

基于ARIMA模型的蔬菜价格预测研究 ——以四川省青椒价格为例

王章琴, 蓝尉太, 张淑亭, 刘芋良, 贾礼平*

乐山师范学院数学与统计学院, 四川 乐山

收稿日期: 2025年6月22日; 录用日期: 2025年7月12日; 发布日期: 2025年7月25日

摘 要

文章通过“全国重点农产品市场信息平台”采集2015年1月~2025年3月四川省青椒市场批发价格数据, 利用ARIMA模型对四川省未来9个月的青椒市场价格做出预测。结果表明: 2025年4月~2025年12月四川省青椒市场价格整体呈现小幅上涨趋势, 价格维持在4.60元/kg~5.00元/kg, 没有明显的波动, 平均价格为4.85元/kg, 较近5年有下降。若2025年四川省青椒产量维持在正常水平, 在排除极端天气等外生变量条件下, 种植户可参考4.60元/kg~5.00元/kg的基准线制定采收计划。

关键词

四川省青椒价格, ARIMA模型, 预测

Research on Vegetable Price Forecasting Based on ARIMA Model

—Taking Sichuan Pepper Price as an Example

Zhangqin Wang, Weitai Lan, Shuting Zhang, Yuliang Liu, Liping Jia*

School of Mathematics and Statistics, Leshan Normal University, Leshan Sichuan

Received: Jun. 22nd, 2025; accepted: Jul. 12th, 2025; published: Jul. 25th, 2025

Abstract

The article through the “national key agricultural products market information platform” collection of January 2015-March 2025 Sichuan Province, green pepper market wholesale price data, the use of ARIMA model in Sichuan Province in the next nine months to make predictions on the market

*通讯作者。

文章引用: 王章琴, 蓝尉太, 张淑亭, 刘芋良, 贾礼平. 基于 ARIMA 模型的蔬菜价格预测研究[J]. 统计学与应用, 2025, 14(7): 221-231. DOI: 10.12677/sa.2025.147199

price of green pepper. The results show that: the overall market price of green pepper in Sichuan Province in April 2025-December 2025 shows a slight upward trend, the price is maintained at 4.60 yuan/kg~5.00 yuan/kg, there is no obvious fluctuation, and the average price is 4.85 yuan/kg, which is a decrease compared with the last five years. If the production of green peppers in Sichuan Province in 2025 is maintained at the normal level, under the condition of excluding exogenous variables such as extreme weather, growers can refer to the baseline of RMB 4.60~5.00/kg to formulate a harvest plan.

Keywords

Sichuan Province Green Pepper Price, ARIMA Model, Forecasting

Copyright © 2025 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

我国农业经济中,作为日常必需品的蔬菜的价格波动一直受着大量的关注。蔬菜是我们日常饮食的核心部分,它的价格波动影响着我们的一日三餐、农民的收益以及社会的稳定程度。四川盆地的蔬菜产业占农产品产业很大一部分,不仅是农民收入的主要来源,还是家家户户百姓的必需品。但随着市场经济的发展与生产方式的变革,蔬菜价格的波动日益加剧,给农民的收入和社会的稳定性带来了不小的挑战。

面对这一挑战,若能及时、准确的预测蔬菜的价格,不仅能帮助农民合理的规划种植品类,作出合理的生产决策,提高收入,还能为政府部门的政策制定提供有力的支持,维持社会的稳定和我们的基本生活需求。但蔬菜价格受到天气、政策和市场需求等多种因素的共同影响[1],导致蔬菜价格的波动呈现出高度的复杂性和不确定性,想要预测并非易事。

为降低预测难度,文章将研究对象确定为:四川省某一特定蔬菜的价格——青椒的价格。尝试运用 ARIMA (Auto Regressive Integrated Moving Average) [2]这一时间序列分析方法,对其价格进行短期预测。

ARIMA 模型在统计预测领域有坚实的理论基础和广泛的应用背景。ARIMA 模型通过整合自回归 AR (AutoRegressive)、差分 I (Integrated)和移动平均 MA (Moving Average)三个核心模块可以有效分解时间序列中的趋势性、季节性与随机波动成分。文章将四川省青椒批发价格当作研究对象,通过将预处理后的数据输入模型来不断改进模型并预测蔬菜的价格,给农业生产决策与市场调控提供数据支撑。

2. 研究数据及研究方法

2.1. 数据来源

研究的数据来自全国重点农产品市场信息平台(网址: ncpscxx.moa.gov.cn)。在符合平台的使用规定和数据保护原则、保证数据的合法性和可靠性的基础上,查询了该平台提供的 2015 年 1 月至 2025 年 3 月期间四川省青椒的批发价格数据并汇总为数据集。该数据集包含了四川省每个月的黑椒批发价格,为预测黑椒价格的 ARIMA 模型提供了准确的数据。

