

基于ARIMA模型的云南省GDP的预测和分析

王 婷

云南财经大学统计与数学学院, 云南 昆明

收稿日期: 2022年12月28日; 录用日期: 2023年1月24日; 发布日期: 2023年1月31日

摘 要

本文选取云南省1978~2020年GDP数据, 进行处理与分析后建立了时间序列分析中的自回归滑动平均求和模型ARIMA(p,d,q), 通过了数据平稳性检验和白噪声检验, 确立了ARIMA(0,2,3)模型, 该模型具有简单实用、预测精度高的特点, 能恰当描述云南省GDP的状况。首先通过该模型预测了云南省2018年和2019年的GDP数据, 与并实际数据进行比较, 预测的平均相对误差较小, 说明该模型预测短时间内的值是比较准确的, 模型拟合良好。最后运用拟合的模型对云南省GDP进行短期预测, 发现2020~2024年云南省GDP呈稳步上升的趋势, 为云南省经济的发展提供参考。

关键词

GDP预测, 时间序列, ARIMA模型

Prediction and Analysis of Yunnan GDP Based on ARIMA Model

Ting Wang

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan

Received: Dec. 28th, 2022; accepted: Jan. 24th, 2023; published: Jan. 31st, 2023

Abstract

This paper selects the GDP data of Yunnan Province from 1978 to 2020. After processing and analyzing, it establishes ARIMA(p,d,q), an autoregressive moving average summation model in time series analysis. Through data stationarity test and white noise test, ARIMA(0,2,3) model is established. This model is simple and practical, with high prediction accuracy, and can properly de-

scribe the GDP situation of Yunnan Province. First, the GDP data of Yunnan Province in 2018 and 2019 are predicted by this model. Compared with the actual data, the average relative error of the prediction is small, which indicates that the model is more accurate in predicting the value in a short period of time and the model fits well. Finally, we use the fitting model to make a short-term forecast of the GDP of Yunnan Province. It is found that the GDP of Yunnan Province will show a steady upward trend from 2020 to 2024, providing a reference for the economic development of Yunnan Province.

Keywords

GDP Forecast, Time Series, ARIMA Model

Copyright © 2023 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

国内生产总值(GDP)指一个国家或者地区在一定时期内运用生产要素所生产的全部最终产品和服务的市场价值,并且该指标是国民经济核算过程中重要的参数指标之一,主要能够综合反映出一个地区或者国家在经济发展方面的实际情况,同时也为国家和地区制定宏观政策、部署战略方针等等提供了一定参考和依据[1]。GDP和通货膨胀率、经济增长率及失业率都是最基础的主要的宏观经济指标。若能对未来GDP做出准确的预测,不仅能为宏观经济健康发展起良好的导向性作用,还有利于政府做出更好的决策。通过它可以判断经济是处于衰退还是处于膨胀阶段,政府是需要采取刺激还是需要采取控制等相关措施。预测准确与否极大地影响政府决策结果的科学性和有效性。因此,如何运用科学有效的方法来对其进行预测具有重要的现实意义[2]。但是在复杂多变的宏观经济环境中,各级决策机构如何进行做出正确经济决策与进行合理规划一直是存在的难题。

目前预测国内生产总值(GDP)的方法有很多。对于短期GDP预测,时间序列分析方法是较为常用的预测方法[3]。其中,单整自回归移动平均(ARIMA)模型比较简单灵活,且可行性较高,目前是应用最广泛的时间序列预测模型之一[4]。见诸文献的代表性成果有:郝香芝[5]利用ARMA技术构建了我国时间序列模型。龚国勇和王莎莎等[6]利用ARIMA模型分别对深圳GDP和中国GDP进行了预测。华鹏等[7]通过ARIMA模型预测广东近五年GDP的增长,为政府制定宏观经济规划起着参考意见,表明ARIMA模型适合用于GDP的预测增长。

