

基于ARIMA模型的陕西省粮食产量分析预测

郝起礼

陕西省土地工程建设集团有限责任公司商洛分公司, 陕西 商洛

收稿日期: 2022年10月19日; 录用日期: 2022年11月18日; 发布日期: 2022年11月25日

摘要

本文基于国家统计局官网关于陕西省2002~2021年的陕西省粮食产量数据, 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0.81%, 远小于5%, 表明模型拟合较好, 利用ARIMA (0, 1, 1)模型对陕西省粮食产量进行预测, 结果表明未来5年陕西粮食产量将持续稳步增长, 平均年净增长值为13.48万吨, 增长率为1.03%。

关键词

粮食产量, 陕西省, 时间序列分析, ARIMA模型

Analysis and Prediction of Grain Yield in Shaanxi Province Based on ARIMA Model

Qili Hao

Shangluo Branch of Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Shangluo Shaanxi

Received: Oct. 19th, 2022; accepted: Nov. 18th, 2022; published: Nov. 25th, 2022

Abstract

Based on the grain yield data of 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 established model is used to compare the predicted value with the actual value in 2021. The relative error is 0.81%, far less than 5%, indicating that the model fits well. The ARIMA (0, 1, 1) model is used to predict the grain yield of Shaanxi Province. The results show that Shaanxi's grain yield will continue to grow steadily in the next five years, with an average annual net growth value of 134,800 tons, with a growth rate of 1.03%.

Keywords

Grain Yield, Shaanxi Province, Time Series Analysis, ARIMA Model



1. 引言

粮食产量是关乎国计民生的大事,中国以世界上9%的耕地来养育全世界22%左右人口的用粮问题[1],供需压力尤为明显,而且“当今世界处于百年未有之大变局”,在动荡的国际形势以及严峻的疫情形势之下,摸清我国未来几年粮食的缺口、预测粮食产量变得尤为重要。2022年中央一号文件《中共中央国务院关于做好2022年全面推进乡村振兴重点工作的意见》更是明确提出“坚持中国人的饭碗任何时候都要牢牢端在自己手中”。随着国际化进程的不断深入,粮食已经不再是传统意义上的农产品,其具备了社会、文化、生态、政治等多重属性,事关国家安全,因此保障粮食安全,深入分析研判供需关系,实现粮食产量的有效预测,对中国的农业经济的发展、产业结构的调整以及农民的创收均有深远意义,同时一定程度上为政府部门合理引导粮食生产提供参考意义[2]。

陕西省作为西北地区重要的农业大省和经济中心,近几年随着经济的快速发展和户籍政策的调整,常住人口数量逐年增加,粮食产量预测对陕西省政府制定相应措施进而保障粮食充足供给十分重要。目前一些学者基于各种模型对中国各个省份乃至全国粮食产量进行了相关研究。谢玉莹[3]通过灰色关联模型筛选出影响粮食产量的主要因素,应用Shapley值理论对3个单一模型进行权重赋值并建立线性组合模型,实现了对粮食产量较为准确的预测。孟强[4]利用时间序列ARIMA(3,1,2)模型拟合了东三省2009~2013年的粮食产量,同时分析预测了东三省2014~2023年未来十年的粮食产量。孙英敏[5]以粮食生产的“投入指标”为评价依据,运用变异系数法,分析陕西省耕地利用转型和粮食产量耦合关系时空演变,结果显示2005~2015年陕西省各地市耕地利用强度逐年增大,但空间布局较分散。刘妍妮[6]基于《陕西省统计年鉴》数据,应用DPS软件建立回归分析模型和GM(1,1)模型,完成未来10年陕西省粮食产量预测,同时分析探讨了粮食产量变化及其影响因素。李志超[7]等通过建立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,但针对陕西省的粮食产量最近进展与未来预测的研究较少,本文将通过应用ARIMA模型,分析预测未来5年陕西省产量的预测值,为陕西省政府部门制定相应措施进而保障粮食充足供给提供一定的参考意义。