2.2. 数据预处理

按照以下步骤进行了预处理来确保数据准确性和可用:

(1) 数据的时间粒度：本文采集“全国重点农产品市场信息平台”2015年1月~2025年3月四川省青椒市场月批发价格数据，数据的时间粒度为月级粒度，适用于对市场青椒价格趋势的长期预测分析，能够捕捉到季节性波动等因素的影响。

(2) 数据清洗：修改了异常值和缺失值。异常值可能是因为数据录入出错或其他特殊情况，缺失值可能是由于数据收集过程中有遗漏。经过检验，原始数据不存在异常值，但存在两个缺失值，本文使用 SPSS “替换缺失值”的功能，并用序列平均值来替换。

(3) 数据转换：原始数据以月份和年份的组合形式呈现(如“Apr-15”表示2015年1月)，利用 SPSS 软件转换为了时间序列数据格式以便进行后续的时间序列分析。

完成这些步骤后就得到了一个高质量、可用于时间序列分析的数据集。

2.3. 研究方法

2.3.1. ARIMA 模型

文章选择 ARIMA (Auto Regressive Integrated Moving Average, 自回归差分移动平均)模型作为主要的预测工具。ARIMA 模型经常应用于经济、医疗、农业等领域的数据预测。

ARIMA 模型是通过识别时间序列数据中的自回归(AR)部分、差分(I)部分和移动平均(MA)部分，来构建预测模型。自回归部分描述数据与其历史值之间关系；差分部分用于消除数据中趋势性和季节性成分；移动平均部分考虑了随机误差项的影响。

P 阶自回归模型[3]为

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t \quad (1)$$

其中， t 表示使用第几期历史值来预测； y_t 是当前值； μ 是常数项； p 是阶数； γ_i 是自相关系数。

移动平均模型[4]为

$$y_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2)$$

ARIMA 模型为

$$y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

其中， Y 是正在考虑的时间序列数据， φ_1 到 φ_p 是模型用来描述当前值与过去 p 个时间点的值之间的关系的参数； θ_1 到 θ_q 是模型用来描述当前值与过去 q 个时间点的误差之间的关系的参数； ε_t 是在时间点的误差项； c 是在 t 时间点的误差项。

2.3.2. ARIMA 预测模型

按照如下步骤构建 ARIMA 预测模型：

(1) 模型识别：绘制时间序列图、自相关函数(ACF)图和偏自相关函数(PACF)图分析数据的趋势性、季节性和随机性成分。

(2) 模型选择：基于模型识别的结果，选择合适的 ARIMA 模型参数(即 p 、 d 、 q 值)，其中 p 表示自回归项的阶数， d 表示差分次数， q 表示移动平均项的阶数。

(3) 模型估计：采用最大似然估计法[4]，对选定的 ARIMA 模型进行参数估计，计算得出模型的各项系数。

(4) 模型检验：采用白噪声检验残差是否存在自相关性，拉格朗日乘数检验模型残差序列是否存在序

列相关。若模型存在显著误差或不符合预期，将重新进行模型选择和估计，直至模型符合预期。

(5) 预测应用：利用构建好的 ARIMA 模型对青椒价格进行短期预测，分析所得预测结果，为农民的生产决策、政府部门的政策制定以及大家的购买行为提供科学依据和指导。

3. 四川省青椒价格波动特征分析

3.1. 青椒价格季节波动特征

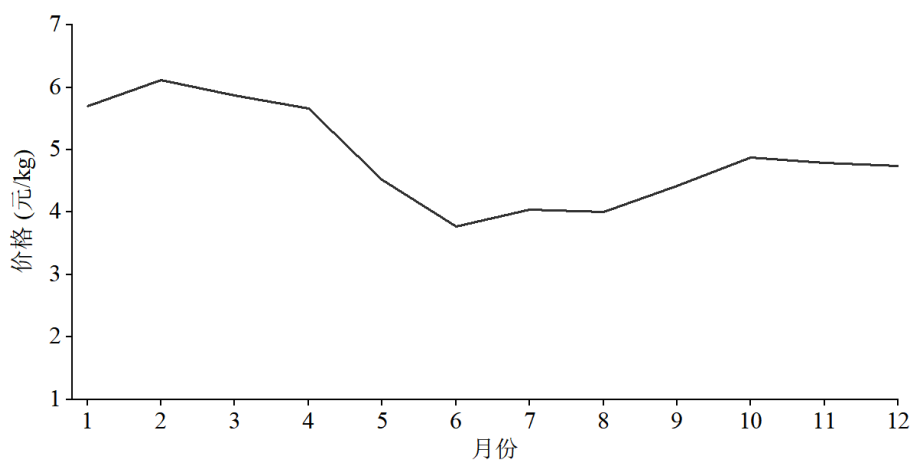


Figure 1. Monthly average price of green peppers in Sichuan province from January 2015 to March 2025