除常用的ARIMA方法外,许多学者也通过从其他角度和方法对国民生产总值预测中的预测进行了论证分析。张海燕[8]对比了ARIMA模型和HP过滤模型分析GDP的增长,但发现HP过滤模型中的滤波是一种特殊固定的滤波,因此可能产生伪趋势,甚至产生序列中原本并不存在的周期,所以并不适合用于对GDP的预测。彭乃弛等[9]通过ARIMA-GM-BP来对GDP进行预测对比,发现GM模型预测误差偏大,但ARMA模型与ARIMA-GM-BP模型预测误差都较小,适合做关于GDP的预测。曾波等[10]利用灰色组合预测模型(GM模型)对中国2008年GDP进行了模糊预测,但模糊预测分析往往具有不确定性和片面性。

综上所述,本文选择时间序列分析方法中的ARIMA模型,该模型具有简单实用、预测精度高的特点,能恰当描述云南省GDP的状况。选取云南省1978年~2020年GDP数据,建立模型进行分析预测,

具有较高的模型可信度。

2. ARIMA 模型介绍

时间序列分析方法主要是用来解决具有季节性，周期性及平稳性的时间序列问题的。求和自回归移动平均模型是最基本的模型之以，简记为 ARIMA(p,d,q)模型：

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

式中， $\nabla^d = (1-B)^d$ ； $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ，为平稳可逆 ARMA(p,q)模型的自回归系数多项式； $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ，为平稳可逆 ARMA(p,q)模型的移动平均系数多项式。

求和自回归移动平均模型这个名字的由来是：d 阶差分后序列可以表示为：

$$\nabla^d x_t = \sum_{i=0}^d (-1)^i C_d^i x_{t-i}$$

式中， $C_d^i = \frac{d!}{i!(d-i)!}$ ，即差分后序列等于原序列的若干序列值的加权和，对差分平稳序列又可以拟合自回归移动平均(ARMA)模型，所以称它为求和自回归移动平均模型。

