# 基于SARIMA模型的河北省居民人均 消费性支出预测

## 杨晓彤

云南财经大学统计与数学学院,云南 昆明

收稿日期: 2023年1月22日: 录用日期: 2023年2月12日: 发布日期: 2023年2月27日

## 摘要

本文基于2013年第1季度~2022年第3季度河北省居民消费性支出的季度数据,利用Rstudio软件进行分析,以2013第1季度~2021年第4季度的数据为训练集,以2022年的数据为测试集。原始序列为非平稳序列,1阶4步差分后,序列趋于平稳,故采用SARIMA加法模型进行预测,并根据AIC信息准则确定模型的拟合效果最优,且其预测结果通过了显著性检验,能够为各级政府提出平衡供需市场的策略提供一定的参考价值。

#### 关键词

SARIMA模型,居民消费性支出,消费预测

# Prediction of Per Capita Consumer Expenditure in Hebei Province Based on SARIMA Model

## **Xiaotong Yang**

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

Received: Jan. 22<sup>nd</sup>, 2023; accepted: Feb. 12<sup>th</sup>, 2023; published: Feb. 27<sup>th</sup>, 2023

#### **Abstract**

Based on the quarterly data of consumer spending of residents in Hebei Province from the first quarter of 2013 to the third quarter of 2022, this paper uses Rstudio software for analysis, taking the data from the first quarter of 2013 to the fourth quarter of 2021 as the training set, and the

文章引用: 杨晓彤. 基于 SARIMA 模型的河北省居民人均消费性支出预测[J]. 统计学与应用, 2023, 12(1): 186-195. DOI: 10.12677/sa.2023.121020

data from 2022 as the test set. The original series is non-stationary, and the series tends to be stable after the first order four-step difference, so the SARIMA additive model is used for prediction, and the fitting effect of the model is determined to be the best according to AIC information criteria, and the prediction results have passed the significance test, which can provide a certain reference value for governments at all levels to propose strategies to balance the supply and demand market.

#### **Keywords**

SARIMA Model, Consumer Expenditure of Residents, Consumption Forecast

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. 研究背景

在畅通国内国际双循环的背景下,准确预测居民消费现状和发展趋势,适时调整制定合理的消费战略和政策,促进扩大内需是做大做强国内大循环的关键。居民人均消费性支出作为反映居民生活水平的指标,在一定程度上能够反映居民的需求情况。供需市场面临多方冲击并不是封闭的,如何在短期中预测居民消费性支出,怎样才能选择最优的预测,是弄清居民消费需求现状、促使国内供需平衡、做好国内大循环的重要前提。

2022年前三季度,河北省居民人均消费性支出为30160元,位于我国各省、直辖市、自治区以特别行政区的中等水平,因此以河北省居民消费性支出为研究对象具有一般代表性。

#### 2. 文献综述

目前,已有许多学者对居民消费支出的发展规律展开了详细的研究,根据研究方法的不同,主要可分为时间序列模型和非时间序列模型两类。

关于非时间序列模型对居民消费支出的发展规律的研究, 封永刚等[1]构建重庆市 1997~2010 年农村居民人均消费支出的灰色预测模型,采用灰色关联分析法进行分析研究,发现 GM(1,1)模型无论从精度还是有效度考虑都可以很好地预测农村居民消费支出; 夏法刚[2]利用灰色预测法对福建省 1993~2008 年农村居民人均实际消费支出进行研究,证实了 GM(1,1)模型同样适用于福建省农村居民消费情况的预测;陶晓婷[3]选取 1981~2015 年的网络数据,分别使用 Lasso、A-Lasso、E-Net、AE-Net、BP 神经网络、A-Lasso和 BP 神经网络的组合模型以及 AE-Net和 BP 神经网络的组合模型对城镇居民消费支出进行预测,并比较他们之间的预测效果,研究结果显示组合模型的预测结果比四种单一的 Lasso类方法和 BP 神经网络的预测结果更接近真实值,对比两个组合模型,AE-Net和 BP 神经网络的组合模型的预测结果更胜一筹;郭倩和王效俐[4]以 1995~2015 年的中国农村居民人均生活消费统计数据为研究对象,采用 Matlab 软件构建了基于 BP 网络的生活消费预测模型,通过预测误差分析可知,该模型可较为准确的预测生活消费趋势。

通过对这些文献的梳理,可以发现学者们在对消费支出的预测中,无论是选择灰色预测还是 Lasso 类方法,亦或是 BP 神经网络,虽然其预测结果都较为理想,但这类非时间序列模型建模过程较为复杂,预测结果的时效性不太理想。然而,时间序列模型凭借其通俗易懂、建模轻松、中短期预测卓有效果的特点在研究此类问题中占有一定优势。

