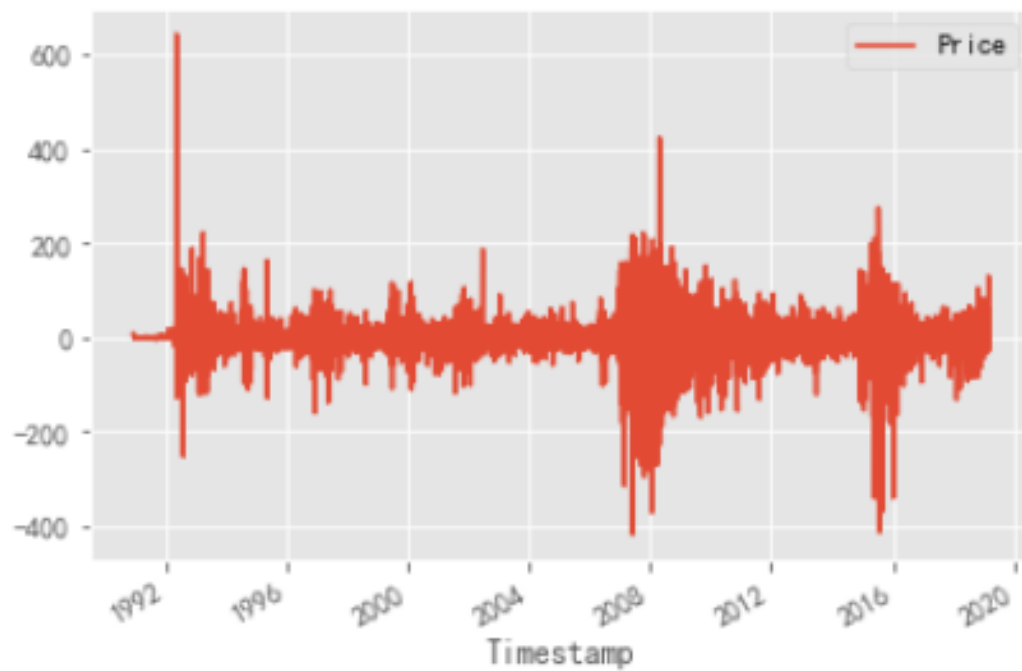
### 3. 检查原始序列（自相关图）

```
plot_acf(Shanghai Stock Exchange Index, lags=80).show()
```



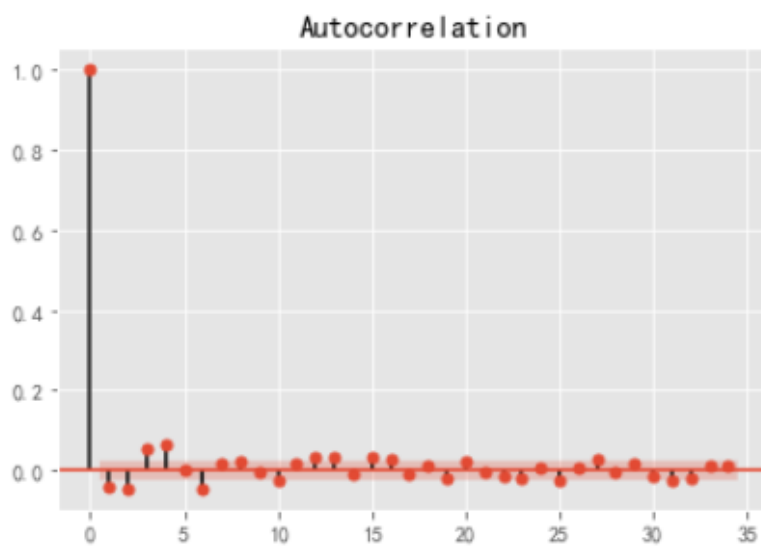
### 4. 一阶差分序列（时序图）

```
d1_sale=sale.diff(periods=1, axis=0).dropna()
plt.figure(figsize=(10,5))
d1_sale.plot()
plt.show()
```