3. ARIMA 模型对云南省 GDP 的预测分析

3.1. 数据来源

本文选取云南省 1978 年~2020 年的 GDP 最新数据，来源于 2020 年云南省统计年鉴，数据真实有效，无缺失值。

3.2. 数据平稳性检验及处理

依据云南省 1978 年至 2020 年的 GDP 数据，得到时间序列图，如图 1 所示。

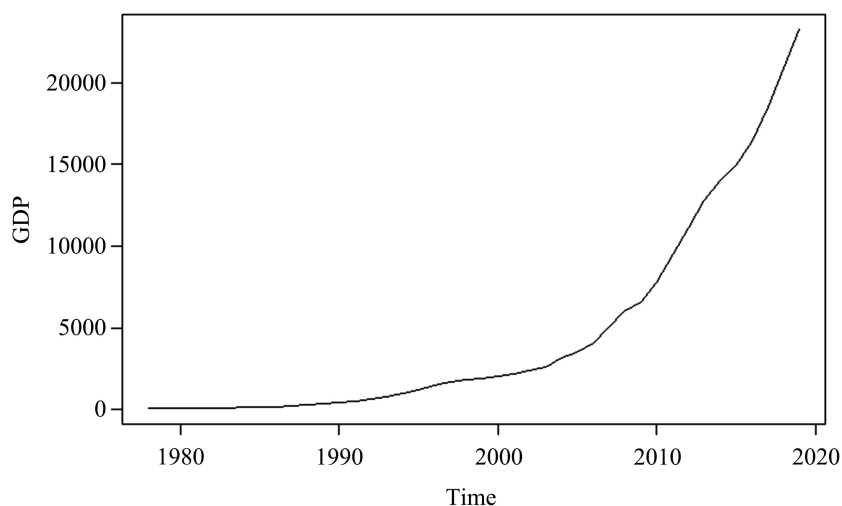


Figure 1. Time series of Yunnan's GDP from 1978 to 2020

图 1. 1978~2020 年云南省 GDP 的时间序列图

从图 1 可以看出, 时序图呈明显的上升趋势, 但是没有周期性和季节性的波动。显而易见, 该时间序列是非平稳的。接着, 对这个非平稳的序列进行一阶差分, 得到一次差分后的结果如图 2 所示, 图中仍然存在明显增长趋势, 于是继续对一阶差分后的序列进行二阶差分, 得到二次差分后的折线图, 如图 3 所示, 从图 3 可以看出不再有持续上升或下降的趋势, 而是围绕 0 值上下波动。对二阶差分后的序列进行单位根检验, 得到 $p = 0.01 < \alpha = 0.05$, 说明该时间序列达到平稳。

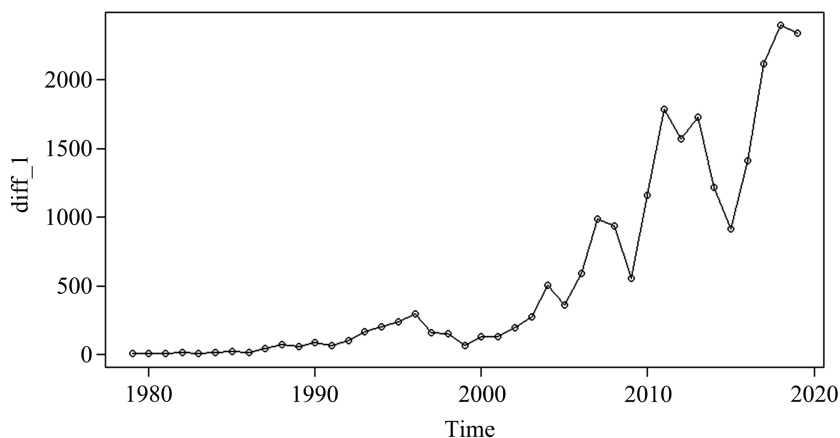


Figure 2. First order differential GDP line chart

图 2. 一阶差分 GDP 折线图

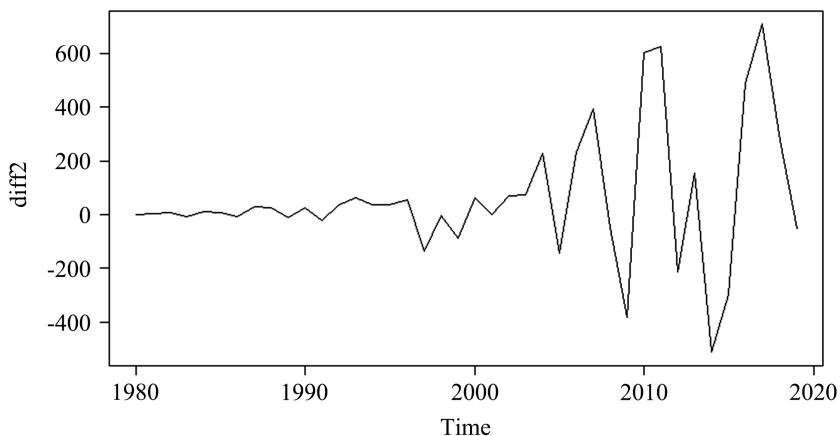


Figure 3. Second order differential GDP line chart

图 3. 二阶差分 GDP 折线图

3.3. 拟合 ARIMA 模型

由上述结果可知, 二阶差分后的云南省 GDP 序列是平稳的。接着, 通过 R 软件得到二次差分后序列的自相关图和偏自相关图, 如图 4, 图 5 所示。

分析二阶差分后的自相关及二阶差分后的偏自相关图可知, 自相关系数在 3 阶以后逐渐趋于零, 偏自相关系数显示拖尾, 因此选取 $p = 0, q = 3$ 。但是若只拟合 ARIMA(0,2,3) 这一个模型, 存在一定主观性, 于是为了消除相应的误差, 同时建立多个 ARIMA 模型, 分别选择: ARIMA(0,2,1), ARIMA(0,2,2), ARIMA(0,2,3), ARIMA(1,2,3), 并且依次得到每个模型的 AIC 值, 见表 1, 从表中数据可以得出, ARIMA(0,2,3) 模型的 AIC 值最小, 因此该模型相对是最优的。

