

# Analysis and Prediction of Yunnan CPI Series —Based on SARIMA Model

Zhuoran Li, Xiaoyu Sun

Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan  
Email: 271460302@qq.com

Received: Jun. 7<sup>th</sup>, 2016; accepted: Jun. 27<sup>th</sup>, 2016; published: Jun. 30<sup>th</sup>, 2016

Copyright © 2016 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

---

## Abstract

Taking Yunnan Province as an example, monthly data nearly 20 years were used on CPI forecast modeling. Analysis shows that the CPI data present seasonal cycle of 12. Through the establishment of the seasonal ARIMA models in this article, we predict the CPI of Yunnan Province in 2016 in the second quarter will gradually rise on the basis of the first quarter. And it is able to maintain within the scope of stable growth.

## Keywords

CPI, Seasonal ARIMA, Prediction

---

# 云南省CPI序列的分析与预测

## —基于SARIMA模型

李卓然, 孙晓宇

云南财经大学统计与数学学院, 云南 昆明  
Email: 271460302@qq.com

收稿日期: 2016年6月7日; 录用日期: 2016年6月27日; 发布日期: 2016年6月30日

---

## 摘要

本文以云南省为例, 运用近20年来的月度数据对CPI进行建模预测。分析表明, CPI数据呈现周期为12

的季节性; 文章通过建立季节性ARIMA模型, 预测2016年第二季度云南省CPI将在第一季度的基础上逐渐上升, 且能够保持在稳定增长的范围内。

## 关键词

CPI, 季节性ARIMA, 预测

## 1. 引言

价格指数反映不同时期内一组商品价格的变动情况。而与人们生活息息相关的, 同时也是人们最关注的价格指数为 CPI, 即居民消费价格指数。它反映一组代表性商品和服务项目价格水平变化趋势和变动幅度的统计指标, 以零售量或居民消费量为权数, 反映居民家庭生活消费品和服务价格水平的情况。如今 CPI 在衡量通货膨胀中起着举足轻重的作用, 是滞后性数据, 也是市场经济活动与政府货币政策的一个重要参考指标。虽然不能简单地凭 CPI 的增长或减少对经济形势下定论, 但 CPI 的增速的高低可以表明通货膨胀或紧缩的情况, 足以引起决策部门的重视。更重要的是, 它是能够反映居民真实购买力和生活水平的指标。

多年来 CPI 以及通货膨胀率的持续增长已经引起了诸多学者的重视。王维, 范彦伟(2012)基于混沌神经网络算法对全国的 CPI 进行了分析和预测, 结果表明 CPI 不是完全的随机行为, 它受到诸多非线性因素的支配与控制, CPI 的波动表现出一定的混沌规律[1]。荣文静(2011)对中国的 CPI 序列用 ARIMA 模型进行了拟合, 同时通过傅里叶变换和灰色预测的方法进行分析研究, 结论显示灰色预测具有较高的精度[2]。在对 CPI 变化情况的研究中, 学者更集中研究的是全国的 CPI 变化, 对各省的 CPI 变动的研究关注较少。云南省地处中国西南边陲, 是中国连通东南亚、南亚最便捷的陆上渠道。尤其在近 20 年在中央对外开放方针的指引下云南省对外经济贸易迅速发展, 同时随着第三产业的发展, 旅游业已逐渐成为云南省经济的另一支柱产业, 然而这带来的是省内的 CPI 指数不断上涨。例如 2015 年云南省 CPI 涨幅超过全国平均水平, 在全国排名第 4。在对云南省 CPI 序列的分析中, 考虑到月度 CPI 数据可能呈现出一定的周期性, 因此选用季节性 ARIMA (即 SARIMA)模型进行拟合并预测。

## 2. 数据及模型描述

### 2.1. 数据介绍

本文选取云南省 1995 年 1 月到 2016 年 3 月的 CPI 数据进行分析建模。为使选择模型更加准确, 首先选用前 252 个月的数据(截止到 2015 年 12 月)进行分析建模, 并根据模型对 2016 年 1 月到 3 月的数据进行预测, 计算均方根误差, 选择预测效果最好的模型对未来 3 个月的数据进行预测。绘制数据变动趋势图(图 1)可知, 1995 年至 2000 年这五年间云南省 CPI 有大幅度的下降, 2000 年之后除了有小幅度的上升趋势外还呈现出了一定的周期性, 与猜想一致。接下来将原始数据进行 STL 分解, 得到图 2。

