

# 基于ARIMA模型的云南省短期CPI预测

何宇婷

云南财经大学, 云南 昆明

收稿日期: 2021年12月24日; 录用日期: 2022年1月14日; 发布日期: 2022年1月27日

## 摘要

居民消费价格指数(CPI)是反映居民生活中消费品和服务价格变动情况的宏观经济指标。本文选取云南省2010年1月至2021年7月的月度CPI数据建立  $ARIMA(1,1,0)(2,0,0)_{12}$  模型, 对云南省未来12个月的CPI进行预测, 并提出相关建议, 为云南省政府有关部门制定政策提供支持。

## 关键词

ARIMA模型, 消费价格指数, 云南省, R

# The Short-Term CPI Forecast of Yunnan Province Based on ARIMA Model

Yuting He

Yunnan University of Finance and Economics, Kunming Yunnan

Received: Dec. 24<sup>th</sup>, 2021; accepted: Jan. 14<sup>th</sup>, 2022; published: Jan. 27<sup>th</sup>, 2022

## Abstract

The Consumer Price Index (CPI) is a macroeconomic indicator that reflects the changes in the prices of consumer goods and services in the lives of residents. This paper selects the monthly CPI data of Yunnan Province from January 2010 to July 2021 to establish  $ARIMA(1,1,0)(2,0,0)_{12}$  model, predicts the CPI of Yunnan Province in the next 12 months, and puts forward relevant suggestions to provide support for relevant departments of the Yunnan Provincial Government to formulate policies.

## Keywords

ARIMA Model, Consumer Price Index, Yunnan Province, R

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

## 1. 引言

2020年新冠疫情蔓延使世界经济遭受重创，中国经济也面临严峻考验，物价水平大幅下跌，然而猪瘟盛行让猪肉价格猛涨，带动国内消费价格指数持续上涨。目前，世界经济慢慢复苏，大宗商品价格持续高涨，在如此复杂的经济环境中，CPI作为通货膨胀的重要指标，研究CPI并对其进行预测很有必要。

改革开放之后，云南省的经济得到快速发展，根据云南统计局公布的数据，近十年云南省月度CPI都保持稳定，在2020年2月CPI达到近年小高峰106.3，但在疫情之后，云南省的月度CPI波动下跌。如今云南省经济已经慢慢走向复苏，本文通过历史数据对云南省消费价格指数建立ARIMA模型，以预测未来一段时间云南省的CPI状况，为云南省政府制定宏观经济政策提供相关参考和依据。

## 2. 文献综述

在对CPI的预测研究文献中，学者一般采用ARIMA模型、马尔科夫链模型、灰色模型对CPI进行预测。根据研究的方法不同，可以通过以下几种模型对CPI展开预测研究。

### 2.1. ARIMA 模型

ARIMA模型的基本思路是首先对原始数据进行平稳性和纯随机性检验，若序列不平稳先差分，后检验差分序列是否为白噪声，若为平稳非白噪声序列则拟合模型，然后对模型进行检验并展开预测。杨颖梅(2015)以北京市1998年1月至2013年5月的CPI月度数据为训练集，建立自回归单整移动平均模型ARIMA(12, 1, 8)，成功预测未来北京市7期的CPI月度数据[1]。孙晓丹(2021)同样建立ARIMA(13, 0, 0)模型对我国CPI月度数据进行分析并预测2021年我国通货膨胀情况，研究结果表明CPI指数具有较长的滞后期，当经济复苏后，CPI指数持续回升使得通货膨胀水平稳定增长[2]。

### 2.2. 马尔科夫链模型

部分学者采用马尔科夫链模型对CPI进行预测。关李娜(2021)将河北省CPI数据依据CPI与通货膨胀的关系划分为四个区间，分别为通货紧缩、正常、通货膨胀、严重通货膨胀，并以此建立马尔科夫链模型预测河北省2020和2021年的通货膨胀都为正常水平[3]。赵雪妍(2021)利用马尔科夫模型对陕西省1998~2019年CPI数据进行分析，建立的预测模型通过了Markov检验并表明预测结果的误差控制在合理范围内，实验结果显示陕西省2021年CPI增速将保持稳定，可能在[0%, 1%)之间[4]。