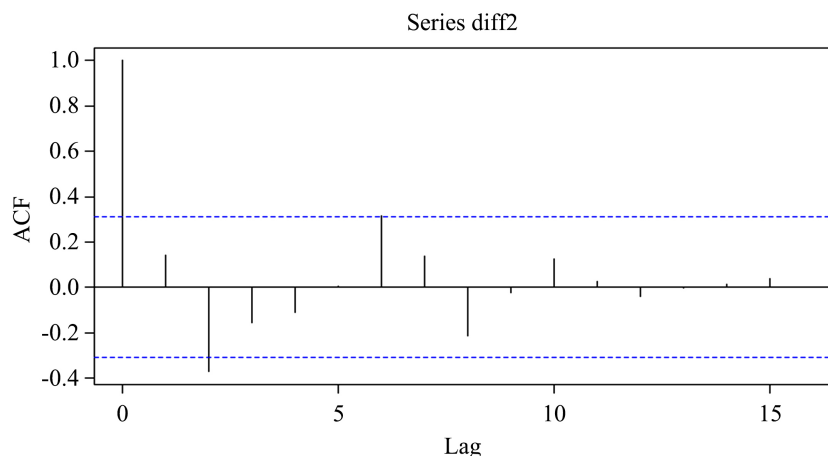


Figure 4. Autocorrelation after second-order difference

图 4. 二阶差分后的自相关图

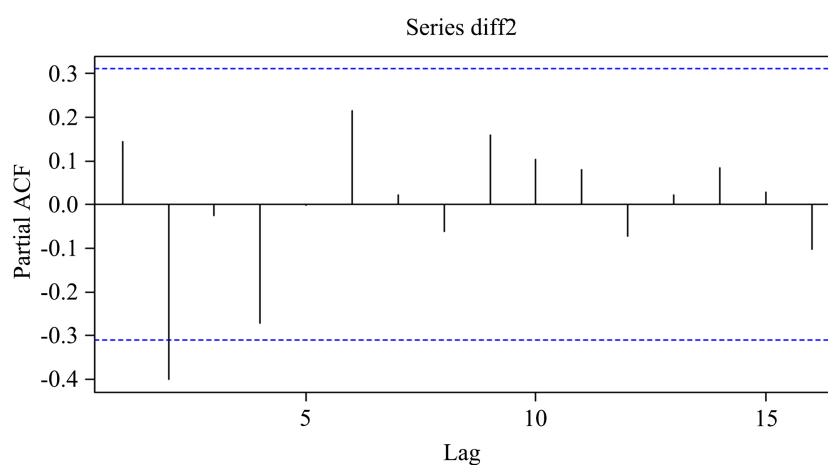


Figure 5. Partial autocorrelation diagram after second-order difference

图 5. 二阶差分后的偏自相关图

Table 1. Comparison of ARIMA models

表 1. ARIMA 模型比较

模型	AIC 值
ARIMA(0,2,1)	554.95
ARIMA(0,2,2)	542.29
ARIMA(0,2,3)	541.06
ARIMA(1,2,3)	541.33

3.4. 模型检验

根据模型识别得到最佳的模型是 ARIMA(0,2,3)，接下来对该模型的残差进行白噪声检验，检验结果如图 6 所示。关于残差序列的白噪声检验结果(图 6 的左下图)，可以看出各阶延迟下白噪声的检验统计量的 p 值显著大于 0.05 (虚线为 0.05 参考线)。我们可以认为这个拟合模型的残差序列属于白噪声序列，即拟合模型显著成立。

