

基于ARIMA模型的陕西省农林牧渔业增加值分析预测

李劲彬^{1,2,3,4,5}, 花东文^{1,2,3,4,5}

¹陕西地建土地工程技术研究院有限责任公司, 陕西 西安

²陕西省土地工程建设集团有限责任公司, 陕西 西安

³自然资源部退化及未利用土地整治工程重点实验室, 陕西 西安

⁴陕西省土地整治工程技术研究中心, 陕西 西安

⁵自然资源部土地工程技术创新中心, 陕西 西安

收稿日期: 2022年9月25日; 录用日期: 2022年10月24日; 发布日期: 2022年10月31日

摘要

本文基于国家统计局官网关于陕西省2002年~2021年的农林牧渔业增加值数据, 通过时间序列分析建立ARIMA模型进行拟合分析, 利用建立的模型对2021年的预测值和实际值进行比较, 相对误差为2.46%, 小于5%, 表明模型拟合良好, 最后利用ARIMA (0, 2, 1)模型对陕西省农林牧渔业增加值进行预测, 结果表明未来五年该产业将持续稳步增长, 且平均增长率为6.89%。

关键词

农林牧渔业, 陕西省, 时间序列分析, ARIMA模型, 分析预测

Analysis and Prediction of Added Value of Agriculture, Forestry, Animal Husbandry and Fishery in Shaanxi Province Based on ARIMA Model

Jinbin Li^{1,2,3,4,5}, Dongwen Hua^{1,2,3,4,5}

¹Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an Shaanxi

²Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an Shaanxi

³Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources, Xi'an Shaanxi

⁴Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an Shaanxi

⁵Land Engineering Technology Innovation Center, Ministry of Natural Resources, Xi'an Shaanxi

Received: Sep. 25th, 2022; accepted: Oct. 24th, 2022; published: Oct. 31st, 2022

Abstract

Based on the added value data of agriculture, forestry, animal husbandry and fishery in Shaanxi Province from 2002 to 2021 on the official website of the National Bureau of Statistics, the ARIMA model is established through time series analysis for fitting analysis. The model is used to compare the predicted value and the actual value in 2021. The relative error is 2.46%, less than 5%, indicating that the model is well fitted. Finally, the ARIMA (0, 2, 1) model is used to predict the added value of agriculture, forestry, animal husbandry and fishery in Shaanxi Province. The results show that the industry will continue to grow steadily in the next five years, with an average growth rate of 6.89%.

Keywords

Agriculture, Forestry, Animal Husbandry and Fishery, Shaanxi Province, Time Series Analysis, ARIMA Model, Analysis and Forecast

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

国内生产总值(GDP)是指一定时期内, 一个国家或者地区所有常住单位生产活动的最终成果, 其是衡量国家或地区经济发展状况和发展水平的重要指标。农林牧渔业增加值[1]是指农、林、牧、渔及农林牧渔专业及辅助性活动生产货物或提供服务活动而增加的价值, 为该产业现价总产值扣除现价中间投入后的余额, 它能够反映一定时期内农林牧渔业的生产总规模以及总成果, 在国内生产总值中占有重要地位[2], 是计算国内生产总值的基础, 而且其变化对地方经济发展调控具有重要的参考价值。“十四五”时期是我国全面建设社会主义现代化强国的第一个五年, 意义非凡。《陕西省国民经济和社会发展第十四个五年规划和二零三五年远景目标纲要》提出到 2035 年陕西省经济综合实力大幅提升, 人均生产总值较 2020 年翻一番。党的十八大以来, 陕西省持续坚持农业农村优先发展战略, 扎实开展脱贫攻坚、乡村振兴工作, 稳步推进现代农业建设、乡村环境治理等工程, 农业农村经济发展稳中有进, 粮食产能稳步提升, 全省目前处于创新驱动和投资拉动并重阶段, 合理预测农林牧渔业增加值对于陕西省进一步采取措施促进农业现代化发展、提升农民收入、改善人居环境至关重要。

杨智凯[3]等选取广西统计年鉴中 1995~2019 年 GDP 数据, 通过 SAS 统计软件分析拟合, 建立了 ARIMA (1, 2, 1)模型, 并且通过对比预测值和真实值, 发现模型拟合效果良好。张梓[4]以贵州省 1978~2020 年 GDP 数据为基础, 通过 Python 处理、检验, 建立最优 ARIMA (0, 1, 1)模型, 并且对未来 5 年贵州省 GDP 进行预测, 结果显示模型相对误差在 7%以内, 预测效果较好。李志超[5]等选取上海市月度居民消费价格指数数据, 通过分析分别建立 ARIMA (3, 1, 7)模型, GM (1, 1)模型以及一元 23 阶多项式回归模型