图 1. 2015 年 1 月~2025 年 3 月四川省青椒月度平均价格

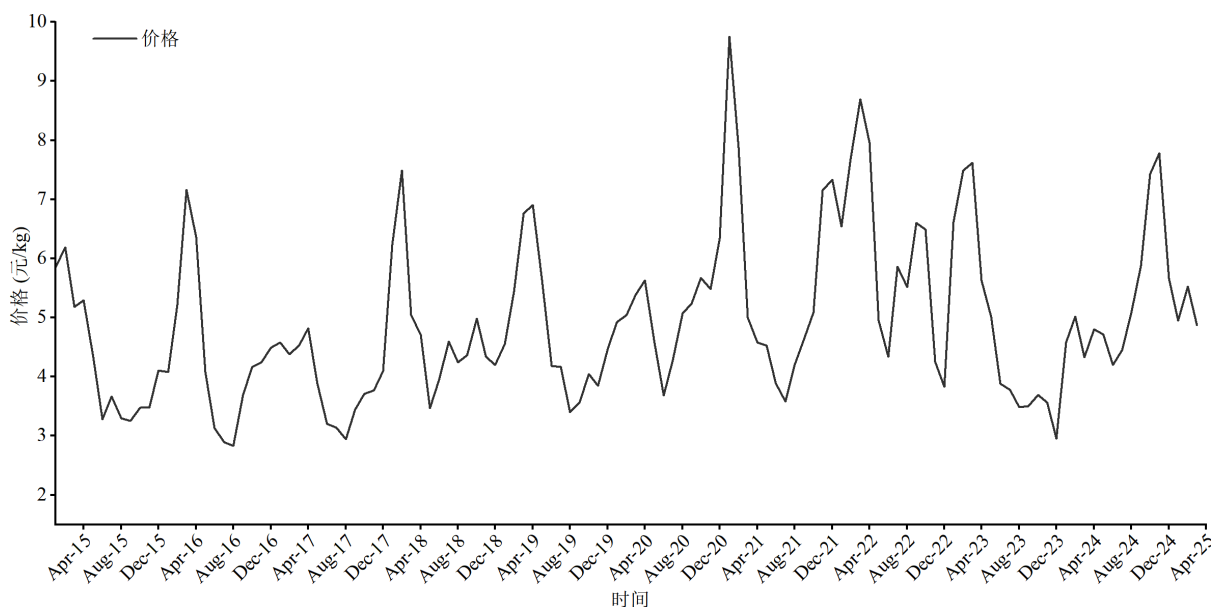


Figure 2. Green pepper prices in Sichuan province from January 2015 to March 2025

图 2. 2015 年 1 月~2025 年 3 月四川省青椒价格

对 2015 年 1 月至 2025 年 3 月四川省青椒各月度价格采取了算术平均的处理方法，可以看出青椒价格呈现出显著的季节性波动特征。由图 1 可知：1~4 月，青椒价格处于全年高位，2 月价格达到年度价格最大值；5~8 月价格呈下降趋势，6 月价格达到年度价格最低值；9~12 月青椒价格又开始缓慢上升。整

体上呈现出“冬高夏低”的周期性波动，青椒上市以前的2月份价格达到最高点，为6.12元/kg；青椒大量上市以后的6月份达到价格最低值，为3.77元/kg。

从图2可知，四川省青椒价格季节波动特征显著。2015、2018年波峰出现在2月；2016、2022、2023年波峰出现在3月；2017、2019年波峰出现在4月；2020、2021、2024年波峰分别出现在12、1、11月。2016、2017、2019年波谷出现在8月；2020、2022、2024年波谷出现在6月；2015、2018、2021、2023年波谷分别出现在9、5、7、12月。青椒价格波动较为平稳。

青椒价格的季节性波动与其生长周期和气候条件有关。从供求关系来看，5至8月份是青椒的收获期，青椒集中上市，市场供给充足。同时，夏季高温多雨会加速青椒的腐烂，商家会主动降价促销来减少库存积压，进一步导致青椒价格降低。9至12月份，青椒供给逐渐减少，价格开始上涨。到了冬季，青椒受到低温等环境因素的限制，供应量进一步减少，价格维持在一个较高水平。1~2月春节期间人们对蔬菜的需求量增加，而青椒作为一种常见蔬菜，需求大大增加，从而推动了价格上涨，所以每年1、2月份青椒价格往往会达到一年中的最高价。