### 2.3. 灰色预测模型

张徐荣和袁之焕(2018)采用灰色预测模型(1, 1)对我国2011~2015年的CPI变动趋势进行预测，结果显示灰色预测相对误差小，模拟精度高[5]。张洋洋等(2020)基于5个月的数据将改进的GM(1, 1)模型用

于山东省的 CPI 预测中, 预测精度为 I 级[6]。

## 2.4. 文献评述

综上所述, 灰色预测模型操作简单, 不需要很多的样本数据就能得到精度较高的预测模型, 特别是对具有短期相关性的数据效果更好。但是马尔科夫链模型的应用需要满足以下三个条件, 系统状态转移概率矩阵不变; 预测期内系统状态总数不变; 状态转移只与时间间隔有关, 只有通过检验的数据才能应用马尔科夫链模型。大部分模型基于年度历史数据对 CPI 进行预测, 本文采用云南省月度 CPI 历史数据, 且 ARIMA 模型结构简单, 并不需要其他外生变量就能对 CPI 做出较好的拟合。因此, 本文采用季节性 ARIMA 模型对云南省的月度消费价格指数建立模型并作出预测。

## 3. 理论模型

### 3.1. SARIMA 模型介绍

在实践中, 有些时间序列除了表现出普通的非平稳性以外, 在周期上也表现出季节性非平稳性, 可以对非季节性部分建立普通的 ARMA( $p, d, q$ ), 而对季节性部分建立 ARMA( $P, D, Q$ ), 更为一般的模型如下:

$$\phi(L)\phi_s(L^s)(L^s)(1-L)^d(1-L^s)^D X_t = \Theta(L)\Theta_s(L^s)\varepsilon_t$$

这就是所谓的 ARIMA( $p, d, q$ )( $P, D, Q$ )<sub>s</sub> 模型, 其中有

$$\phi_s(L^s) = 1 - \varphi_1 L^s - \varphi_2 L^{2s} - \dots - \varphi_p L^{ps}$$

$$\Theta_s(L^s) = 1 + \vartheta_1 L^s + \vartheta_2 L^{2s} + \dots + \vartheta_Q L^{Qs}$$

分别对应季节性的 AR( $P$ )和 MA( $Q$ )部分。

$$\phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p$$

$$\Theta(L) = 1 + \vartheta_1 L + \vartheta_2 L^2 + \dots + \vartheta_Q L^Q$$

分别对应非季节性的 ar( $p$ )和 ma( $q$ )。

### 3.2. 平稳性检验

对序列的平稳性检验一般有两种方法, 一是根据图像直接作出判断, 但是该方法容易主观臆测; 二是通过构造统计量进行假设检验, 常用 ADF 检验。ADF 检验原理如下:

对任一 AR( $p$ )过程, 有

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t$$

等价变换为:

$$x_t - x_{t-1} = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t$$

整理上式为:

$$\nabla x_t = (\phi_1 + \dots + \phi_p - 1)x_{t-1} - (\phi_2 + \dots + \phi_p)\nabla x_{t-1} - \dots - \phi_{p-1}\nabla x_{t-p+1} + \varepsilon_t$$

简记为:

$$\nabla x_t = \rho x_{t-1} + \beta_1 \nabla x_{t-1} + \cdots + \beta_{p-1} \nabla x_{t-p+1} + \varepsilon_t$$