#### 5. 一阶差分序列（自相关图）

```
plot_acf(d1_sale, lags=34).show()
```



#### 6. 确定阶数（p\q 值）

（一）使用 AIC 方法

```

pmax=int(len(d1_sale)/600)
qmax=int(len(d1_sale)/600)

aic_matrix=[]
for p in range(pmax+1):
    tmp=[]
    for q in range(qmax+1):
        try:
            tmp.append(ARIMA(sale, (p,1,q)).fit().aic)
        except:
            tmp.append(None)
    aic_matrix.append(tmp)
aic_matrix=pd.DataFrame(aic_matrix)
p,q=aic_matrix.stack().idxmin()

```

通过 AIC 方法得出的最优 p 值为 11，q 值为 9

### （二）使用人工识图方法

图 4-4、4-5、4-6 为使用人工识图进行参数调优产生的自相关图、lags=20 的偏自相关图、lags=34 的偏自相关图；由图可知自相关图，1 阶截尾；偏自相关图，拖尾。则  $ARIMA(p, d, q)=ARIMA(0, 1, 1)$ 。

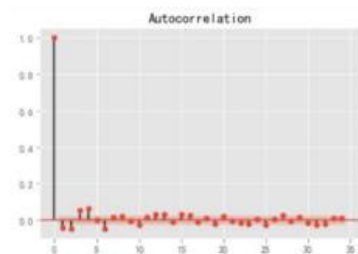


图 4-4

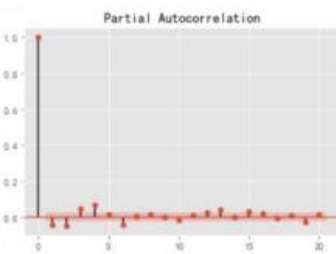


图 4-5

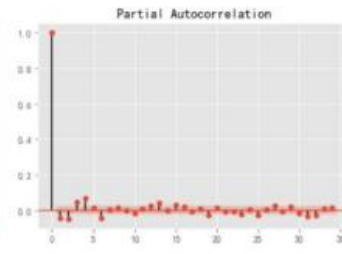
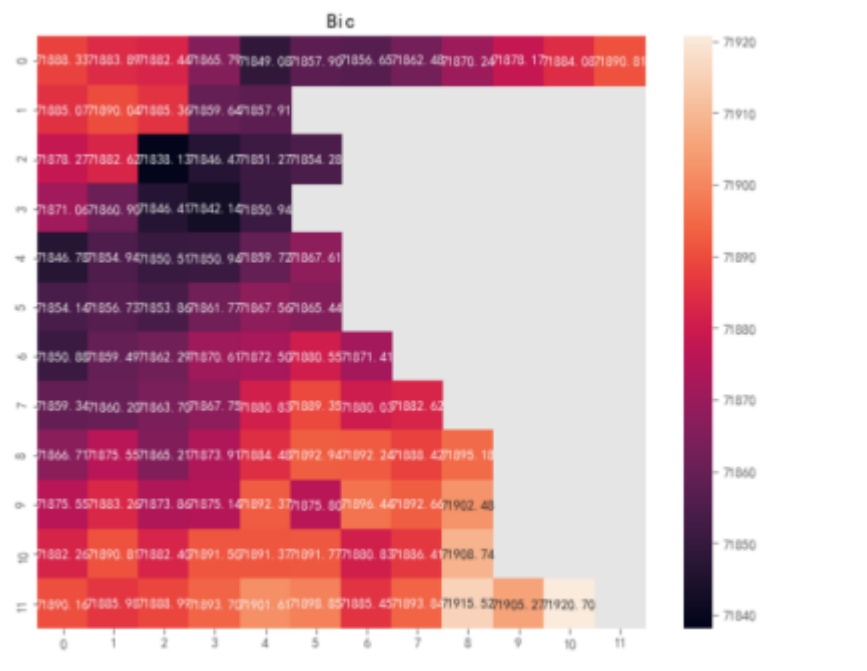


