

# Model and Forecast for China's GDP from 1953 to 2015 Based on the ARIMA Model

Xingyu Lang, Xin Li, Yiming Sun, Xiaodong Bai\*

Dalian Minzu University, Dalian Liaoning  
Email: \*baixd518@126.com

Received: Jun. 10<sup>th</sup>, 2017; accepted: Jun. 25<sup>th</sup>, 2017; published: Jun. 30<sup>th</sup>, 2017

---

## Abstract

This paper uses random time series method to analyze the data of China's GDP from 1953 to 2015. By drawing, model recognition, parameter estimation, model estimation and model fitting, we establish GDP time series model and utilize it to predict the economic development of China in the next 3 years, and thus provide the basis for the government to formulate the economic development strategy.

## Keywords

Gross Domestic Product, Time Series, ARIMA Model, Forecast

---

# 基于ARIMA模型对中国1953~2015年GDP的分析建模及预测

郎星宇, 李 鑫, 孙弋茗, 白晓东\*

大连民族大学, 辽宁 大连  
Email: \*baixd518@126.com

收稿日期: 2017年6月10日; 录用日期: 2017年6月25日; 发布日期: 2017年6月30日

---

## 摘 要

本文从《中国统计年鉴》中选取中国1953年至2015年共63年的GDP作为数据, 运用时间序列分析的基本分析方法——随机时序分析, 对数据进行绘图分析、模型识别、参数估计、模型估计、模型拟合和检验, 建立GDP时间序列模型, 并对中国未来3年的经济发展做出预测, 为政府制定经济发展战略提供依据。

\*通讯作者。

## 关键词

国内生产总值GDP, 时间序列, ARIMA模型, 预测

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 前言

上世纪 80 年代初, 中国开始研究联合国国民经济核算体系的国内生产总值(GDP)指标。中国于 1985 年开始建立 GDP 核算制度。1993 年, 中国正式取消国民收入核算, GDP 成为国民经济核算的核心指标。23 年国家统计局宣布中国将改进 GDP 核算与数据发布制度, 取消容易引起误解的预计数, 建立定期修正和调整 GDP 数据的机制, 在发布 GDP 数据的同时发布相关的重要数据, 必要时还将公布核算方法。这是中国提高 GDP 数据的准确性和透明度, 向国际通行办法迈进的重要一步[1]。

2014 年国家统计局将积极稳妥的推进国家统一核算地区生产总值, 深化固定资产投资统计, 加快改进能耗统计进一步完善社会消费品零售统计, 同时将精心组织实施第三次全国经济普查认真做好普查登记。尽快制定经济核算图, 指定全国统一的核算办法, 为 2015 年正式实施全国统一的核算 GDP 来打下一个基础。此举将有效消除近 10 年来各省 GDP 总和与国家统计局核算的全国 GDP 存在较大出入的情况。

## 2. 问题分析

中国 1953~2015 年的 GDP 数据可以看成是一个时间序列, 首先通过其时间序列图判断其是否为平稳时间序列, 若不是则需要先进行对数据的预处理, 之后初步提出分析模型, 不断通过对模型识别、参数估计以及显著性检验等建立最优时间序列模型, 并用来预测中国未来 3 年 GDP 的走势。

## 3. 名词解释

1) AIC 准则: 该准则适用于模型的适用性检验。其定义为: 样本外预测误差方法的有效估计量, 但受自由度约束较为严重, 常用于预测模型的选择。

2) GDP: 是指在一定时期内一个国家或地区经济中所生产出的全部最终产品和劳务的价值, 常被公认为衡量国民经济发展、判断宏观经济运行状况的最佳指标。

3) ARIMA 模型: 自回归模型和滑动平均模型的组合, 便构成了用于描述平稳随机过程的自回归滑动平均模型 ARMA, ARIMA 模型的实质就是差分运算与 ARMA 模型的组合。公式为:

$$\begin{cases} \phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t. \\ E(x_t \varepsilon_t) = 0, \forall s < t \end{cases}$$