其中:

$$\rho = \phi_1 + \phi_2 + \cdots + \phi_p - 1$$

$$\beta_j = -\phi_{j+1} - \phi_{j+2} - \cdots - \phi_p, j = 1, 2, \cdots, p-1$$

若序列  $\{x_t\}$  平稳, 则

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1$$

等价于  $\rho < 0$ 。

若序列非平稳, 则至少存在一个单位根, 有

$$\phi_1 + \phi_2 + \cdots + \phi_p = 1$$

等价于  $\rho = 0$ 。在 AR(p)过程单位根检验的假设条件可以确定为:

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

构造 ADF 检验量:

$$\tau = \hat{\rho} / s(\hat{\rho})$$

其中  $s(\hat{\rho})$  为参数  $\rho$  的样本标准差。

### 3.3. 纯随机性检验

模型的显著性检验一般通过对序列残差进行白噪声检验[7]。

原假设:

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_p = 0, \forall m > 1$$

备择假设:

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

LB 统计量由 Ljung 和 Box 证明出其近似服从自由度为  $m$  的卡方分布, 数学表达式为:

$$LB = n(n+2) \sum_{k=0}^m (\hat{\rho}_k^2 / n - k) \sim X^2(m)$$

其中,  $m$  为延迟期数,  $n$  为观测期数。若 LB 统计量小于显著性水平, 认为原序列为非白噪声序列, 数据具有可提取的信息, 可以继续拟合模型。

## 4. 数据来源

本文采用数据来自《云南统计年鉴》, 选取云南省 2010 年 1 月至 2021 年 7 月的月度居民消费价格指数为样本数据, 利用 ARIMA 模型进行 CPI 预测。

根据图 1 可知, 近十年里云南省月度 CPI 的顶峰是 2020 年 2 月 106.34, 2014 年至 2018 年初云南省月度居民消费价格指数波动下跌, 2018 年小幅上涨, 后经历新冠疫情、猪肉价格上涨、经济慢慢复苏等复杂环境, 让云南省的月度居民消费价格指数急剧上涨后断崖式下跌。

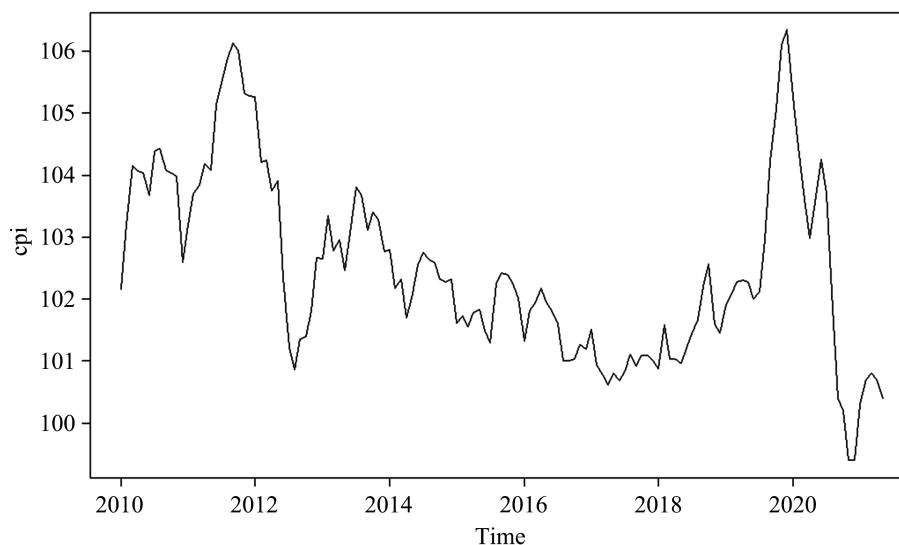


Figure 1. Raw data timing diagram

图 1. 原始数据时序图

## 5. 建模分析

### 5.1. 数据平稳性和纯随机性检验

首先对原始时间序列进行 ADF 检验，检验结果见表 1。在 5% 显著性水平下，无漂移项无趋势项模型、有漂移项无趋势项模型、有漂移项有趋势项模型均没有通过 ADF 检验，因此认为原始时间序列是不平稳的。