对 CPI 指数进行短期预测,结果显示 ARIMA (3, 1, 7)模型和灰色预测模型 GM (1, 1)预测精度相近且较高。ARIMA 模型应用较为广泛,但针对陕西省的相关经济指标研究缺较少,本文将通过应用 ARIMA 模型,分析预测未来 5 年陕西省农林牧渔业的变化,为陕西省经济发展以及宏观调控提供一定的参考意义。

2. ARIMA 模型

ARIMA 模型,全称为 Autoregressive Integrated Moving Average Model,即差分自回归移动平均模型,由 Box 和 Jenkins 于上世纪七十年代提出,是一种著名的时间序列预测方法。ARIMA (p, d, q)模型由三部分组成,分别是自回归过程 AR (p),单整 I (d)以及移动平均过程 MA (q),其中 p 表示自回归阶数, d 为将原始非平稳时间序列转化为平稳时间序列所进行的差分次数, q 为移动平均阶数[6]。其原理是首先将原始非平稳时间序列通过差分转化为平稳时间序列,然后将因变量仅对它滞后值(阶数)以及随机误差项的现值和滞后值进行回归,从而完成模型建立,再对模型进行识别和检验,通过后模型便可用于相关指标预测。模型表达式如下:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (\phi_p \neq 0, \theta_q \neq 0) \quad (1)$$

ARMA 模型主要针对平稳时间序列,因此针对一组原始时间序列数据首先需要进行平稳性检验,通常有两种方法,一是做出原始数据时间序列图直接观察判断,二是通过数学方法计算得到。前者由于需要人为去观察判断,对判断者的经验要求较高,因而平稳性检验通常采用数学方法计算完成,主要有 ADF 检验、KPSS 检验等。通过检验如果原始时间序列不平稳,那么需要对其进行差分,然后再进行平稳性检验,如此反复,直至数据平稳。处理后的数据通过平稳性检验后,需要确定模型 ARIMA 中的参数 p 、 d 、 q ,其中 d 已知,自回归阶数 p 和移动平均阶数 q 通常通过自相关检验 ACF 图和偏自相关检验 PACF 图进行判断。

3. 实例分析

3.1. 数据来源

为较好研究陕西省未来五年农林牧渔业增加值发展趋势,同时确保数据的真实性和准确性,本文选取数据来源于国家统计局官网(<https://data.stats.gov.cn/easyquery.htm?cn=E0103>),以陕西省 1992 年至 2021 年生产总值数据(如表 1)为基础,通过应用时间序列分析方法建立 ARIMA 模型,分析预测其变化趋势。

Table 1. The value-added data of agriculture, forestry, animal husbandry and fishery in Shaanxi Province from 1992 to 2021 (unit: 100 million yuan)

表 1. 陕西省 1992~2021 年农林牧渔业增加值数据(单位: 亿元)

年份	1992	1993	1994	1995	1996	1997	1998	1999
农林牧渔业增加值	116.7	148.2	171.8	217.3	250.6	255.4	266.9	254.6
年份	2000	2001	2002	2003	2004	2005	2006	2007
农林牧渔业增加值	258.2	263.6	282.2	302.7	372.3	435.8	484.8	592.6
年份	2008	2009	2010	2011	2012	2013	2014	2015
农林牧渔业增加值	753.7	790.8	989.5	1237.5	1372.5	1528.6	1637.8	1675.6
年份	2016	2017	2018	2019	2020	2021	/	/
农林牧渔业增加值	1778.9	1830.6	1927.8	2098.2	2381.8	2532.2	/	/

数据来源: 国家统计局官网(1992~2021 年)。

3.2. 平稳性检验及处理

通过绘制陕西省 1992~2021 年农林牧渔业增加值及其差分时间序列图(图 1), 可以判定该时间序列及其一阶差分序列是非平稳的, 而二阶差分相对平稳, 为了进一步证实直观判断, 我们选择对原始时间序列及其不同阶数的差分序列同时进行 ADF 检验和 KPSS 检验。对于 ADF 检验, 原始时间序列的结果返回值均为 0, 表示拒绝原假设, 序列不平稳, 相应的 KPSS 检验的返回值为 1, 同样表明序列不平稳。其一阶差分序列 ADF 检验显示不平稳而 KPSS 检验显示平稳, 其二阶差分序列的 ADF 检验返回值为 1, KPSS 检验的返回值 0, 均表明序列平稳。