由图 2 可知, 通过 STL 分解, 原始序列存在着周期为 12 的季节性, 并且从趋势来看, 自 2000 年以后有轻微的上升趋势。季节影响对时间序列的研究意义重大。澳大利亚统计局指出: 因为季节影响不但掩盖了时序最基础和最真实的变化, 而且也覆盖了研究者比较感兴趣的确定的非季节特征, 所以观察到的数据必须经过季节调整。因此, 在 CPI 时间序列研究中选择模型时应考虑季节影响。通常进行季节调整的方法有回归法和移动平均法, 但这两种方法的季节分量都是确定的, 因此这里选用 SARIMA 方法[3]。

### 2.2. 模型描述

SARIMA 方法是在 ARIMA 方法上推广而来的。1931 年, ARMA 统计模型是数学家 Walker 在 AR

CPI of Yunnan province 199501-201603

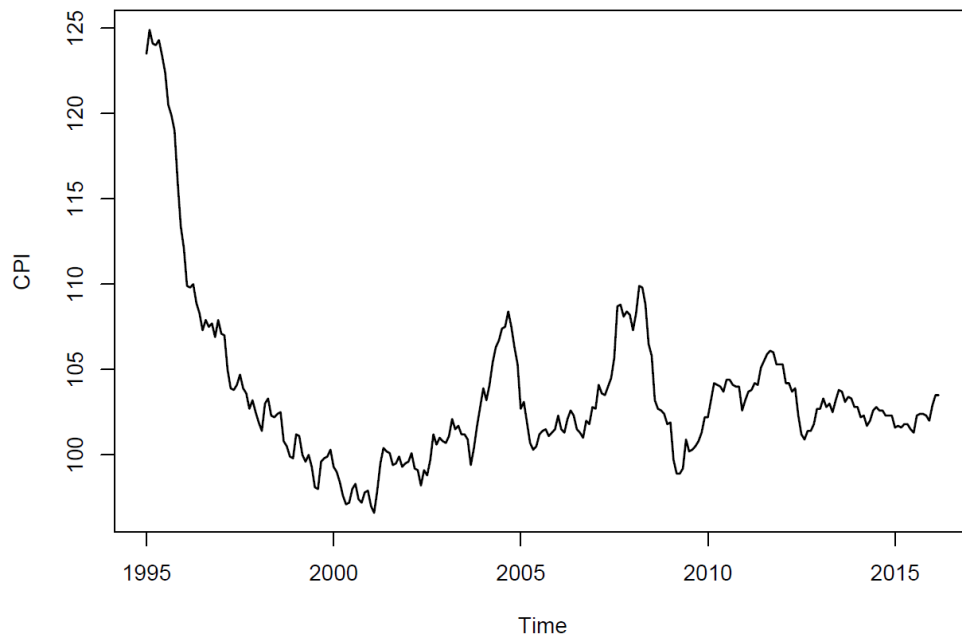


Figure 1. Time series map of Yunnan Province during 1995-Jan to 2016-Mar  
 图 1. 1995 年 1 月~2016 年 3 月云南省 CPI 时序图