## 4. 模型的假设

1) 中国 1953~2015 年没有出现或遭遇重大改革等事件。

2) 采集的数据统计口径相同无太大误差。

流程图: 如图 1 所示。

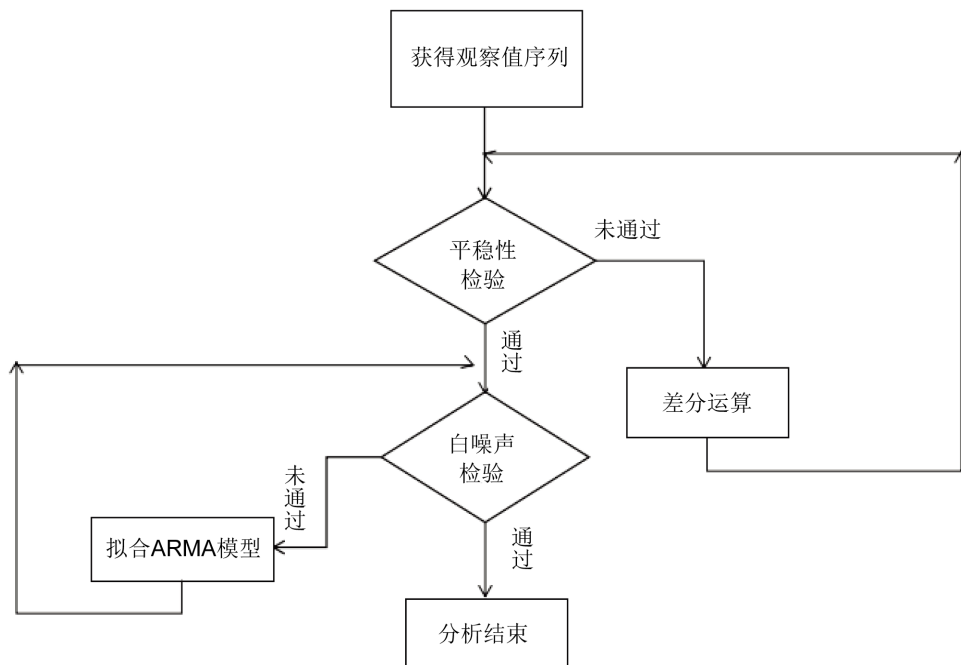


Figure 1. Flow diagram  
图 1. 流程图

## 5. 时间序列模型的建立

### 5.1. 序列平稳性检验

若一个随机过程的统计特性不随时间的推移而变化，则称它为平稳随机过程。平稳性是一些时间序列具有的统计特征，对数据进行平稳性检验是分析时间序列的关键步骤。平稳时间序列有两种定义，根据限制条件的严格程度，分为严平稳时间序列和宽平稳时间序列。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[2]。

#### 5.1.1. 时序图检验及自相关图检验

由图 2 可以看出该序列有明显的递增趋势，所以它一定不是平稳序列。从图可以看出 GDP 在 1978 年之前一直很平稳的在 2000~3000 亿元波动，但 1978 年后具有很明显的上升趋势，到 2000 年以后更是迅猛增长，可以看出原始序列显然是非平稳的。这很符合中国的国情，改革开放以后中国的经济快速发展，故 GDP 也指数式快速增长[3]。

图 3 中我们发现序列的自相关系数递减到零的速度相当缓慢，在很长的延迟时期里，自相关系数一直为正，而后又一直为负，在自相关图中显示三角对称性，这是具有单调趋势的非平稳序列的一种自相关图形式。该序列并不平稳。

#### 5.1.2. 差分

为了能够对序列进行分析，要使其平稳化。故将选择差分法，对序列进行平稳化处理，从而进一步分析预测。由差分的选择我们可以知道序列蕴含着曲线趋势，通常低阶(二阶或三阶)差分就可以提取出曲线趋势的影响，我们对原始数据进行一、二阶差分，并验证其平稳性。我们先进行一阶差分，即相距一期的两个序列值之间的减法运算( $\nabla x_t = x_t - x_{t-1}$ )。如图 4 所示。