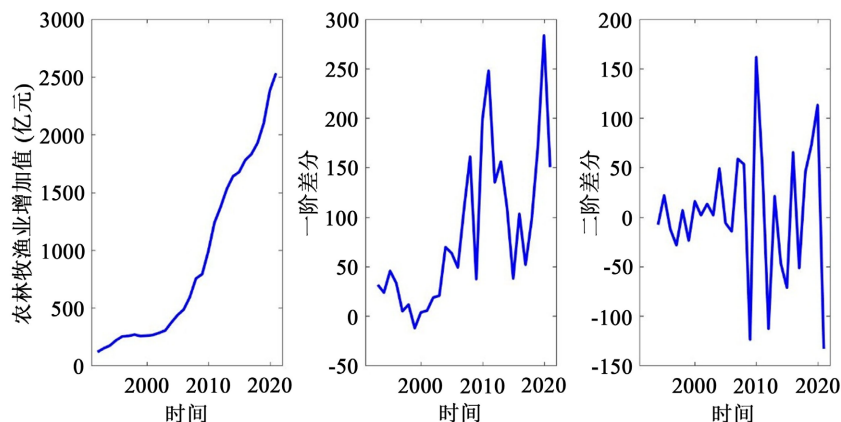


Figure 1. The added value of agriculture, forestry, animal husbandry and fishery in Shaanxi Province and its differential time series

图 1. 陕西省农林牧渔业增加值及其差分时间序列图

3.3. ARIMA (p, d, q)模型的建立与检验

原始时间序列经过二阶差分后的序列通过了平稳性检验, 则阶数 $d = 2$, 因此模型 ARIMA (p, d, q) 剩余仅需确定 p, q 值。通过绘制陕西省农林牧渔业增加值二阶差分数据的自相关检验 ACF 和偏自相关检验 PACF 图(图 2), 发现自相关函数从 0 阶后衰减趋于零, 而偏自相关函数在 1 阶以及 15 阶之后衰减趋于零。为进一步确定模型阶数, 应用 AIC 信息准则, 对多个备选模型进行建模对比。

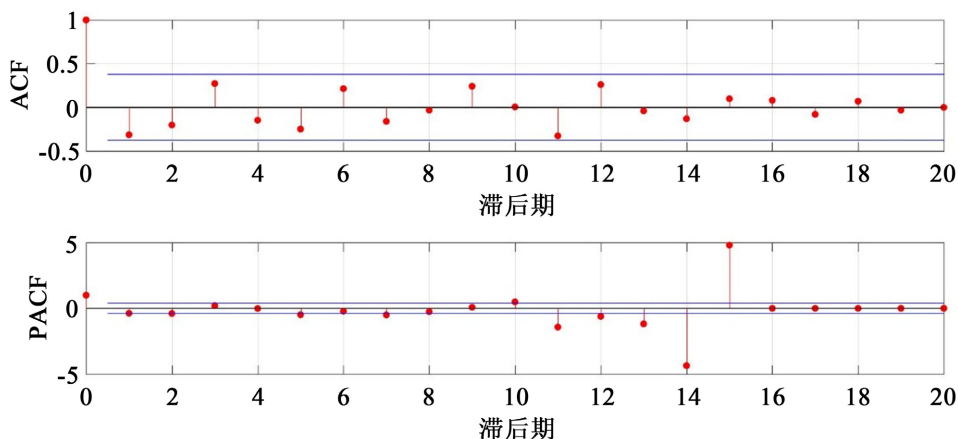


Figure 2. Autocorrelation and partial autocorrelation plots

图 2. 自相关和偏自相关图

表 2 为 ARIMA 模型参数表, 主要包含三方面内容, 分别是模型参数、Q 统计量、信息准则。首先 ARIMA 模型在应用前原始时间序列的二阶差分序列需要通过白噪声检验, 其检验采用 Q 统计量进行, 通常采用 Q_6 来分析, 如果其对应的 p 值大于 0.1, 则表明通过白噪声检验。最后对不同的备选模型进行比较, 选择 AIC 和 BIC 值较小的模型, 最终确定最优模型为: ARIMA (0, 2, 1), 其公式为:

$$x_t = 5.710 - 0.525\varepsilon_{t-1}。$$

Table 2. ARIMA model parameter table

表 2. ARIMA 模型参数表

项	常数项 ϕ_0	MA 参数 θ_1	Q 统计量 Q_6 (p 值)	AIC	BIC
值	5.710	-0.525	6.698 (0.350)	314	318

3.4. 未来 5 年农林牧渔业增加值预测

图 3 为采用 ARIMA (0, 2, 1)模型后, 陕西省农林牧渔业增加值的模型拟合以及未来 5 年变化的预测图, 由图 3 的拟合部分可以看出模型的拟合值曲线与真实值曲线吻合度高, 说明模型拟合效果好, 可以较为真实的预测未来五年陕西省农林牧渔业增加值。由预测部分可以看出, 随着预测时间的延长, 预测值的误差范围喇叭口型扩大, 表明 ARIMA 模型适合短期预测分析。