Decomposition of additive time series

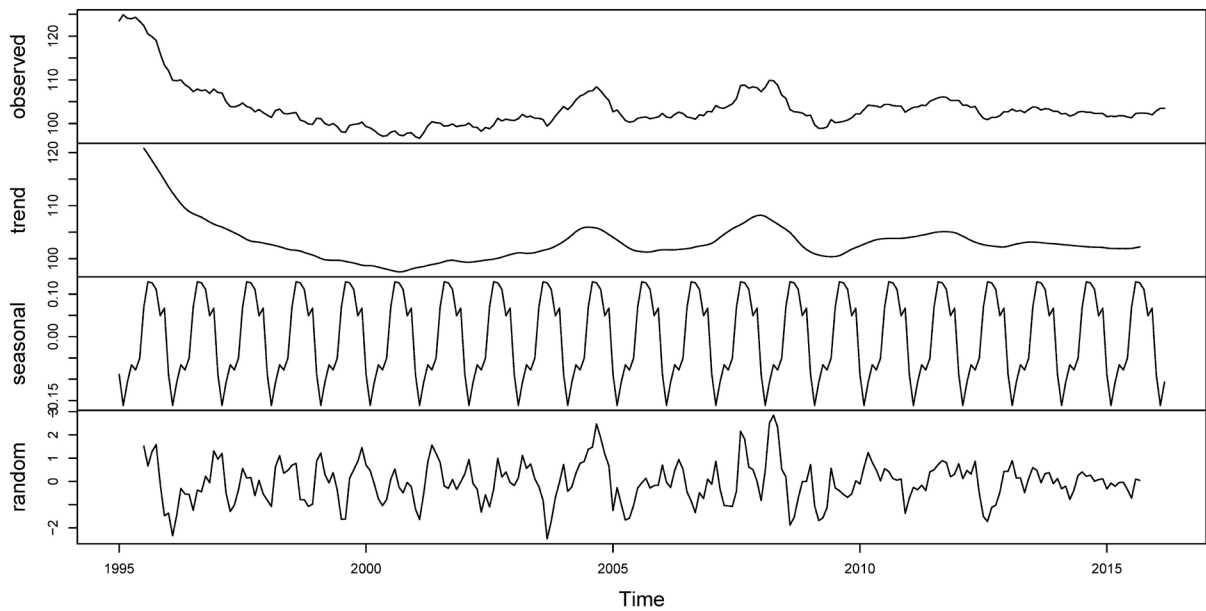


Figure 2. STL decomposition of Yunnan Province time series during 1995-Jan to 2016-Mar  
 图 2. 1995 年 1 月~2016 年 3 月云南省 CPI 时序 STL 分解图

模型的基础之上提出的。假设静态时间序列  $\{x_t, t = 0, \pm 1, \pm 2, \pm 3, \dots\}$  的均值为 0, ARMA 模型的形式如下:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (1)$$

简写为:

$$\phi(L)(x_t - \mu) = \theta(L)\varepsilon_t \quad (2)$$

其中,  $p, q$  是 ARMA 的阶数, ARMA 模型的参数分别是  $\phi_i (i=1, 2, \dots, p)$  和  $\theta_i (i=1, 2, \dots, q)$ 。如果看时间序列  $x_t$  是非平稳的, 则需要通过  $d$  次差分使其变为平稳序列。因此在 ARMA 模型基础上, 科学家 Box 和 Jenkins 提出了 ARIMA( $p, d, q$ )模型[4], 它由自回归过程(AR( $p$ )), 单整( $I(d)$ ), 以及移动平均过程(MA( $q$ ))三个过程组成。

当时间序列呈现出季节性时, 构造的模型中应当加入季节因素, 便可以得到著名的 Box-Jenkins 乘积季节模型, 即季节性的 ARIMA 模型(SARIMA)。季节性 ARIMA 模型由季节性效应模型和 ARIMA 模型两部分构成。序列  $x_t$  具有季节效应时, 应使用季节差分算子消除季节性。季节 ARIMA 模型即 SARIMA( $p, d, q$ )( $P, D, Q$ )<sup>s</sup> 模型的一般形式为:

$$\phi_p(L)\Phi_p(L^s)(1-L)^d(1-L^s)^D x_t = \theta_q(L)\Theta_q(L^s)\alpha_t \quad [5] \quad (3)$$

其中, 季节自回归阶数用  $P$  来表示, 季节平均移动阶数用  $Q$  来表示,  $s$  表示季节性周期, 季节  $P$  阶自回归算子是  $\phi_p(L)\Phi_p(L^s)$ ,  $\Theta_q(L^s)$  是  $Q$  阶移动平均算子,  $\alpha_t$  表示均值为 0, 方差为  $\sigma^2$  的正态分布。

### 3. 实证分析

#### 3.1. 数据平稳性检验

ARIMA 模型的拟合都是针对平稳序列展开的, 而现实生活中很多实际数据都是非平稳的。通常情况下, 非平稳的时间序列可以通过取对数或者差分变换将其转换成平稳序列。一般检验时间序列的平稳性的方法可分为两大类, 一类是通过描绘时间序列的时序图和自相关图, 按照图上所呈现的特点进一步判断分析。另外一类是通过计算某一反映序列平稳性的特征统计量, 再对该序列进行平稳性判断。第一类方法相对于第二类来说简单易操作, 但带有一定的主观性, 因此在进行平稳性分析时通常选择第二类方法。