3.2. 青椒价格不规则波动特征

通过剔除原始序列中的季节变动和循环变动趋势，我们得到了青椒价格的不规则波动图。从图3可以看出，2015~2019年间青椒价格波动相对平缓。2020~2022年间青椒价格波动剧烈。这一现象可能与新冠疫情的爆发和反复有关。2020年初疫情突然爆发导致餐饮行业需求减少，青椒价格骤降。2021年初春节期间，人们对蔬菜的需求量显著增加，但受到疫情防控的影响，蔬菜运输成本上升，青椒价格又再次上涨。2022年局部的疫情反复又进一步加剧了青椒价格的波动。同时，市场上价格的剧烈波动也会对后续年份的市场供需产生连锁反应。2020年青椒价格处于历年相对较低水平，农户种植青椒的意愿降低，从而导致次年青椒的价格上涨。而2021年的高价可能刺激农户扩大青椒的种植面积，造成2022年青椒供给过剩，价格下跌。2023年以后，青椒价格的波动幅度与之前相比明显减小，这可能是由于青椒市场供需关系逐渐趋向稳定、市场机制不断健全以及外部经济环境变化减少，使得青椒价格逐渐稳定。

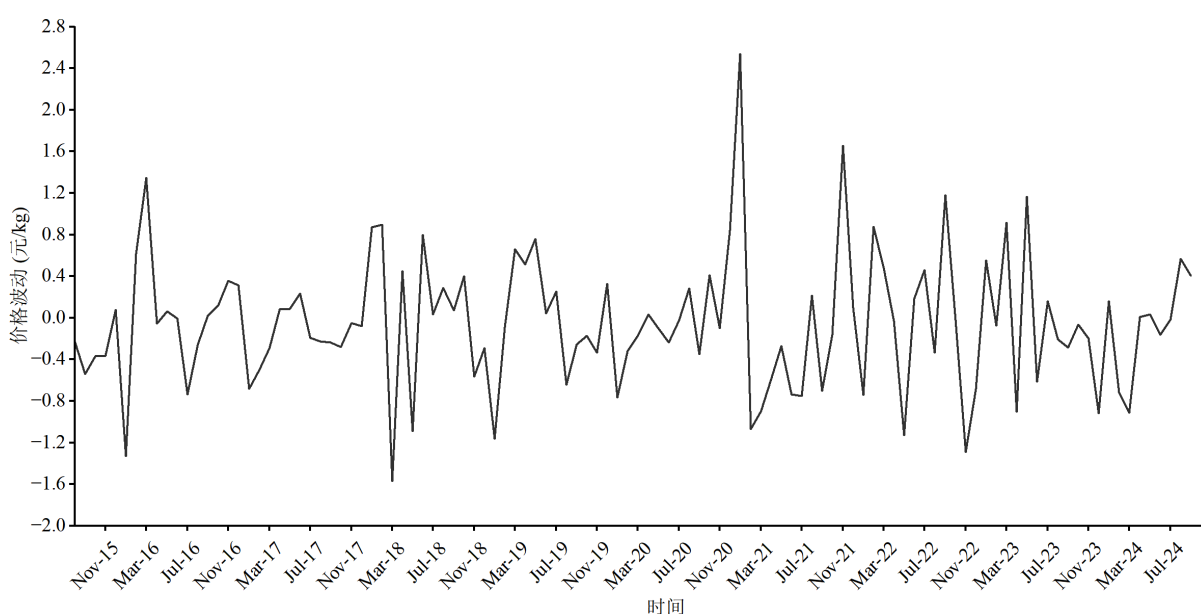


Figure 3. Irregular fluctuations in green pepper prices in Sichuan province from January 2015 to March 2025

图3. 2015年1月~2025年3月四川省青椒价格的不规则波动

4. 基于 ARIMA 的蔬菜价格预测实证结果分析

4.1. 平稳性检验

构建 ARIMA 模型必须保证选取的序列具有平稳性, 因此首先对序列进行平稳性检验。ADF 检验结果由表 1 可见, 针对价格, 当差分阶数为 0 时, ADF 值分别小于不同检验水平下的其他三个临界值, 且原时间序列数据的显著性 P 值为 0.000, 即 $P=0.000<0.01$, 所以有高于 99% 的把握拒绝原假设(原假设: 此序列不是平稳序列), 此时序列平稳, 可以不对原始数据进行差分处理, 则模型 $d=0$ 。

Table 1. ADF test results for original data

表 1. 原始数据 ADF 检验结果

差分阶数	t	P	临界值		
			1%	5%	10%
0	-4.863	0.000	-3.487	-2.886	-2.580
1	-6.372	0.000	-3.491	-2.888	-2.581
2	-7.578	0.000	-3.493	-2.889	-2.581