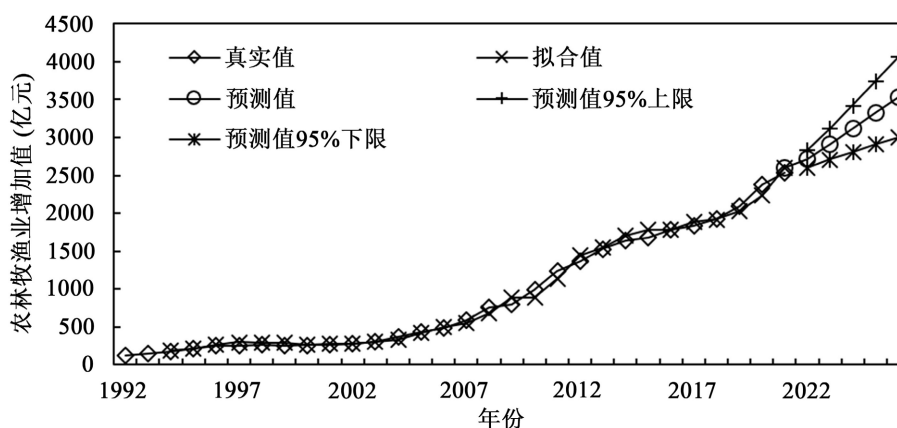


Figure 3. Model fitting and prediction of the added value of agriculture, forestry, animal husbandry and fishery in Shaanxi Province

图 3. 陕西省农林牧渔业增加值模型拟合和预测

Table 3. The five-year forecast table of the added value of agriculture, forestry, animal husbandry and fishery in Shaanxi Province

表 3. 陕西省农林牧渔业增加值五年预测表

年份	预测值(亿元)	实际值(亿元)	年净增长值(亿元)	增长率(%)
2021	2594.5	2532.2	-	-
2022	2721.1	-	188.9	7.46
2023	2915.6	-	194.5	7.15
2024	3115.9	-	200.3	6.87
2025	3321.9	-	206.0	6.61
2026	3533.6	-	211.7	6.37

由表 3 可知, 2021 年陕西省农林牧渔业增加值的预测值为 2594.5 亿元, 与实际值 2532.2 亿元相比, 相对误差为 2.46%, 小于 5%, 预测精度为 97.54%。未来五年(2022~2026 年)陕西省农林牧渔业增加值在没有自然灾害、大的突发公共事件以及政策等因素影响下, 会持续稳步增长, 五年平均增长率为 6.89%。

4. 结论

本文根据陕西省 1992~2021 年农林牧渔业增加值数据, 开展 2022~2026 年预测研究, 通过绘制并分析时间序列图, 发现原始时间序列不平稳, 而平稳性是 ARMA 模型预测的基础, 所以应用 ADF 检验和 KPSS 检验发现, 原始时间序列的二阶差分序列是平稳的, 可以进行模型建立, 最后应用 AIC 信息准则, 确定模型 ARIMA (0, 2, 1)。应用该模型预测发现, 2022~2026 年农林牧渔业增加值平均年净增长值为 200.3 亿元, 平均增长率为 6.89%, 陕西省农林牧渔业增加值将持续稳步增长。

参考文献

- [1] <http://www.stats.gov.cn/tjsj/ndsj/2021/html/zb12.htm>.
- [2] 景倩. ARMA 模型在我国农林牧渔业发展中的应用[J]. 消费导刊, 2008(7): 189-190.
- [3] 杨智凯, 范彦勤, 印海廷, 彭博. 基于 ARIMA 模型对桂林市 GDP 的预测研究[J]. 桂林航天工业学院学报, 2021, 26(4): 477-483.
- [4] 张梓. 基于 ARIMA 模型的贵州省 GDP 分析与预测[J]. 国土与自然资源研究, 2022(5): 39-41.
- [5] 李志超, 刘升. 基于 ARIMA 模型、灰色模型和回归模型的预测比较[J]. 统计与决策, 2019, 35(23): 38-41.
- [6] 严彦文. 基于 ARIMA 模型的山东省 GDP 的分析与预测[J]. 数学的实践与认识, 2018, 48(4): 285-292.