这里所说的特征统计量平稳性检验即是进行单位根检验, 常用的检验方法有 DF 检验、ADF 检验、PP 检验以及 KPSS 检验四种方法。传统上使用前三种检验方法的较多, 但它们存在的共同的缺陷在于它们的零假设为“有单位根”, 因此, 在数据量不够或者缺乏足够证据时, 往往无法拒绝零假设, 从而顺理成章地得到“有单位根”的结论。实际上, 再拒绝原假设时, 只能得到“没有足够证据说明没有单位根”(没有足够证据证明平稳)的结论, 并非“有证据说不平稳”的结论。而 KPSS 与其恰恰相反, 它是 Kwitkowski 等人提出的。与其他检验不同的是, 它的零假设是平稳或趋势平稳的, 检验显著则意味着可能不平稳, 得到的结论更加准确, 近年来被采用得比较广泛。本文采用 KPSS 检验方法进行单位根检验。通过 R 软件对原序列以及其经过 1、2 次差分的序列进行单位根检验, 检验结果如表 1 所示。

由表 1 可知, 在滞后阶数为 2 时检验统计量的 P 值为 0.1 大于  $\alpha (\alpha = 0.05)$ , 不拒绝原假设, 认为序列平稳, 因此该序列为 2 阶单整。该序列进行一般二阶差分 and 分别进行一般差分和季节性差分的单位根检验结果相同, 但根据 AIC 准则最终确定的 SARIMA 模型中的  $d$  应为 2,  $D$  为 0。