2. ARIMA 模型

ARIMA模型,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,其中AR、MA、I分别表示自回归模型、移动平均模型和差分,其主要针对的是平稳时间序列,但实际生活中大部分时间序列均是非平稳的,因此需要通过差分将其转化为平稳时间序列,然后通过建立ARIMA模型完成数据建模及预测。ARIMA(p, d, q)模型由AR、MA、I三部分组成,其中 p, d, q 分别表示自回归阶数、差分次数以及移动平均阶数[8]。模型表达式如下:

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

ARIMA模型的建立流程为首先检验时间序列的平稳性,若非平稳则需进行差分使其平稳,其次是模型ARIMA(p, d, q)的识别,接下来是模型的检验,即残差白噪声检验,最后是模型的具体应用。

3. 实例分析

3.1. 数据来源

本文数据均来自国家统计局官网(<https://data.stats.gov.cn/easyquery.htm?cn=E0103>), 数据真实、可靠。为深入研究陕西省粮食产量的发展趋势, 并预测未来 5 年的粮食产量, 笔者以陕西省 2002~2021 年粮食产量(表 1)为依据, 通过应用时间序列模型 ARIMA 对产量数据进行分析拟合预测。

Table 1. The grain yield data of Shaanxi Province from 2002 to 2021 (unit: 10,000 tons)

表 1. 陕西省 2002~2021 年粮食产量数据(单位: 万吨)

| | | | | | | | |
|------|---------|---------|---------|---------|---------|---------|---------|
| 年份 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| 粮食产量 | 1005.6 | 968.40 | 1040.00 | 1043.00 | 1041.90 | 1071.98 | 1120.25 |
| 年份 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 粮食产量 | 1149.06 | 1185.98 | 1207.39 | 1255.92 | 1210.55 | 1183.53 | 1204.67 |
| 年份 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | - |
| 粮食产量 | 1263.96 | 1194.20 | 1226.00 | 1231.13 | 1274.83 | 1270.43 | - |

数据来源: 国家统计局官网(2002~2021 年)。

3.2. 平稳性检验及处理

图 1 为 2002~2021 年陕西省粮食产量及其一阶差分时间序列图, 依据经验结合图像初步判断陕西省粮食产量时间序列是非平稳的, 而其一阶差分时间序列相对平稳。为进一步验证判断, 笔者通过应用 MATLAB 软件对原始时间序列及其一阶差分时间序列进行 ADF 和 KPSS 检验。结果显示, 原始时间序列的 ADF 检验和 KPSS 检验返回值分别为 0 和 1, 均表明该序列不平稳, 而其一阶差分时间序列的 ADF 检验和 KPSS 检验返回值分别为 1 和 0, 均表明时间序列平稳。

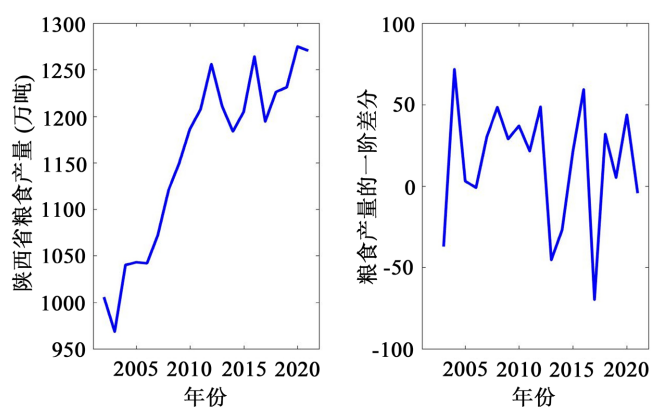


Figure 1. Time series graph of grain yield and its first-order difference in Shaanxi Province

图 1. 陕西省粮食产量及其一阶差分时间序列图

3.3. 模型建立

陕西省粮食产量时间序列经一阶差分后产生的时间序列通过了平稳性检验, 因此 ARIMA (p, d, q) 模型阶数 d 值可以确定为 1, 剩余仅需确定 p, q 值。图 2 为陕西省粮食产量的一阶差分序列的自相关(ACF)