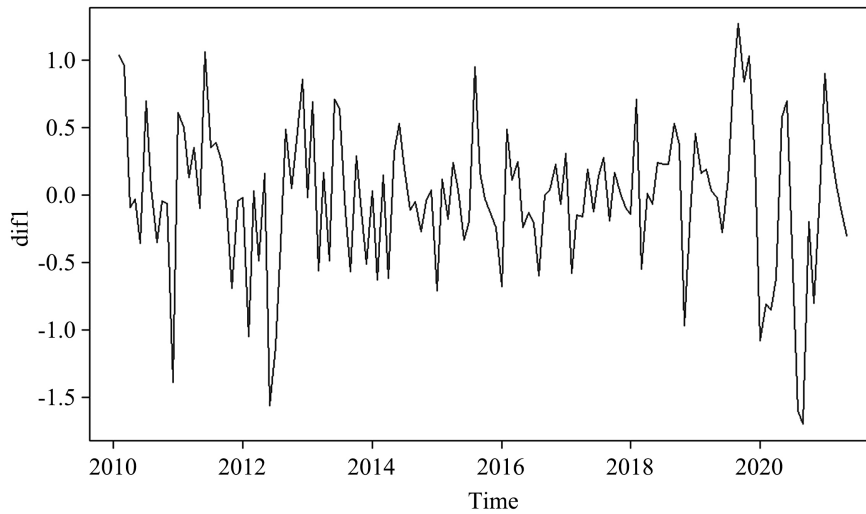
Table 1. ADF inspection of original sequence

表 1. 原始序列 ADF 检验

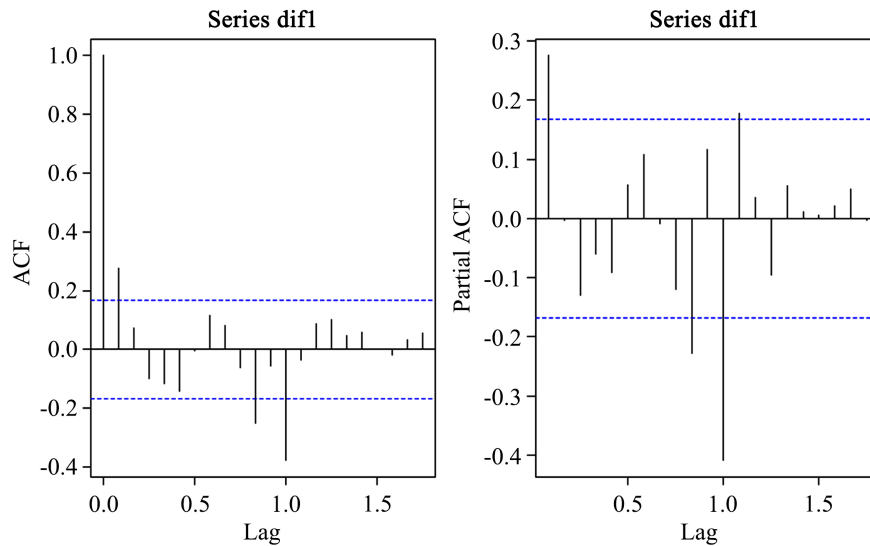
Lag	无漂移项无趋势项		有漂移项无趋势项		有漂移项有趋势项	
	ADF	P.value	ADF	P.value	ADF	P.value
0	-0.306	0.556	-1.90	0.3663	-2.52	0.3574
1	-0.437	0.518	-2.71	0.0801	-3.22	0.0874
2	-0.558	0.479	-2.83	0.0606	-3.26	0.0800
3	-0.509	0.497	-2.46	0.1488	-2.95	0.1811
4	-0.537	0.487	-2.34	0.1961	-2.82	0.2327