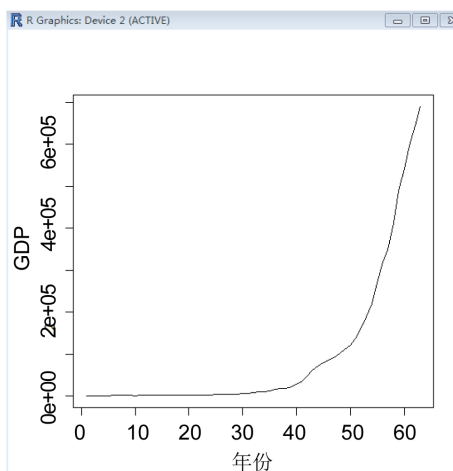


Figure 2. Sequence diagram  
图 2. 时序图

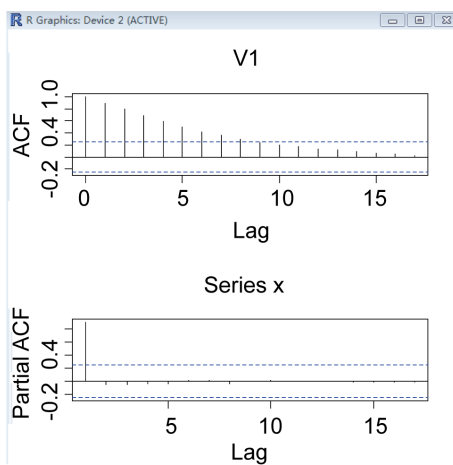


Figure 3. Autocorrelogram  
图 3. 自相关图

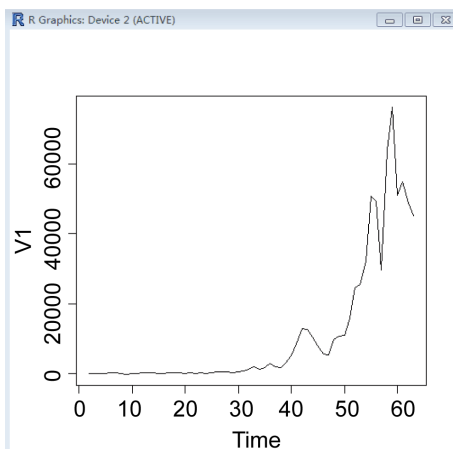


Figure 4. After the difference of sequence  $\{x_t\}$  sequence diagram  
图 4. 差分后的序列  $\{x_t\}$  的时序图

1 阶差分序列时序图显示, 1 阶差分提取了原序列中的部分长期趋势, 但长期趋势信息提取不充分, 1 阶差分后序列中仍蕴涵长期递增的趋势, 于是对 1 阶差分后序列再做一次差分运算。检验结果表明差分后的序列是非平稳的, 故还要再次进行差分计算, 即二阶差分:  $\nabla^2 x_t = \nabla x_t - \nabla x_{t-1}$ , 见图 5。

由图 6 可知, 我们可以确定二阶差分后序列平稳, 时序图显示 2 阶差分比较充分地提取了原序列中蕴涵的长期趋势, 使得差分后序列不再呈现确定性趋势。所以, 我们认为 ARIMA 模型的差分阶数  $d$  等于 2。

## 5.2. 纯随机性检验

在将数据平稳化之后, 还要判断序列是否有分析价值, 必须对序列进行纯随机性检验, 即白噪声检验。为了判断序列是否有分析价值, 必须对序列进行纯随机性检验, 即白噪声检验, 因此在建模之前需要进行纯随机性检验。若是到平稳的白噪声序列, 则该序列没有分析价值; 若是平稳非白噪声序列, 可进行模型拟合。