关于时间序列模型对居民消费支出的发展规律的研究,李洪雄和汪浩瀚[5]选取 1978~2007 年的 GDP、CPI 以及居民消费支出为样本,分别构建了向量自回归模型和向量误差修正模型对生产总值和居民消费支出进行预测,结果显示 VECM 模型的预测值比 VAR 模型的预测值更接近实际值;曹飞[6]选取 1980~2007年云南农村居民人均消费支出为样本,通过构建自回归移动平均模型对云南农村居民消费支出进行预测,综合比较 ARIMA(3,1,0)模型和 ARIMA(3,1,1)模型,发现 ARIMA(3,1,0)模型的预测结果更为理想;王妍婷[7]构建 ARIMA 模型对湖北省 1980~2014年城镇居民人均消费支出进行研究,综合比较 ARIMA(1,1,1)模型、ARIMA(1,1,0)模型和 ARIMA(0,1,1)模型,发现 ARIMA(1,1,1)模型的拟合优度最高、AIC 值最低,说明 ARIMA(1,1,1)模型对湖北城镇居民消费支出的预测效果更理想;石捡情和杨世娟[8]选取 2009年 1月到 2017年 4月的每月居民消费价格指数为研究对象,构建 ARIMA 模型进行研究,发现 ARIMA(4,1,1)模型的预测效果理想;刘忠广和刘德欣[9]以 1980~2014年河南省城镇居民人均消费支出为实验组构建ARIMA 疏系数模型进行预测,以 2015~2018年的数据为对照组检验模型预测效果,研究表明ARIMA(4,2,1)模型能够较好地预测河南城镇居民的消费趋势。

综上所述,已经有较多学者利用包括时间序列模型在内的各类模型对居民消费进行了预测研究,但考虑季节因素的研究微乎其微,对河北省居民消费支出的研究也寥寥无几。因此,本文在前人研究的基础上,利用 SARIMA 模型对河北省居民人均消费性支出进行短期预测分析。

## 3. 理论方法介绍

#### 3.1. 单位根检验

检验序列的平稳性是构建 ARIMA 模型的必要环节,序列平稳可以避免伪回归的出现,模型的构建才有意义。一般可以利用时序图、自相关图和偏自相关图对序列的平稳性进行初步判断,也可以运用 ADF 检验进一步判断序列的平稳性[10]。主要包含三种类型的模型:

无漂移项无趋势项模型:

$$\Delta x_{t} = \delta x_{t-1} + \sum_{i=1}^{m} \beta_{i} \Delta x_{t-1} + \varepsilon_{t}$$

有漂移项无趋势项模型:

$$\Delta x_{t} = \alpha + \delta x_{t-1} + \sum_{i=1}^{m} \beta_{i} \Delta x_{t-1} + \varepsilon_{t}$$

有漂移项有趋势项模型:

$$\Delta x_{t} = \alpha + \beta t + \delta x_{t-1} + \sum_{i=1}^{m} \beta_{i} \Delta x_{t-1} + \varepsilon_{t}$$

原假设为  $H_0$ :  $\delta = 0$ ,表示序列至少存在一个单位根,当 ADF 检验的统计值小于临界值或其对应的 P 值小于显著性水平,则拒绝原假设,认为所检验序列平稳。

#### 3.2. 季节性差分自回归移动平均(SARIMA)

求和自回归移动平均模型是差分运算与 ARMA 模型的结合物,用来拟合初始非平稳经过差分后呈现平稳特征的数据,简记为 ARIMA(p,d,q),其中 p 和 q 的含义与 ARMA 模型一致,d 表示序列差分的阶数。 ARIMA 模型的具体形式为:

$$\left(1 - \sum_{i=1}^{p} \varphi_i B^i\right) \left(1 - B\right)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t$$