将原始 CPI 月度数据进行一阶差分，差分之后时序图见图 2。

一阶差分后的自相关图和偏自相关图见图 3，都呈现出 1、12 阶显著，且图像近似三角函数，表明一阶差分后的序列具有季节效应。故对 1 阶差分后的序列再进行 12 步差分，然后对 1 阶 12 步差分的序列进行单位根检验，单位根检验结果如表 2 所示。在 5% 显著性水平下，三种模型的 P 值均小于或等于 0.01，表明 1 阶 12 步差分后的云南省月度 CPI 数据已经实现了平稳。



**Figure 2.** Timing diagram after 1st order difference  
**图 2.** 一阶差分后时序图



**Figure 3.** ACF and PACF diagrams after first-order difference  
**图 3.** 一阶差分后的 ACF、PACF 图

**Table 2.** ADF test after CPI's first-order 12-step difference  
**表 2.** CPI 的 1 阶 12 步差分后的 ADF 检验

Lag	无漂移项无趋势项		有漂移项无趋势项		有漂移项有趋势项	
	ADF	P.value	ADF	P.value	ADF	P.value
0	-8.78	0.01	-8.76	0.01	-8.73	0.01
1	-6.31	0.01	-6.29	0.01	-6.28	0.01
2	-6.25	0.01	-6.24	0.01	-6.23	0.01
3	-5.87	0.01	-5.86	0.01	-5.86	0.01
4	-5.87	0.01	-5.88	0.01	-5.87	0.01

数据具有平稳性质之后再通过“Ljung-Box”进行纯随机性检验，以判断平稳时间序列是否为白噪声。检验结果见表 3 所示。

**Table 3.** Pure randomness test

**表 3.** 纯随机性检验

指标名称	延迟 6 阶	延迟 12 阶
X-squared	15.953	61.157
p-value	0.01401	1.388e-08

根据表 3 的纯随机性检验结果，延迟 6 阶和 12 阶的 LB 统计量 P 值均未超过显著性水平 0.05，表明 1 阶 12 步差分后的云南省月度 CPI 数据是平稳非白噪声序列，可以用 ARIMA 建立模型。

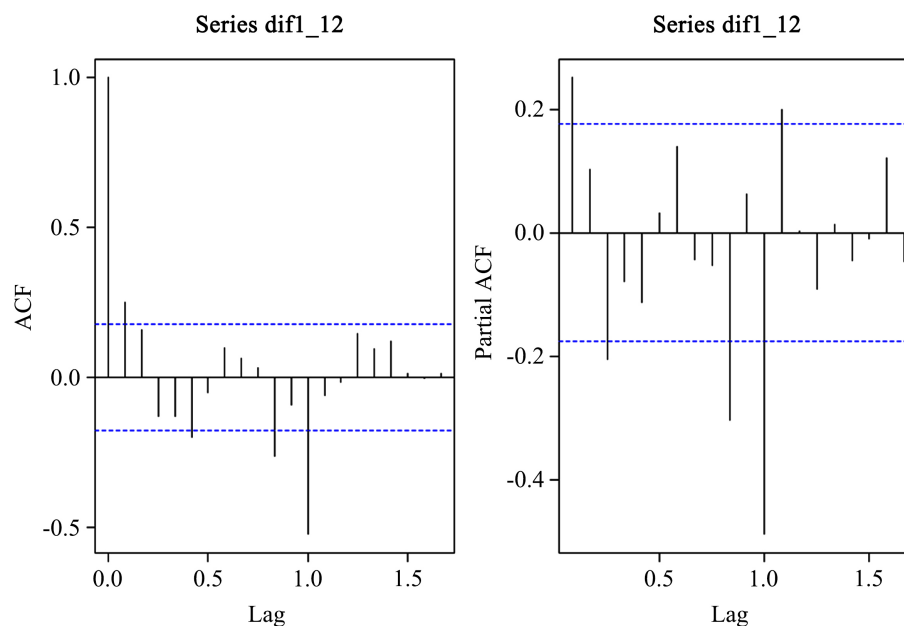
## 5.2. 模型建立

### 5.2.1. auto.arima 函数定阶

采用 auto.arima 函数对模型进行定阶，R 软件自动拟合的模型为  $ARIMA(1,1,0)(0,0,2)_{12}$ ，拟合结果如表 4 所示，3 个系数中只有两个系数大于其两倍标准差，1 个系数不显著。