和偏自相关(PACF)图,从 ACF 图可以发现自 0 阶后衰减趋于零,而 PACF 图在 1 阶和 12 阶之后衰减趋于零,通过 AIC 信息准则(值越小越好),我们可以进一步确定模型阶数。表 2 为 AIC 和 BIC 值较小的模型参数统计表,得到的最优模型为 ARIMA (0, 1, 1), 其公式为: $x_t = 14.789 - 0.367\varepsilon_{t-1}$ 。

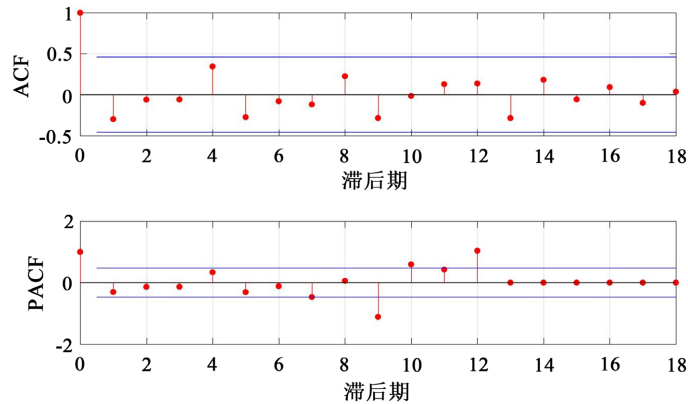


Figure 2. Autocorrelation and partial autocorrelation plots of first difference sequences

图 2. 一阶差分序列的自相关和偏自相关图

Table 2. Model parameter statistics table

表 2. 模型参数统计表

| 项 | 常数项 ϕ_0 | MA 参数 θ_1 | AIC | BIC |
|---|--------------|------------------|-------|-------|
| 值 | 14.789 | -0.367 | 194.5 | 197.4 |

3.4. 模型检验

为了验证所建模型的可靠性,需要对模型残差进行白噪声检验,ARIMA 模型要求残差不存在自相关性,即模型为白噪声,可通过 Q 统计量检验进行判断。表 3 为模型 Q 统计量表,通常选取 Q_6 进行分析,其对应的 p 值远大于 0.1,即模型残差序列为白噪声,表明模型通过检验,基本满足要求,最终确定预测模型为 ARIMA (0, 1, 1)。

Table 3. Model Q statistic table

表 3. 模型 Q 统计量表

| 项 | 统计量 | p 值 | 项 | 统计量 | p 值 |
|-------|-------|-------|----------|--------|-------|
| Q_1 | 0 | 0.993 | Q_7 | 6.111 | 0.527 |
| Q_2 | 0.071 | 0.965 | Q_8 | 6.316 | 0.612 |
| Q_3 | 0.072 | 0.995 | Q_9 | 9.717 | 0.374 |
| Q_4 | 1.905 | 0.753 | Q_{10} | 9.896 | 0.450 |
| Q_5 | 3.913 | 0.562 | Q_{11} | 10.947 | 0.448 |
| Q_6 | 5.261 | 0.511 | Q_{12} | 11.787 | 0.463 |

3.5. 粮食产量预测

图 3 为模型 ARIMA (0, 1, 1)关于陕西省粮食产量拟合情况以及未来 5 年粮食产量预测情况,由图可以看出真实值和拟合值吻合度高,而且随着时间的推移,陕西省粮食产量逐渐增加,并且随着预测年限