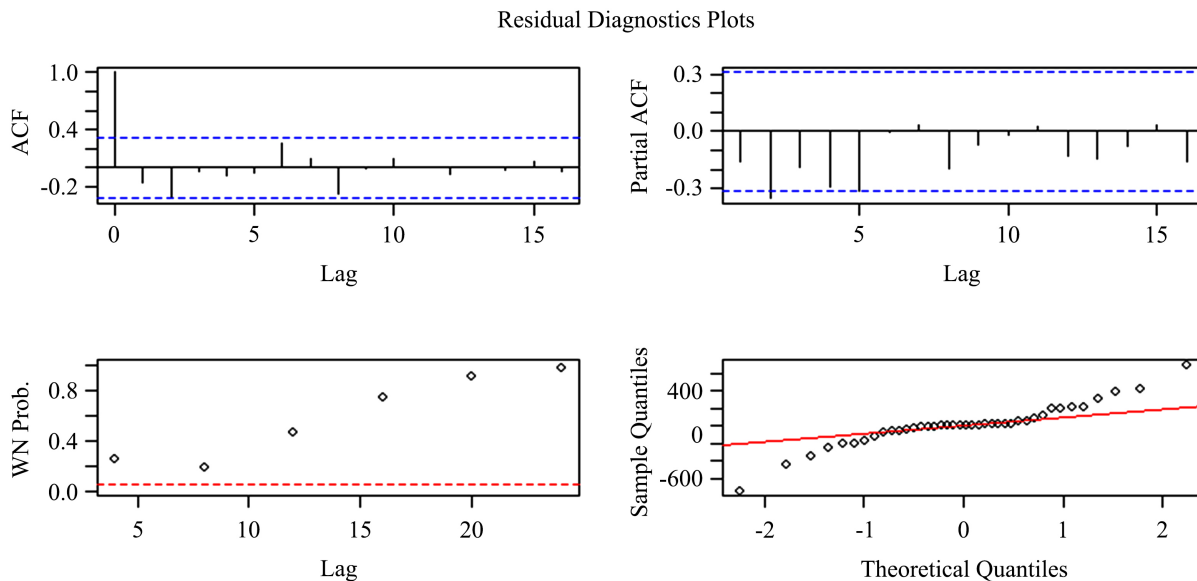


Figure 6. Significance test of GDP series fitting model in Yunnan province

图 6. 云南省 GDP 序列拟合模型显著性检验图

3.5. 模型的预测

由以上分析结果，通过拟合的 ARIMA(0,2,3)模型对云南省 GDP 进行预测，预测 2018 年和 2019 年云南省国内生产总值的数据，结果如表 2 所示。