### 5.2.2. 观测法定阶

1 阶 12 步差分后的自相关图及偏自相关图见图 4 所示。



**Figure 4.** ACF and PACF diagrams after 1st-order 12-step difference

**图 4.** 1 阶 12 步差分后的 ACF、PACF 图

根据 1 阶差分及 1 阶 12 步差分的自相关图及偏自相关图，拟合模型  $ARIMA(1,1,0)(2,0,0)_{12}$ 、 $ARIMA(1,1,0)(1,0,0)_{12}$ ，模拟结果见表 4 所示。

**Table 4.** Model fitting results  
**表 4.** 模型拟合结果

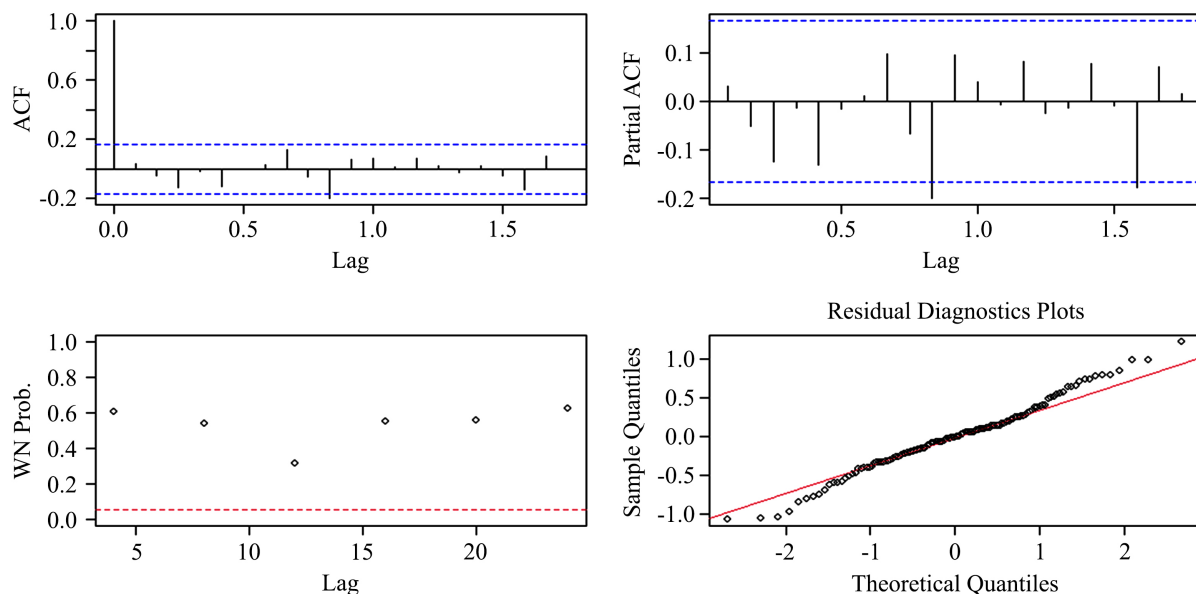
模型	模拟结果			
	参数	Ar1	Sma1	Sma2
ARIMA(1,1,0)(0,0,2) <sub>12</sub>	估计值	0.3241	-0.6996	-0.1775
	S.E.	0.0833	0.0956	0.1064
	AIC 值		171.87	
	参数	Ar1	Sar1	Sar2
ARIMA(1,1,0)(2,0,0) <sub>12</sub>	估计值	0.3249	-0.6983	-0.3363
	S.E.	0.0846	0.0931	0.0976
	AIC 值		170.12	
	参数	Ar1	Sar1	Sar2
ARIMA(1,1,0)(1,0,0) <sub>12</sub>	估计值	0.3527	-0.5345	0
	S.E.	0.0829	0.0809	0
	AIC 值		178.94	