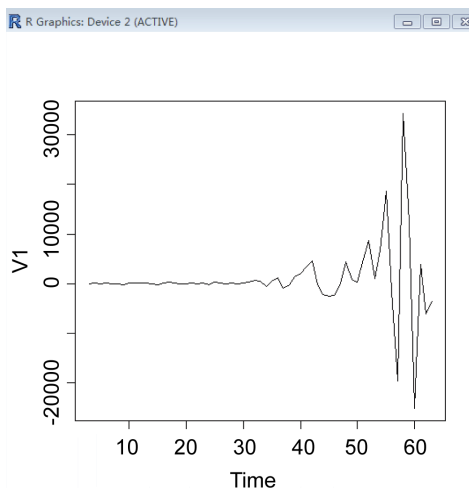


Figure 5. Second order difference sequence diagram

图 5. 二阶差分时序图

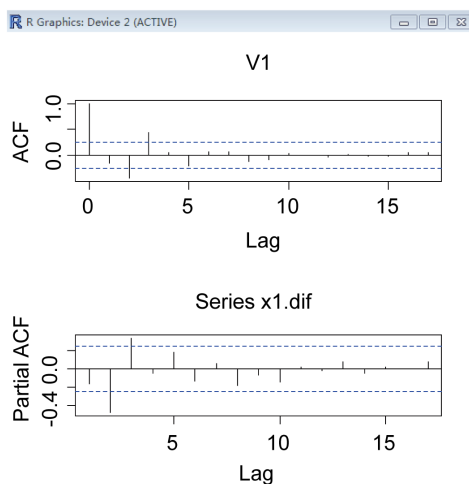


Figure 6. Auto correlation and partial autocorrelation map after difference

图 6. 差分后的自相关和偏自相关图

原假设：延迟期数小于或等于  $m$  期的序列值之间相互独立。

备择假设：延迟期数小于或等于  $m$  期的序列值之间有关联性。

当  $P$  值小于置信水平时，拒绝检验假设；而当  $P$  值大于置信水平时，不拒绝检验假设。在二阶差分后白噪声检验如图显示：

```
> for(i in 1:2) print(Box.test(x1.dif,type="Ljung-Box",lag=6*i))

Box-Ljung test

data:  x1.dif
X-squared = 30.368, df = 6, p-value = 3.346e-05

Box-Ljung test

data:  x1.dif
X-squared = 32.287, df = 12, p-value = 0.00125

> |
```

LB 统计量的  $P$  值 = 0.00125，小于 0.05，所以可以断定二阶差分序列属于非白噪声序列。结合前面的平稳性检验结果，说明该序列是平稳非白噪声序列，可进行模型拟合。

### 5.3. 时间序列模型的建立

#### 5.3.1. 系统定阶

为了尽量避免因个人经验不足而导致的模型识别不准确的问题，R 提供了 `auto.arima` 函数。如前所述，ARIMA 模型有  $(p, d, q)$  3 个参数，在此我们采用 `auto.arima` 函数定阶。

```
> auto.arima(x)
Series: x
ARIMA(2,2,1)

Coefficients:
          ar1      ar2      ma1
      -0.7156  -0.5553   0.7918
s.e.      0.1151   0.1102   0.0887

sigma^2 estimated as 31309493:  log likelihood=-612.16
AIC=1232.33   AICc=1233.04   BIC=1240.77
```

经过上述过程，拟合 ARIMA (2, 2, 1) 模型。

#### 5.3.2. 拟合模型检验

模型的显著性检验主要是检验模型的有效性。一个模型是否显著有效主要看他提取的信息是否充分。一个好的拟合模型应该能够提取观察值序列中几乎所有的样本相关信息，换言之，拟合残差项中将不再蕴涵任何相关信息，即残差序列应该为白噪声序列，这样的模型称为显著有效模型。反之，如果残差序列为非白噪声序列，那就意味着残差序列中还残留着相关信息未被提取，这就说明拟合模型不够有效，通常还需要选择其他模型，重新拟合。所以模型的显著性检验即为残差序列的白噪声检验[4]。

```

> x.fit1<-arima(x,order=c(2,2,1))
> x.fit1

Call:
arima(x = x, order = c(2, 2, 1))

Coefficients:
      ar1      ar2      ma1
-0.7156 -0.5553  0.7918
s.e.   0.1151   0.1102  0.0887

sigma^2 estimated as 29769682:  log likelihood = -612.16,  aic = 1232.33
> for(i in 1:2) print(Box.test(x.fit1$residual,lag=6*i))

      Box-Pierce test

data:  x.fit1$residual
X-squared = 1.2923, df = 6, p-value = 0.9721

      Box-Pierce test

data:  x.fit1$residual
X-squared = 2.9591, df = 12, p-value = 0.9958

```

LB 统计量  $P=0.9721$  和  $P=0.9958$  均显著大于  $\alpha$  ( $\alpha=0.05$ ), 可知残差通过了白噪声检验, 即认为残差序列为白噪声序列, 该拟合模型显著有效。

我们可以根据自相关图和偏自相关图拟合(3, 2, 1)模型, 检验结果:

```

> x.fit2<-arima(x,order=c(3,2,1))
> x.fit2

Call:
arima(x = x, order = c(3, 2, 1))

Coefficients:
      ar1      ar2      ar3      ma1
-0.6467 -0.5086  0.0866  0.7543
s.e.   0.1719   0.1383  0.1597  0.1206

sigma^2 estimated as 29608603:  log likelihood = -612.02,  aic = 1234.04
> for(i in 1:2) print(Box.test(x.fit2$residual,lag=6*i))

      Box-Pierce test

data:  x.fit2$residual
X-squared = 1.4732, df = 6, p-value = 0.9613

      Box-Pierce test

data:  x.fit2$residual
X-squared = 2.9354, df = 12, p-value = 0.996