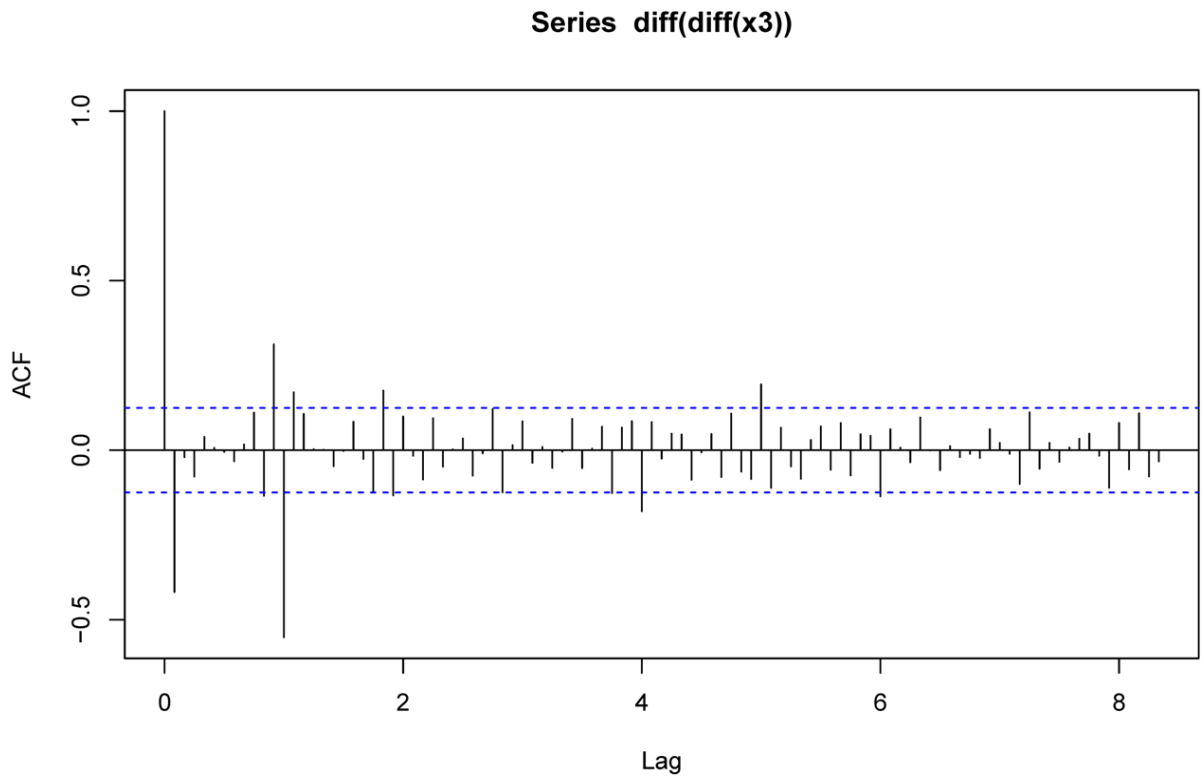
2 阶差分后的序列自相关和偏自相关图为图 3, 图 4。

根据 ACF 和 PACF, 两者都呈拖尾性质, 应使用 ARIMA 模型, 并且可判断参数  $p = 2, q = 3$ , 及 ARIMA(2,2,3)比较合适。实际建模过程中, 还需建立临近阶数的 ARIMA 模型进行比较, 选择最优模型。

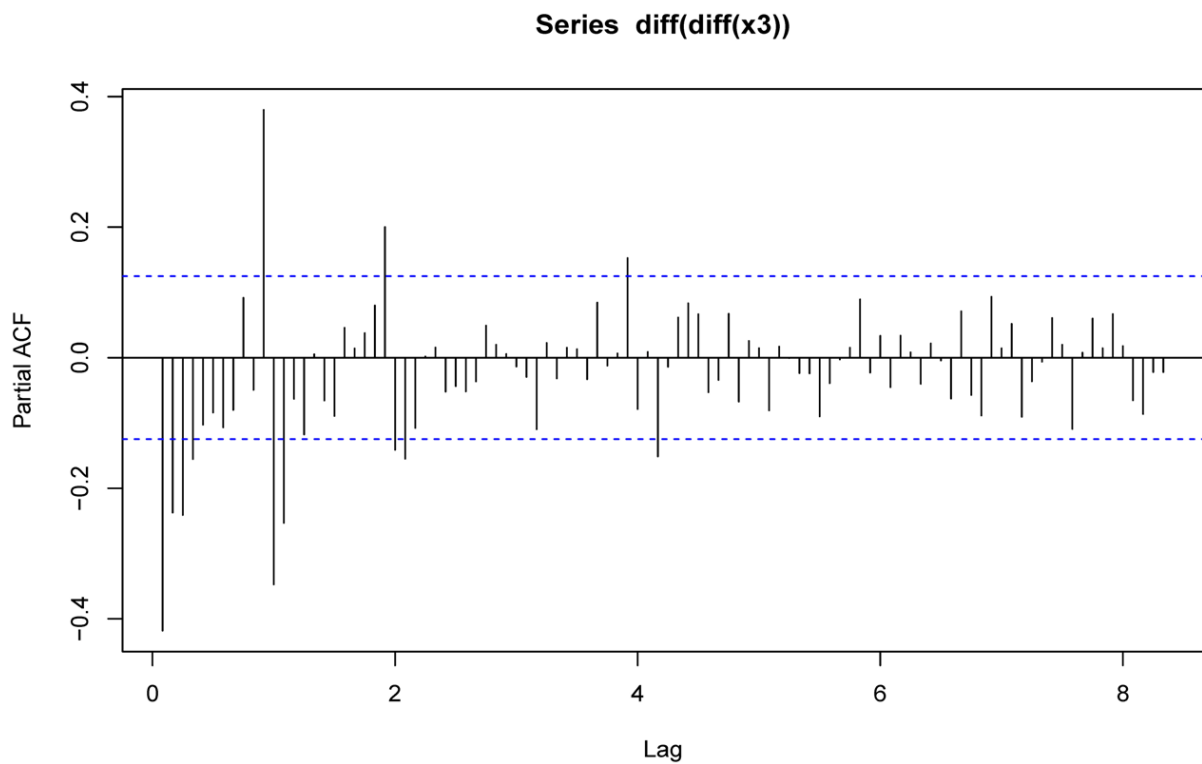
由于季节性模型的  $P, Q$  判断较为困难, 但根据文献, 超过 2 的情况很少, 可以对  $P, Q$  分别取 0,1,2 由低阶到高阶逐步实验, 再根据最小信息准则和模型诊断结果判断和确定最优模型。