图 4-6

### （三）使用 BIC 方法

调优结果如图所示：



## 7. 建立模型

```
model=ARIMA(sale, (2,1,2)).fit()
model.summary2()
```

Model:	ARIMA	BIC:	71838.1328
Dependent Variable:	D.Price	Log-Likelihood:	-35893.
Date:	2021-09-26 09:59	Scale:	1.0000
No. Observations:	6892	Method:	css-mle
Df Model:	5	Sample:	1
Df Residuals:	6887		3
Converged:	1.0000	S.D. of innovations:	44.208
No. Iterations:	25.0000	HQIC:	71811.253
AIC:	71797.1041		

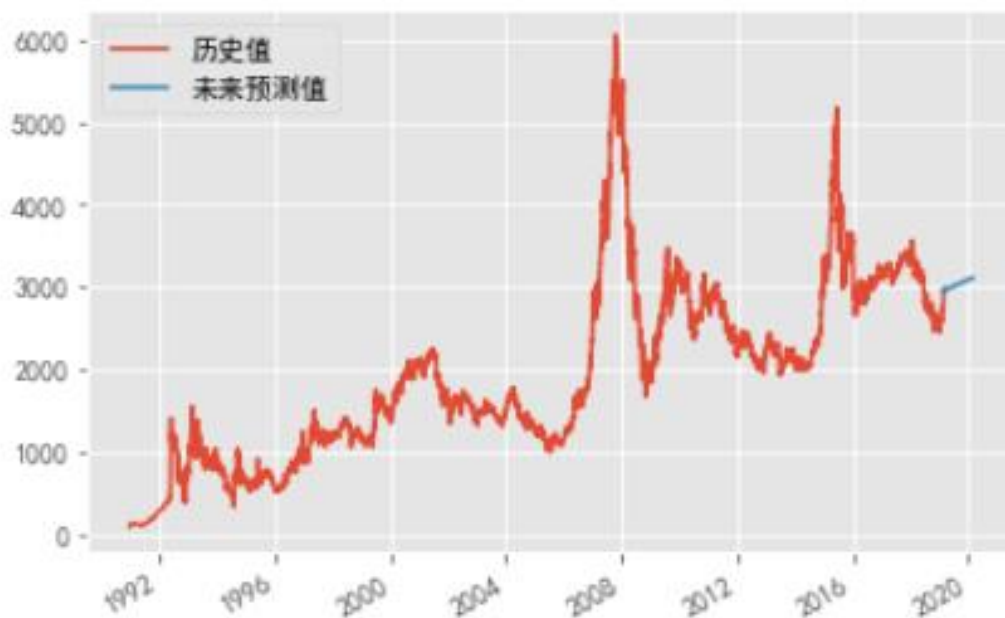
	Coef.	Std.Err.	t	P> t	[0.025	0.975]
const	0.4152	0.5038	0.8242	0.4098	-0.5722	1.4026
ar.L1.D.Price	0.0933	0.0452	2.0657	0.0389	0.0048	0.1819
ar.L2.D.Price	-0.8007	0.0460	-17.3887	0.0000	-0.8909	-0.7104
ma.L1.D.Price	-0.1422	0.0488	-2.9174	0.0035	-0.2378	-0.0467
ma.L2.D.Price	0.7575	0.0507	14.9288	0.0000	0.6580	0.8569

## 8. 未来预测

最后预测出未来 70 天的上证指数走向趋势：

```
forecast=pd.Series(model.forecast(70)[0],index=pd.date_range('2019-2-28',periods=70,freq='D'))
```

```
data=pd.concat((sale,forecast),axis=0)
data.columns=['历史值','未来预测值']
plt.figure(figsize=(100,20))
data.plot()
plt.show()
```



## 五. 结论

在本次研究中，我们通过过去几年来上证指数的数据绘制的时序图，进行差

分运算，使其符合平稳性要求。再通过 AIC、BIC、人工识图等参数优化方法，使其通过白噪声检验，进行模型建立并成功完成了对上证指数的预测。