```

拟合(3, 2, 2)模型，检验结果如下：

```
> x.fit3<-arima(x,order=c(3,2,2))
> x.fit3

Call:
arima(x = x, order = c(3, 2, 2))

Coefficients:
      ar1      ar2      ar3      ma1      ma2
 0.0688  0.0058  0.5023  0.0352 -0.5585
s.e.  0.3186  0.2377  0.1736  0.3377  0.2835

sigma^2 estimated as 29138857:  log likelihood = -611.55,  aic = 1235.11
> for(i in 1:2) print(Box.test(x.fit3$residual,lag=6*i))

      Box-Pierce test

data:  x.fit3$residual
X-squared = 1.1445, df = 6, p-value = 0.9795

      Box-Pierce test

data:  x.fit3$residual
X-squared = 3.4202, df = 12, p-value = 0.9918
```

P 值均显著大于  $\alpha$ ，可知残差均通过了白噪声检验，即认为残差序列为白噪声序列。

### 5.3.3. 模型优化

AIC 准则为选择最优模型提供了有效的规则，在所有通过检验的模型中使得 AIC 函数达到最小的模型为相对最优模型。

由图可知，ARIMA (2, 2, 1)模型使得 AIC 值最小，为 1232.33，所以 ARIMA (2, 2, 1)模型为相对最优模型。

## 5.4. 预测未来 3 年 GDP 走势及分析

所谓预测，就是要利用序列以观测到的样本值对序列在未来某个时刻的取值进行估计。目前对平稳序列最常用的预测方法是线性最小方差预测。线性是指预测值为观测值序列的线性函数，最小方差是指预测方差达到最小[5] [6]。

利用此模型对国内生产总值之后 3 年的 GDP 进行预测结果如下：

```
> x.fore<-forecast(x.fit1,h=3)
> x.fore

      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
2016      742438.7 735446.4 749431.1 731744.8 753132.6
2017      791907.4 775793.5 808021.3 767263.3 816551.4
2018      839566.3 815571.4 863561.2 802869.3 876263.3
```

可以从图 7 看出，模型拟合效果很好，国内生产总值在未来 3 年内还是会稳定地上涨。



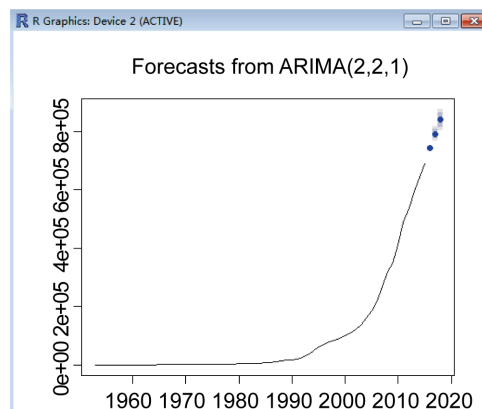


Figure 7. Prediction chart

图 7. 预测图

## 6. 模型的评价

### 6.1. 模型的优点

1) 本文巧妙运用流程图,将建模思路完整清晰的展现出来。

2) 国民生产总值(GDP)受经济基础、人口增长、资源、科技、环境等诸多因素的影响,这些因素之间又有着错综复杂的关系,因此,运用结构性的因果模型分析和预测 GDP 往往比较困难。将历年的 GDP 作为时间序列,根据过去的数据得出其变化规律,建立预测模型,有着重要的意义。

3) 建立模型的过程中运用了 R 统计软件,结果准确并提高了工作效率。

### 6.2. 模型的缺点

1) 本文没有考虑 GDP 数据的周期性和季节性,如果能从这方面详细研究,一定能对 GDP 序列的发展变化规律做出更为准确地分析。

2) 对 GDP 的预测本文只建立了 ARIMA 这一种模型,缺乏与其他模型比较。

3) 本文仅以 GDP 的变化为视角,并在该视角下,力图达到对经济运行较为准确的预测。但由该模型得出的预测结果只是一个预测值,而国民经济是一个复杂多变的动态系统,当国家的宏观政策发生调整,发展环境发生改变时都会使宏观经济指标出现相应地变化。