4.2. 模型系数选取

通过分析自相关(ACF)和偏自相关图(PACF)来确定模型的 q 和 p 。图 4 显示在滞后 1 处有显著的尖峰, 之后相关性逐渐衰减到零, 从滞后 3 处开始, 后面大部分都在范围内符合 MA(q)模型截尾特征, 所以 $q=4$ 。而图 5 在滞后 1 处也有显著尖峰, 滞后 1 之后, 值迅速下降并在更高滞后期接近零, 这种在滞后 1 后快速截尾的模式表明自回归(AR)成分应为 $p=1$ 。综上, 模型 $q=4$ 、 $p=1$, 建立的模型为 ARIMA(1,0,4)。

ACF图

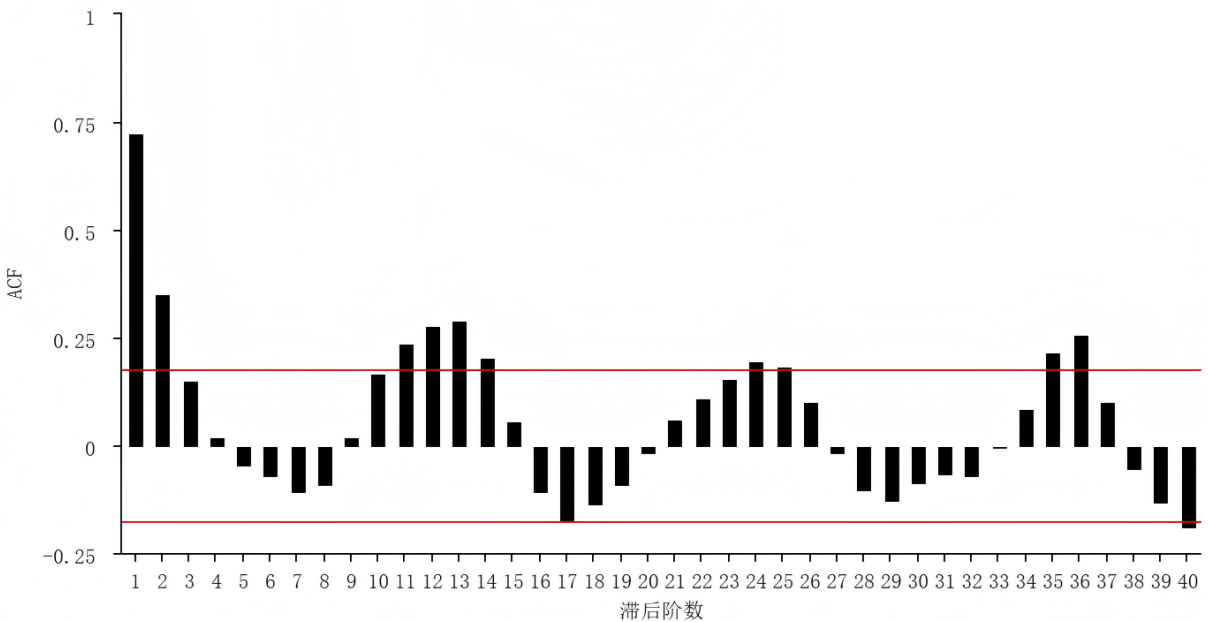


Figure 4. Diagram of ACF

图 4. ACF 图

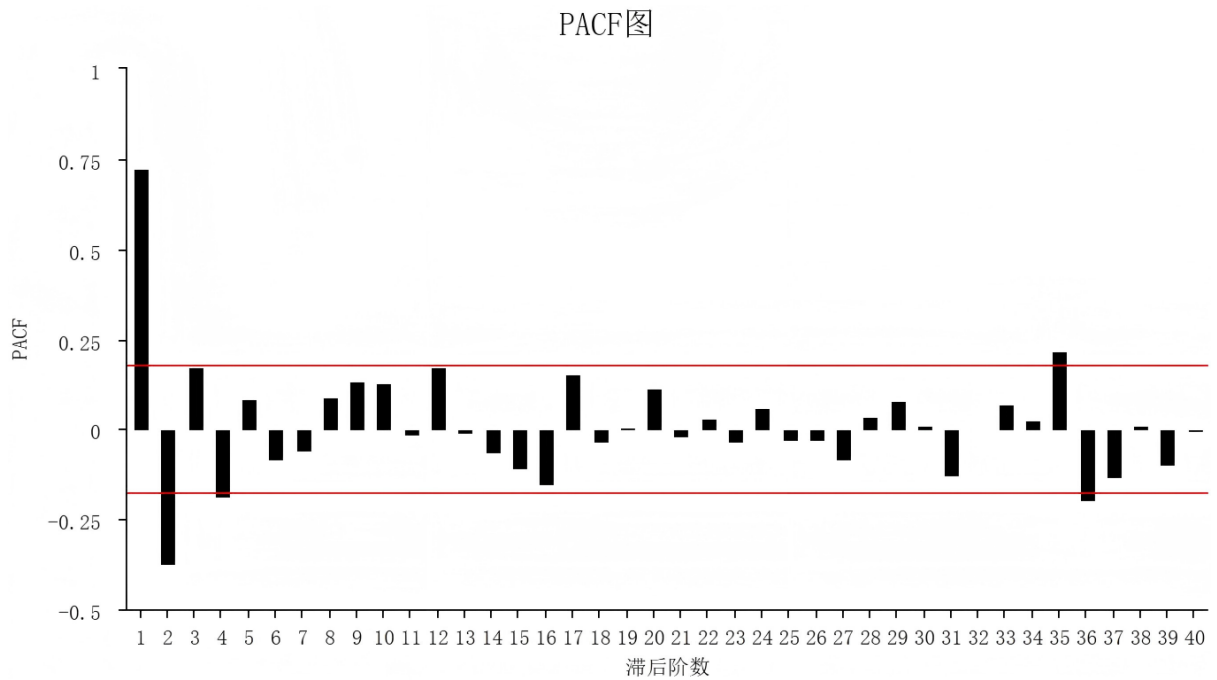


Figure 5. Diagram of PACF
图 5. PACF 图

4.3. 模型检验

Table 2. Q-statistics table for the model
表 2. 模型 Q 统计量表

排序	稿件	件数
项	统计量	p 值
Q1	0.002	0.965
Q2	0.002	0.999
Q3	0.003	1.000
Q4	0.014	1.000
Q5	0.065	1.000
Q6	0.253	1.000
Q7	0.328	1.000
Q8	2.110	0.977
Q9	2.152	0.989
Q10	4.753	0.907
Q11	5.369	0.912
Q12	5.457	0.941
Q13	10.917	0.618
Q14	11.111	0.677
Q15	12.126	0.669

ARIMA 模型要求模型残差为白噪声,即残差不存在自相关,可通过 Q 统计量的 P 值进行白噪声检验(原假设:残差是白噪声),若 P 值超过 0.1,则认为残差是白噪声。结果如表 2 所示, Q 统计量的 P 值均大于 0.1,所以接受原假设,即模型为白噪声。