#### 3.2. 模型拟合和比较

通过多次拟合, 比较合适的模型参数估计结果如表 2 所示。



**Figure 3.** The ACF after 2 order difference  
**图 3.2** 阶差分后的 ACF



**Figure 4.** The PACF after 2 order difference  
**图 4.2** 阶差分后的 PACF

Table 1. KPSS unit root test results

表 1. KPSS 单位根检验结果

滞后阶数	零假设	统计量	P 值	结论
0	平稳(Level)	0.7601	0.01	不平稳
	趋势平稳(Trend)	0.5803	0.01	不平稳
1	平稳(Level)	0.4891	0.04783	不平稳
	趋势平稳(Trend)	0.1774	0.02273	不平稳
2	平稳(Level)	0.0196	0.1	平稳
	趋势平稳(Trend)	0.0154	0.1	平稳

Table 2. Each model estimation parameters

表 2. 各模型估计参数

模型变量	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8	模型 9
AR(1)	—	—	—	0.1562	0.1607	-0.4136	0.3312	-0.6302	-0.8803
AR(2)	—	—	—	0.0509	0.0521	0.3991	0.0958	—	-0.0057
AR(3)	—	—	—	—	—	—	—	—	0.5321
SAR(1)	-0.4920	-0.4975	-0.5214	-0.5190	-1.0753	-0.2650	-0.5121	-1.0851	-0.5145
SAR(2)	-0.2209	-0.2178	-0.2321	-0.2304	-0.3489	—	-0.2257	-0.3522	-0.2299
SMA(1)	-0.3572	-0.3384	-0.303	-0.3073	0.2698	-0.5474	-0.3138	0.2762	-0.3222
SMA(2)	—	—	—	—	-0.3518	—	—	-0.3556	—
MA(1)	-0.8297	-0.7476	-0.7425	-0.8985	-0.8966	-0.3217	-1.0748	-0.1039	0.1838
MA(2)	—	-0.1108	-0.0718	—	—	-0.7669	0.0841	-0.5452	-0.7305
MA(3)	—	—	-0.0541	—	—	0.2187	0.0624	-0.1220	-0.7093
MA(4)	—	—	—	—	—	—	—	—	0.4291
AIC	487.17	486.08	487.32	487.17	487.04	492.44	488.66	489.03	484.06

由表 2 可知, 模型 2 和模型 9 这两个模型的 AIC 最小。AIC 准则只是一个参考指标, 接下来比较两者的预测的平均误差率。其计算公式为:

$$NMSE = E\left(\frac{|x - xp|}{x}\right) \quad (4)$$

其中  $x$  为 CPI 的真实值,  $xp$  为模型的拟合值。通过计算, 模型 2 的平均误差率为 0.0032, 模型 9 的均方根误差为 0.0039。相比之下, 模型 2 的拟合效果更好且模型形式比较简单, 因此选择模型 2。

### 3.3. 模型诊断

对模型  $ARIMA(0,2,2) \times (2,0,1)^{12}$  进行诊断, 结果如图 5 所示,

图 4 中呈现了该模型拟合残差的广义方差检验, Ljung-Box 检验的 p 值, 相应的 ACF 及残差序列。广义方差检验和 Ljung-Box 检验的 p 值均大于临界值 0.05, 且 ACF 呈现出没有自相关的特性, 并且残差图显示出拟合的残差在 0 值上下波动, 且逐渐趋于稳定, 因此该模型无自相关, 无异方差, 拟合得比较理想。