ARIMA(1,1,0)(2,0,0)<sub>12</sub> 模型的系数均大于两倍标准差，表明系数显著，且根据 AIC 准则，拟选择 ARIMA(1,1,0)(2,0,0)<sub>12</sub> 模型，模型表达式为：

$$(1 - 0.3294B)(1 + 0.6983B^{12} + 0.3363B^{24})(1 - B)X_t = \varepsilon_t$$

**5.2.3. 模型检验**

对 ARIMA(1,1,0)(2,0,0)<sub>12</sub> 模型的残差进行检验，见图 5 所示，残差的白噪声检验 P 值全都位于显著性水平 5% 之上，表明模型是显著的。



**Figure 5.** Model significance test plot  
**图 5.** 模型显著性检验图



再通过 Ljung-Box 对模型显著性进行检验, 检验结果见表 5 所示。

**Table 5.** Model significance LB test

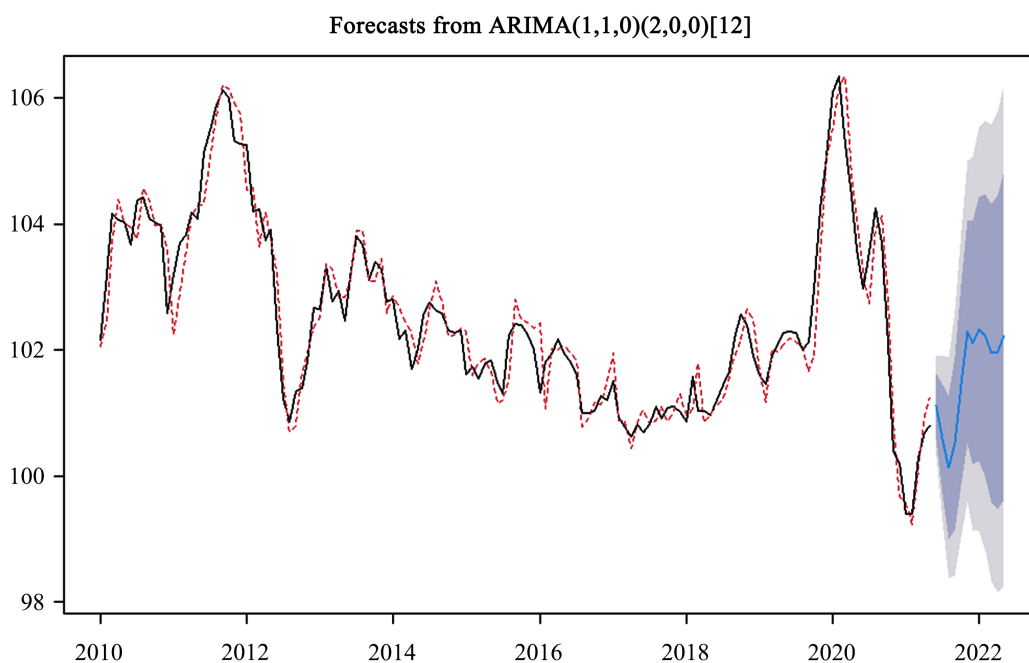
**表 5.** 模型显著性 LB 检验

滞后阶数	1	2	3	4	5	6
P 值	0.7123	0.7888	0.427	0.5881	0.4435	0.5725

滞后 6 阶的 P 值均大于显著性水平 0.05, 接受残差为白噪声的原假设, 说明  $ARIMA(1,1,0)(2,0,0)_{12}$  模型显著。

#### 5.2.4. 模型预测

本文以 2010 年 1 月至 2021 年 7 月的云南省月度 CPI 指数为实验集, 建立  $ARIMA(1,1,0)(2,0,0)_{12}$  模型, 预测云南省未来 12 期的月度居民消费价格指数。预测效果图见图 6。