拉格朗日乘数(Breush-Godfrey LM 检验)用于检验模型残差序列是否存在序列相关。LM 检验原假设 H0 为序列不存在序列相关,备择假设 H1 为序列存在序列相关;如果对应 P 值小于 0.05,即拒绝原假设,意味着存在序列相关,反之 P 值大于 0.05,即接受原假设,意味着序列不存在序列相关。表 3 中 LM 检验的两个统计量的 P 值均大于 0.05,所以模型残差不存在序列相关。

Table 3. LM test for residual terms
表 3. 残差项 LM 检验

F 统计量	0393	P 值	0.947
T × R ² 统计量	4.194	P 值	0.938

4.4. 敏感性分析

所有时序数列通过检验后,利用 ARIMA 模型做出预测。根据表 4 的 ARMA(1,4)模型参数表,可得模型公式为

$$y(t) = 7.085 - 0.446 \times y(t-1) + 1.576 \times \varepsilon(t-1) + 1.065 \times \varepsilon(t-2) + 0.538 \times \varepsilon(t-3) + 0.246 \times \varepsilon(t-4) \quad (4)$$

其中,常数项为 7.085,代表四川省青椒价格的基础水平或长期价格;自回归系数为-0.446,表示当前月青椒价格约有 44.6%受上个月价格的影响,并且系数小于 1 又接近 0.5,说明价格有一定惯性,但会逐渐回归均值,也体现出农产品价格的连续性特征。移动平均系数 θ_1 影响最大(1.576),后续阶数依次减弱。

Table 4. Parameter table of ARIMA(1,4) model
表 4. ARIMA(1,4)模型参数表

项	符号	系数	标准误	z 值	p 值	95% CI
常数项	c	7.085	1.684	4.209	0.000	3.786~10.385
AR 参数	φ_1	-0.446	0.312	-1.431	0.152	-1.057~0.165
MA 参数	θ_1	1.576	0.320	4.926	0.000	0.949~2.203
	θ_2	1.065	0.383	2.778	0.005	0.313~1.816
	θ_3	0.538	0.184	2.928	0.003	0.178~0.898
	θ_4	0.264	0.08	3.303	0.001	0.107~0.421