### 3.4. 预测

在上述模型的基础上, 对云南省 2016 年第一季度的 CPI 值进行预测, 共计 4 个月。预测值及预测图分别如表 3 和图 6 所示。

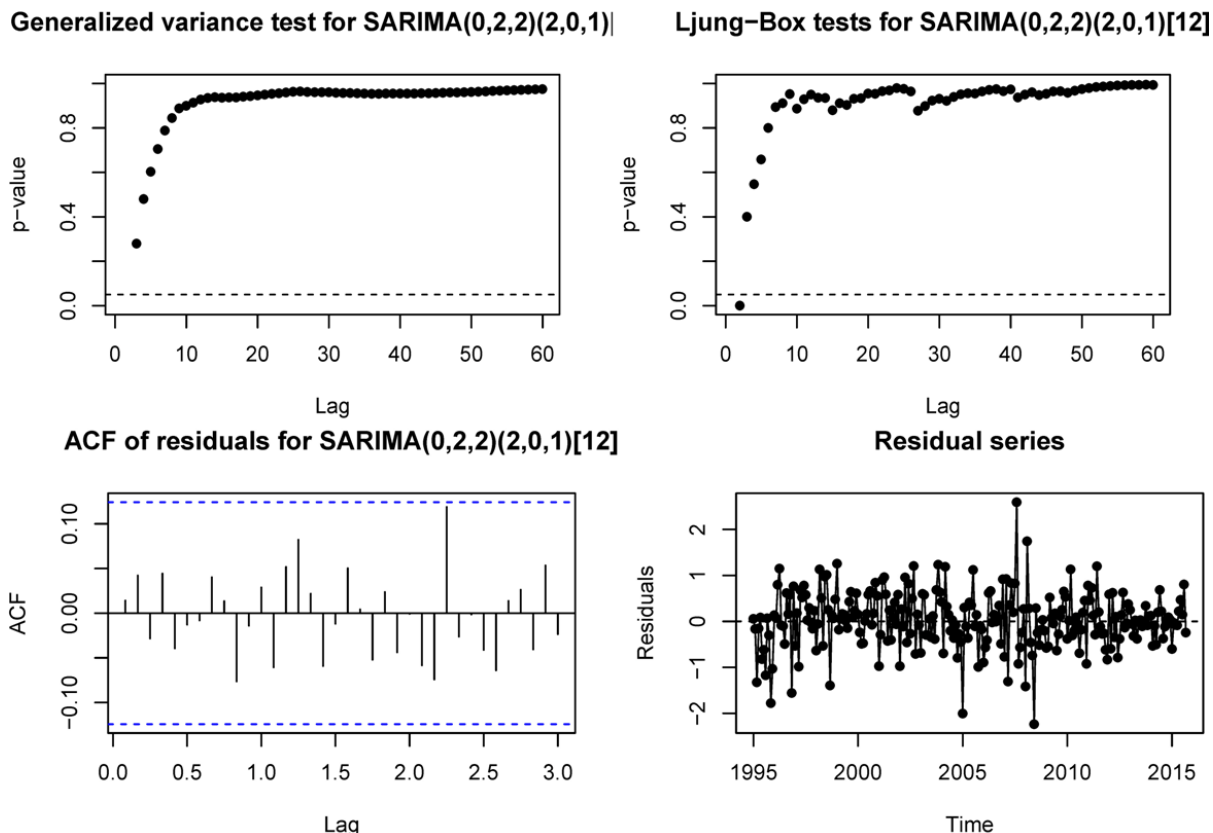


Figure 5. Test results of model  $ARIMA(0,2,2) \times (2,0,1)^{12}$

图 5. 模型  $ARIMA(0,2,2) \times (2,0,1)^{12}$  检验结果

Table 3. Prediction results based on model 2

表 3. 基于模型 2 的预测结果

时间	真实值	点预测值	95%置信下限	95%置信上限	预测误差
15-Jul	101.3	101.15	99.9636	102.3365	0.001481
15-Aug	102.3	101.3888	99.4454	103.3322	0.008907
15-Sep	102.4	101.4861	98.8613	104.1108	0.008925
15-Oct	102.4	101.571	98.3326	104.8095	0.008096
15-Nov	102.3	101.4985	97.5986	105.3985	0.007835
15-Dec	102	101.5355	97.0307	106.0404	0.004554
16-Jan	102.9	102.4947	96.9364	107.2012	0.003939
16-Feb	103.5	102.4998	96.3464	107.882	0.009664
16-Mar	103.5	102.4795	95.7917	108.5536	0.009860
16-Apr	—	102.5434	99.28075	105.806	—
16-May	—	102.4029	98.49685	106.309	—
16-Jun	—	102.1894	97.63472	106.744	—