Table 2. Comparison between predicted value and actual value

表 2. 预测值与实际值的比较

	预测值	实际值	相对误差
2018	20602.90	20880.63	1.32%
2019	22719.88	23223.75	2.16%

从表 2 中的结果可以明显看出，预测的平均相对误差较小，说明该模型预测短时间内的值是比较准确的，预测效果较好。

接着，运用该模型预测了云南省 2020 年至 2024 年 5 年的 GDP 数据，预测结果如表 3 所示，预测图如图 7 所示。

Table 3. GDP forecast of Yunnan province from 2020 to 2024

表 3. 云南省 2020~2024 年 GDP 的预测值

年份	预测值
2020	25466.73
2021	27662.59
2022	29850.03
2023	32037.48
2024	34224.92

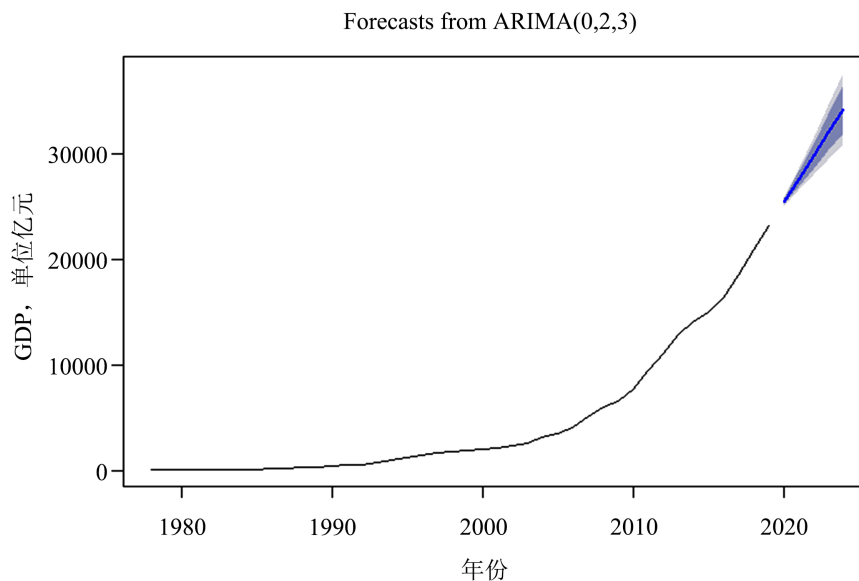


Figure 7. GDP forecast of Yunnan province
图 7. 云南省 GDP 的预测图

预测的结果呈稳步上升的趋势，符合现实背景，但是由于 2020 年新冠肺炎疫情的影响，2020 年的 GDP 肯定会受到影响，所以预测出来的数据只能作为参考作用。

4. 结论与建议

在建模分析预测的过程中，首先建模数据需要满足平稳性的条件，若不满足，则需要进行处理使其通过检验，比如处理的方法有差分和取对数，然后再依据画出的 ACF 图和 PACF 图，确定出模型为 ARIMA(0,2,3)，接着模型要通过参数的显著性检验和残差的白噪声检验，增加模型的可信度，在实际应用过程中，ARIMA 模型预测短期的效果很好，但预测时间的延长会导致模型的误差变大。

比较 2018 年、2019 年的预测值和实际值的过程中会发现还是存在一定误差。特别是 2019 年云南省 GDP 净增量达 3000 多亿，增幅较大，查询相关数据可以发现，这一增幅主要来源于第三产业的大幅增加，这与云南省的旅游业离不开关系。这也为云南的经济发展提供一定建议。

一是预测的 GDP 值除了作一定参考之外，应关注 GDP 的构成，除了继续发展第三产业之外，还应加大第二产业尤其是工业的发展。

二是关注国际贸易。云南省作为西南边陲之地，应注重与东南亚的贸易往来，进一步发展经济，提升经济地位。

参考文献

- [1] 许宪春. 国内生产总值核算的重要意义和作用[J]. 中国统计, 2003, 34(2): 8-9.
- [2] 李连友, 张曦. 对 2006-2010 中国经济预测准确性的评析[J]. 统计研究, 2012, 29(12): 4.
- [3] 李莹. 时间序列分析在山东省 GDP 预测中的应用研究[D]: [硕士学位论文]. 济南: 山东大学, 2011: 1-34.
- [4] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2008.
- [5] 郝香芝, 李少颖. 我国 GDP 时间序列的模型建立与预测[J]. 统计与决策, 2007(23): 4-6.
- [6] 龚国勇. ARIMA 模型在深圳 GDP 预测中的应用[J]. 数学的实践与认识, 2008, 38(4): 54-57.
- [7] 华鹏, 赵学民. ARIMA 模型在广东省 GDP 预测中的应用[J]. 统计与决策, 2010(12): 166-167.

- [8] 张海燕. 我国宏观经济变量周期性波动的动态模型与计量分析[D]: [博士学位论文]. 长春: 吉林大学, 2006.
- [9] 彭乃弛, 党婷. 基于 ARMA-GM-BP 组合预测模型及应用[J]. 统计与决策, 2016(2): 80-82.
- [10] 曾波, 刘思峰, 方志耕, 谢乃明. 灰色组合预测模型及其应用[J]. 中国管理科学, 2009(5): 151.