图 6 为 2015 年 1 月~2025 年 3 月四川省青椒价格的真实值与拟合值的对比图,经比较发现,均方误差为 0.70 平均绝对百分比误差为 0.13,所以预测结果较好。在 95%的置信水平为下,对 2025 年 4 月~2025 年 12 月四川省青椒价格进行预测分析,得到预测结果如表 5 所示。

从表 5 的预测结果可以看出,青椒价格在 2025 年 4 月最低(4.62253),2025 年 8 月最高(4.92563),无明显季节性波动,说明供需关系或成本推动的长期趋势主导。价格整体呈现波动式小幅上涨趋势,从 4 月的 4.62253 上升到 12 月的 4.90101,在 6 月、9 月均有小幅度下降,10 月到 12 月也有微小的回调。而各月置信区间上下限宽度较大,差值在 3.3 左右(如 4 月为 6.25080-2.99426=3.25654),说明模型对价格的预测存在一定的不确定性。且远期预测的置信区间较宽,近期预测的置信区间收窄,表明模型对远期预测的不确定性较大,对近期预测趋于稳定。

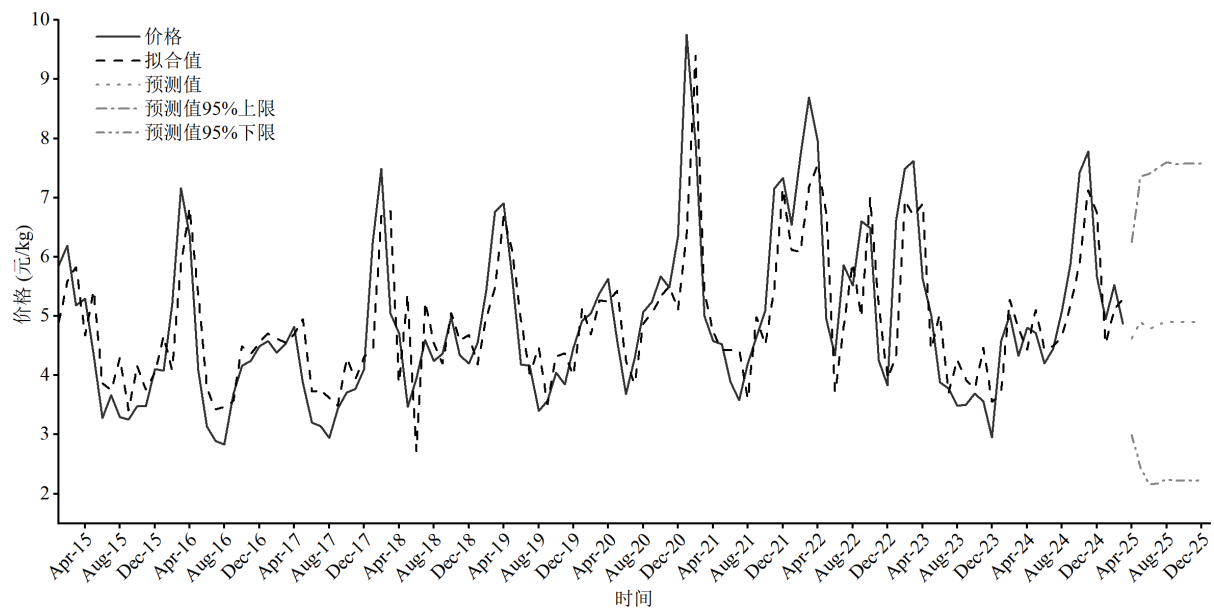


Figure 6. Comparison chart of actual and fitted values of green pepper prices

图 6. 青椒价格真实值与拟合值对比图

Table 5. Prediction results of green pepper prices: next 9 months

表 5. 未来 9 个月青椒价格预测结果

	预测值	95%置信区间上限	95%置信区间下限
2025 年 4 月	4.62253	6.25080	2.99426
2025 年 5 月	4.90660	7.36324	2.44996
2025 年 6 月	4.77994	7.40092	2.15895
2025 年 7 月	4.84250	7.50502	2.17998
2025 年 8 月	4.92563	7.59735	2.25392
2025 年 9 月	4.88856	7.56210	2.21502
2025 年 10 月	4.90509	7.57900	2.23119
2025 年 11 月	4.89772	7.57169	2.22374
2025 年 12 月	4.90101	7.57500	2.22702

Table 6. Sensitivity analysis of model parameters

表 6. 模型参数敏感性分析

模型参数	AIC 值	BIC 值	RMSE
(1, 0, 4)	319.118	338.804	0.8349
(1, 1, 4)	321.063	340.692	0.8496
(2, 0, 4)	319.995	339.492	0.8202
(1, 0, 3)	320.453	338.926	0.8467
(1, 0, 5)	321.135	343.632	0.8348