**Figure 6.** Model fitting and prediction renderings

**图 6.** 模型拟合及预测的效果图

图 6 是模型预测效果图, 虚线表示拟合值, 实线表示观测值。阴影部分的实线为预测值, 深色、浅色阴影分别为序列 80% 和 95% 置信区间。预测效果图显示, 模型拟合值与真实值比较接近, 拟合效果好。由于 ARIMA 模型的预测方差会随着预测步长的增加而逐渐增大, 所以 ARIMA 模型通常用于短期预测。

表 6 是模型的 12 期预测值, 将 2021 年 8 月至 12 月的真实值与预测值进行对比发现, 相对误差随着时间的推移逐渐增加, 前 5 期的平均误差为 4.45%, 表  $ARIMA(1,1,0)(2,0,0)_{12}$  模型在预测云南省短期月度 CPI 时效果好。

**Table 6.** Model prediction results and real value comparison results table  
**表 6.** 模型预测结果与真实值对比结果表

时间	预测值	Lo95	Hi95	真实值	相对误差	平均误差
2021.08	104.1978	103.2910	105.1047	99.8	4.22%	
2021.09	104.9632	103.4581	106.4684	99.7	5.01%	
2021.10	104.8430	102.8563	106.8298	100	4.62%	4.45%
2021.11	105.0343	102.6446	107.4240	100.8	4.03%	
2021.12	104.9584	102.2197	107.6971	100.4	4.34%	
2022.01	104.7272	101.6777	107.7767	—	—	—
2022.02	104.7323	101.4005	108.0641	—	—	—
2022.03	104.9454	101.3532	108.5376	—	—	—
2022.04	105.2132	101.3783	109.0481	—	—	—
2022.05	105.2186	101.1554	109.2818	—	—	—
2022.06	103.8452	99.5659	108.1246	—	—	—
2022.07	103.3521	98.8671	107.8372	—	—	—

## 6. 结论与建议

从预测结果来看,  $ARIMA(1,1,0)(2,0,0)_{12}$  模型在短期内对云南省月度 CPI 的预测较为准确, 可根据此模型对云南省未来的短期 CPI 走向作出预测。根据预测结果, 未来一年时间内云南省的月度 CPI 将稳中有升, 表明云南省的经济正从疫情中慢慢恢复过来, 为避免通货膨胀降低居民生活质量损害居民利益, 政府应当提前制定相应调控政策以保障物价水平正常波动, 使云南省的经济发展保持稳中向好趋势。

## 参考文献

- [1] 杨颖梅. 基于 ARIMA 模型的北京居民消费价格指数预测[J]. 统计与决策, 2015(4): 76-78.
- [2] 孙晓丹. 基于 ARIMA 模型对我国 CPI 指数的分析预测[J]. 商展经济, 2021(15): 15-17.
- [3] 关李娜. 基于马氏链的河北省居民消费价格指数分析及预测[J]. 现代商业, 2021(27): 107-109.
- [4] 赵雪妍. 陕西省居民消费价格指数预测[J]. 现代商业, 2021(29): 96-98.
- [5] 张徐荣, 袁之焕. 用灰色模型方法预测我国居民消费价格指数变动趋势[J]. 纳税, 2018(8): 175.
- [6] 张洋洋, 邓伟, 宋长钰. 基于时间序列分析的山东省 CPI 预测[J]. 山东师范大学学报(自然科学版), 2020, 35(3): 329-336.
- [7] 王燕. 时间序列分析——基于 R [M]. 北京: 中国人民大学出版社, 2015.