## 7. 总结

本文根据 1953~2015 年中国的国内生产总值(GDP)的统计资料,针对 GDP 的非平稳特征,通过差分变成平稳序列,建立 GDP 时间序列的 ARIMA 模型,并在此基础上用于国内生产总值 GDP 的预测分析。计算结果表明,该模型能较好地解决国内生产总值 GDP 的估计和预测问题,预测精度较为精确。

## 基金项目

辽宁省自然科学基金面上项目资助(201602188);国家级大创项目资助(G201612026032);校级大创项目(XA201609241)。

## 参考文献 (References)

- [1] 刘薇. 时间序列分析在吉林省 GDP 预测中的应用[D]: [硕士学位论文]. 长春: 东北师范大学, 2009
- [2] 王燕. 应用时间序列分析[M]. 第三版. 北京: 中国人民大学出版社.

- [3] 张丽. 天津市人均 GDP 时间序列模型及预测[J]. 北方经济, 2007(6): 44-46.
- [4] 刘自强, 王效岳, 白如江. 基于时间序列模型的研究热点分析预测方法研究[J]. 情报理论与实践, 2016, 39(5): 27-33.
- [5] 朱莹. 基于 ARMA 模型对江苏省 1979-2014 年 GDP 的分析建模及预测[J]. 江苏商论, 2017(2): 60-61.
- [6] 王坚, 汪卫. 一种基于 ARIMA 模型的钢铁产品销售预测方法[J]. 企业信息化技术, 2017(41): 18-23.

## 附录

### 附录 1. 原始数据

	A	B	C	D	E	F	G	H	I
1	年份	国内生产总值		年份	国内生产总值		年份	国内生产总值	
2		(亿元)			(亿元)			(亿元)	
3	1953		824.4	1974		2827.7	1995		61339.9
4	1954		859.8	1975		3039.5	1996		71813.6
5	1955		911.6	1976		2988.6	1997		79715
6	1956		1030.7	1977		3250	1998		85195.5
7	1957		1071.4	1978		3678.7	1999		90564.4
8	1958		1312.3	1979		4100.5	2000		100280.1
9	1959		1447.5	1980		4587.6	2001		110863.1
10	1960		1470.1	1981		4935.8	2002		121717.4
11	1961		1232.3	1982		5373.4	2003		137422
12	1962		1162.2	1983		6020.9	2004		161840.2
13	1963		1248.3	1984		7278.5	2005		187318.9
14	1964		1469.9	1985		9098.9	2006		219438.5
15	1965		1734	1986		10376.2	2007		270232.3
16	1966		1888.7	1987		12174.6	2008		319515.5
17	1967		1794.2	1988		15180.4	2009		349081.4
18	1968		1744.1	1989		17179.7	2010		413030.3
19	1969		1962.2	1990		18872.9	2011		489300.6
20	1970		2279.7	1991		22005.6	2012		540367.4
21	1971		2456.9	1992		27194.5	2013		595244.4
22	1972		2552.4	1993		35673.2	2014		643974
23	1973		2756.2	1994		48637.2	2015		689052.1

### 附录 2. 程序

```

a <- read.table("D:/shuju.txt",sep=",")
a
x <- ts(a,start=1953)
x
plot(x,xlab="年份",ylab="GDP")
par(mfrow=c(2,1))
acf(x);pacf(x)
x.dif<-diff(x)
plot(x.dif)
x1.dif<-diff(x.dif)
plot(x1.dif)
par(mfrow=c(2,1))
acf(x1.dif);pacf(x1.dif)
for(i in 1:2) print(Box.test(x1.dif,type="Ljung-Box",lag=6*i))
auto.arima(x)
x.fit1<-arima(x,order=c(2,2,1))
x.fit1
for(i in 1:2) print(Box.test(x.fit1$residual,lag=6*i))
x.fit1<-arima(x,order=c(2,2,1))
x.fit1

```

```
for(i in 1:2) print(Box.test(x.fit1$residual,lag=6*i))
x.fit2<-arima(x,order=c(3,2,1))
x.fit2
for(i in 1:2) print(Box.test(x.fit2$residual,lag=6*i))
x.fit3<-arima(x,order=c(3,2,2))
x.fit3
for(i in 1:2) print(Box.test(x.fit3$residual,lag=6*i))
x.fore<-forecast(x.fit1,h=3)
x.fore
library(forecast)
plot(x.fore)
```

**期刊投稿者将享受如下服务:**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [sa@hanspub.org](mailto:sa@hanspub.org)