## Forecasts from ARIMA(0,2,2)(2,0,1)[12]

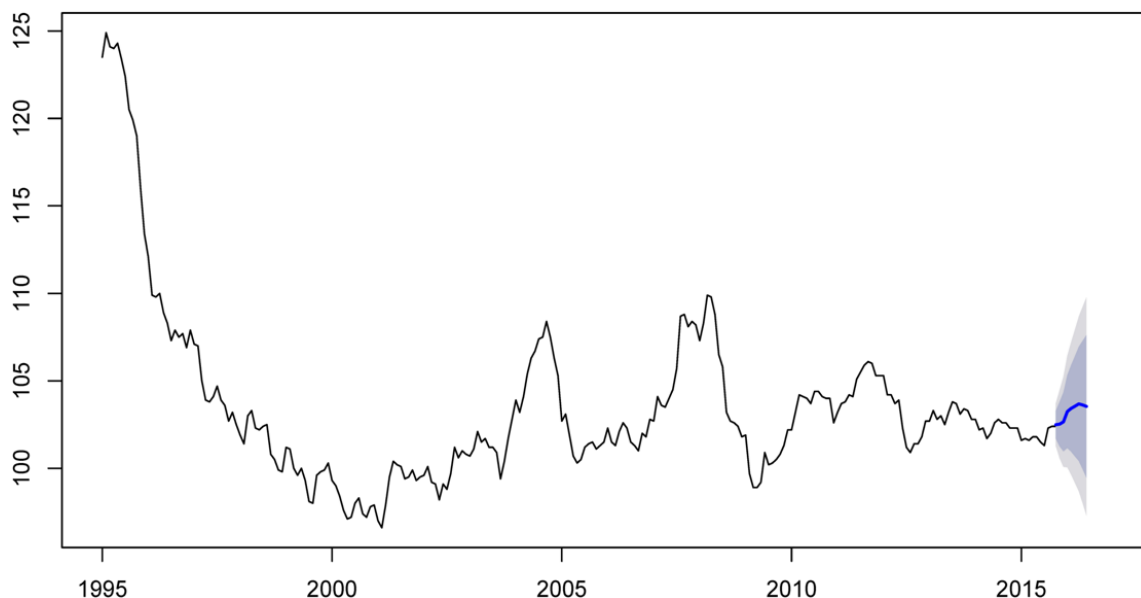


Figure 6. Yunnan province in the first quarter of 2016 CPI forecast

图 6. 云南省 2016 年第一季度 CPI 预测图

由于实际社会中影响 CPI 的因素复杂多变, 点预测值要低于真实值, 根据表 3 中的计算数据, 模型对 CPI 的预测误差均不超过 1%, 控制在可接受的范围内, 因此该模型比较可信。并且云南省 2016 年第二季度 CPI 相对第一季度而言会有所下降, 物价趋于稳定。

#### 4. 结论

由于 CPI 数据呈现一定的季节性, 因此如果不考虑季节因素的影响而使用单一的 ARIMA 模型会使拟合精度以及预测效果大打折扣。AIC 准则虽然在一定程度上可以判断模型拟合优度, 但进行预测时通过计算均方误差选择模型更能提高预测的准确程度。另外, 本文通过建立季节性 ARIMA 模型, 将 CPI 的季节性因素考虑在内, 提高了模型的拟合优度以及预测的可信度。

一般来说, CPI > 3% 的增速称为通货膨胀, CPI > 5% 增速称为严重通货膨胀。一般情况下, 为维持经济的稳定增长, 保持 CPI 在 3% 以内的增速是合理地。根据历年 CPI 变动趋势以及预测的 2016 年二季度 CPI 值, 除旅游高峰期带来的 CPI 会暂时高于 3% 外, 其余月份的 CPI 增速均维持在 3% 以内, 因此云南省的 CPI 呈现出比较平稳的趋势。因此, 除了有关部门需要对市场经济情况准确把握, 制定相应的货币政策等宏观调控外, 旅游旺季, 物价部门应做好相关工作, 避免哄抬物价等不良现象, 从而保证居民生活的稳定。

#### 参考文献 (References)

- [1] 王维, 范彦伟. 基于混沌神经网络算法的 CPI 预测[J]. 统计与决策, 2012(10): 22-24.
- [2] 荣文静. CPI 指数序列的分析及预测[D]: [硕士学位论文]. 成都: 成都理工大学, 2011.
- [3] 张明芳, 等. 居民消费价格指数季节调整实证研究[J]. 财经研究, 2004, 30(3): 133-144.
- [4] 耿娟娟. 基于 S-12-ARIMA 和 SARIMA 模型及其组合模型的 CPI 预测研究[D]: [硕士学位论文]. 成都: 西南石油大学, 2015.
- [5] 雷鹏飞. SARIMA 模型在预测中国 CPI 中的应用[J]. 统计与决策, 2011(5): 28-30.



**再次投稿您将享受以下服务：**

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>