本文 ARIMA 模型的主要参数为(1, 0, 4)，主要根据序列的平稳性、MA (q)模型、AR (q)模型的截尾

特征分别确定模型的 p 、 d 、 q 参数。为了测试模型对参数微调的敏感性，构造了不同参数的 ARIMA 模型，如表 6 所示。可以看出，模型(1, 0, 4)表现最优，AIC 值与 BIC 值均为最低，说明该模型在拟合优度和预测精度优于其他参数组合。参数调整对模型性能影响显著，如当差分次数 d 提高，RMSE 值反而上升，说明原始序列无需过度差分。滞后阶数 q 对模型影响较小，但会对 BIC 值产生一定影响。

4.5. 性能评估

为确认模型预测的准确性，利用季节性自回归积分滑动平均模型[5] (SARIMA)与 ARIMA 做对比，选用不同评估指标来量化模型预测值与真实值的偏差程度。本文所选的评估指标为均方根误差(RMSE)，衡量实际值与预测值间的偏差，值越小越好；均方误差(MSE)，衡量预测误差的平方，值越小越好；平均绝对误差(MAE)，即误差的平均值，值也越小越好；平均绝对百分比误差(MAPE)，即误差的相对比例，值越小越好；决定系数(R^2)。

$$\left\{ \begin{array}{l} RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \\ R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \end{array} \right. \quad (5)$$

其中， n 为样本数量， y_i 为真实值， $\hat{y}_i$ 为预测值， $\bar{y}$ 为平均值。

表 7 为模型评估指标值，可以看出相比于 SARIMA，ARIAM 的 RMSE 值与 MSE 值更低，说明它的预测误差更小，即模型的精确性更高；两模型的 MAE 值非常接近，意味着二者平均误差相似；但 SARIMA 的 MAPE 值略低于 ARIMA，且 R^2 值更高，说明 SARIMA 在相对误差百分比方面表现更好，且对数据的拟合程度更好，更能够解释数据的变化。综上，ARIMA 的精确性更好，但拟合程度略低于 SARIMA。

Table 7. Model performance evaluation

表 7. 模型性能评估

	RMSE	MSE	MAE	MAPE	R^2
ARIMA (1, 0, 4)	0.8349	0.6971	0.6383	13.01%	0.691
SARIMA (1, 0, 4) (1, 1, 1)	0.8501	0.7227	0.6334	12.96%	0.758

5. 结论

通过分析四川省历年青椒市场批发价格，发现 2015 年~2025 年青椒市场批发价格呈现逐步上涨趋势，特别在 2020 年后价格出现明显的上升趋势，虽然在接下来的几年有所回落，但仍然维持在较高水平，显示出一定的市场需求上升。

文章利用 ARIMA 模型对四川省未来 9 个月的青椒价格进行预测分析,通过 SPSS 对模型参数进行计算、调试,并对比多个不同参数的模型,最终确定 ARIMA(1,0,4)为预测模型。根据 2015 年 1 月~2025 年 3 月四川省青椒价格的数据,预测出 2025 年 4 月~2025 年 12 月的价格变化情况,如图 6 所示,四川省未来 9 个月青椒整体价格有小幅上升,但月度平均价格将维持在 4.60 元/kg~5.00 元/kg,2025 年年度平均价格为 4.85 元/kg,较近 5 年有下降趋势,但整体上呈现出“冬高夏低”的周期性波动。最终预测结果表明,若 2025 年四川省青椒产量维持在正常水平,在排除极端天气与市场干预等外生变量条件下,青椒价格预计不会出现剧烈波动,且种植户可参考 4.60~5.00 元/kg 的基准线制定采收计划。

基金项目

2024 年四川省大学生创新创业训练计划项目(S202410649104)。

参考文献

- [1] 罗琴. 重庆市蔬菜价格形成及影响因素分析[J]. 农村经济与科技, 2021, 32(3): 169-172。
- [2] 刘泳君, 谢欣雨, 高凯. 基于 ARIMA 算法的饮料产品需求预测研究[J]. 全国流通经济, 2025(1): 12-15。
- [3] 刘芹, 常仁杰, 毕立雄, 等. 基于时间序列自回归移动平均模型的我国政府卫生支出预测研究[J]. 医学与社会, 2021, 34(11): 64-69。
- [4] 吴铭辉, 高文根, 华峰, 等. 基于最大似然估计的智能电网 FDIA 检测[J]. 四川轻化工大学学报(自然科学版), 2023, 36(2): 38-45。
- [5] 韩宗华, 韦登高. 基于 SARIMA 模型的广西某三甲医院手术量预测研究[J]. 信息与电脑, 2024, 36(24): 181-183。