的增加, 预测值的误差范围逐渐增大, 表明模型拟合效果好, 可较为真实反映未来 5 年陕西省粮食产量变化, 但其预测分析的年限受限, 适合短期分析预测。

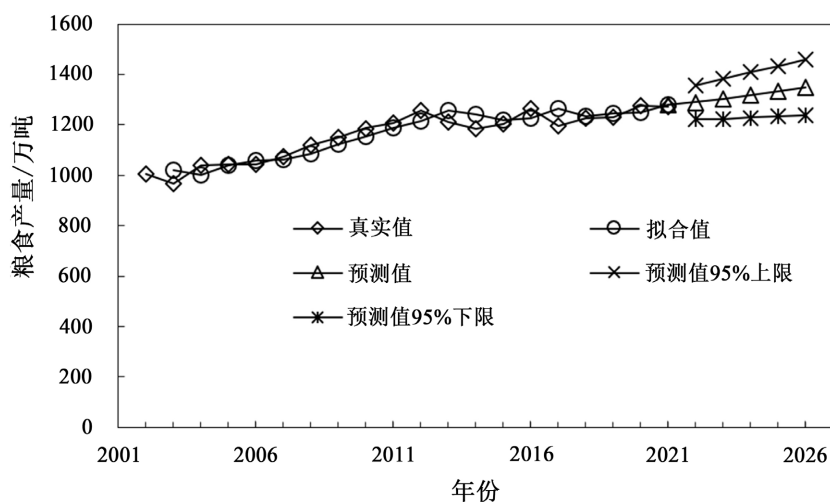


Figure 3. Shaanxi Province's grain yield model fitting and yield forecast for the next 5 years

图 3. 陕西省粮食产量模型拟合和未来 5 年产量预测

Table 4. The forecast table of grain yield in Shaanxi Province in the next five years

表 4. 陕西省粮食产量未来 5 年预测表

| 年份 | 预测值 (万吨) | 真实值 (万吨) | 年净增长值 (万吨) | 增长率(%) |
|------|-------------|-------------|---------------|--------|
| 2021 | 1280.75 | 1270.43 | - | - |
| 2022 | 1289.00 | - | 8.25 | 0.65 |
| 2023 | 1303.79 | - | 14.79 | 1.15 |
| 2024 | 1318.58 | - | 14.79 | 1.13 |
| 2025 | 1333.37 | - | 14.79 | 1.12 |
| 2026 | 1348.16 | - | 14.79 | 1.11 |

由表 4 可以看出 2021 年陕西省粮食产量真实值和预测值分别为 1270.43 万吨和 1280.75 万吨, 相对误差仅为 0.81%, 远远小于 5%, 预测精度高达 99.19%。预测结果显示, 未来 5 年陕西省粮食产量平均年净增长值为 13.48 万吨, 增长率为 1.03%。

4. 结语

本文通过选取陕西省 2002~2021 年粮食产量数据, 经过分析处理检验, 构建 ARIMA (0, 1, 1)模型, 完成了 2022~2026 年陕西省粮食产量预测。以 2021 年的数据进行验证, 结果显示模型精度高达 99.19%。预测结果显示未来五年陕西省粮食产量呈逐年增加趋势, 且平均年净增长值和增长率分别为 13.48 万吨和 1.03%, 为陕西省经济决策提供一定参考。同时随着预测年限的增加, 预测精度有所下降。因此今后在建立 ARIMA 模型时应多研究改进算法, 进一步提高预测精度。

参考文献

- [1] 殷培红, 方修琦, 马玉玲. 21 世纪初我国粮食供需的先空间格局[J]. 自然资源学报, 2006, 21(4): 625-632.
- [2] 杨娜. 粮食产量预测方法研究[D]: [硕士学位论文]. 郑州: 河南工业大学, 2017.
- [3] 谢玉莹. 基于 Shapley 值的陕西省粮食产量预测[J]. 现代食品, 2022, 28(15): 1-4.
- [4] 孟强, 李海晨. 基于 ARIMA 模型的东三省粮食产量时间序列分析[J]. 科学技术创新, 2015(21): 128-129.
- [5] 孙英敏. 陕西省耕地利用转型与粮食产量耦合关系研究[J]. 中国农业资源与区划, 2018, 39(7): 14-20.
- [6] 刘妍妮. 陕西省粮食播种面积及其产量动态与预测[J]. 江西农业学报, 2018, 30(10): 142-146.
- [7] 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.
- [8] 查华, 石舫. 基于 ARIMA 模型对江苏省 GDP 的预测[J]. 兰州文理学院学报(自然科学版), 2022, 36(3): 33-36+54.