其中, X, 为原始序列, B 为滞后算子,  $\varepsilon$ , 为白噪声序列。

SARIMA 模型是对 ARIMA 模型的扩展,它在 ARIMA 模型的基础上提取季节效应,本文采用的是

SARIMA 加法模型,故主要对加法模型进行介绍。

SARIMA 加法模型的基本原理是将序列的总效应分解为长期趋势效应、循环波动效应、季节性变化效应和随机波动效应,四者以累加的形式反映模型的总体趋势,即  $x_t = T_t + C_t + S_t + I_t$ ,其中,  $T_t$ 为长期趋势效应、  $C_t$ 为循环波动效应、  $S_t$ 为季节性变化趋势、  $I_t$ 为随机波动趋势。若原序列趋势性和周期性并存,可记为 SARIMA  $(p,d,q)(P,D,Q)^s$ 模型,其中 p、 d、 q 的含义与 ARIMA 一样, P、 Q 分别表示季节的自回归阶数和移动平均阶数, D 为代表季节差分阶数。 SARIMA 模型的具体形式为:

$$\varphi_{p}(B)\Phi_{p}(B^{s})(1-B)^{d}(1-B^{s})^{D}X_{t}=\theta_{q}(B)\Theta_{Q}(B^{s})\varepsilon_{t}$$

式中, $(1-B^s)^D$ 为季节性差分。

## 3.3. 参数估计方法

本文采用极大似然估计法进行估计:

$$L(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k; x_1, x_2, \dots, x_n) = \max \left\{ p(x_1, x_2, \dots, x_n); \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k \right\}$$

一般情况下时间序列所服从的正态分布是未知的,因而首先假定序列服从多元正态分布:

$$x_t = \phi_1 x_{t-1} + \dots + \phi_n x_{t-n} + \varepsilon_t - \theta_1 \varepsilon_t - \dots - \theta_a \varepsilon_{t-a}$$

其中, 
$$\overline{x} = (x_1, x_2, \dots, x_n)$$
,  $\overline{\beta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ ,  $\sum n = (\tilde{x}'\tilde{x}) = \Omega \sigma_{\varepsilon}^2$ 。

对数似然函数为:

$$\ln\left(\tilde{x},\tilde{\beta}\right) = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma_{\varepsilon}^{2}\right) - \frac{n}{2}\left|\Omega\right| - \frac{n}{2\sigma_{\varepsilon}^{2}}\ln\left[\tilde{x}'\Omega\tilde{x}\right]$$

求偏导,得到方程组:

$$\begin{cases} \frac{\partial}{\partial \sigma_{\varepsilon}^{2}} \ln\left(\tilde{x}, \tilde{\beta}\right) = -\frac{n}{2\sigma_{\varepsilon}^{2}} + \frac{s\left(\tilde{\beta}\right)}{2\sigma_{\varepsilon}^{4}} = 0\\ \frac{\partial}{\partial \tilde{\beta}} \ln\left(\tilde{x}, \tilde{\beta}\right) = -\frac{1}{2} \frac{\partial \ln\left|\Omega\right|}{\partial \tilde{\beta}} + \frac{1}{\sigma_{\varepsilon}^{4}} \frac{\partial s\left(\tilde{\beta}\right)}{2\partial \tilde{\beta}} = 0 \end{cases}$$

式中,  $s(\tilde{\beta}) = \tilde{x}'\Omega^{-1}\tilde{x}$ 。

#### 3.4. 模型检验

#### 3.4.1. 模型显著性检验

检验残差序列是否为白噪声。

H0:  $\rho_1 = \cdots = \rho_m = 0$ , m > 1; H1: 至少存在某个  $\rho_k \neq 0$ ,  $1 \le k \le m$ 。 检验统计量为:

$$LB = n(n+2)\sum_{k=1}^{m} \frac{\widehat{\rho_k^2}}{n-k} \sim x^2(m)$$

若残差序列非白噪声,说明有效信息未被充分提取。

#### 3.4.2. 参数显著性检验

显著性检验的主要目的是检验模型中的参数是否显著非零,假设条件:

H0: 
$$\beta_i = 0$$
,  $1 \le j \le m$ ; H1:  $\beta_i \ne 0$ ,  $1 \le j \le m$ 

检验统计量为:

$$T = \sqrt{n - m} \frac{\widehat{\beta}_{j} - \beta_{j}}{\sqrt{a_{jj} Q(\hat{\beta})}} \sim t(n - m)$$

# 4. 数据来源和处理

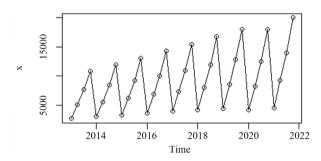
本文以河北省居民消费性支出为研究对象,数据选用河北省居民消费性支出水平 2013 年第一季度至 2022 年第三季度的季度数据,用变量  $X_i$  表示,其中 2013~2021 年的数据用于模型构建,2022 年第一季度至第三季度的数据用于检验模型的预测能力。数据全部来源于国家统计局网。

# 5. 河北省居民消费性支出水平实证分析

## 5.1. 平稳性检验

#### 5.1.1. 时序图检验

绘制河北省居民消费性支出 2013 年第一季度至 2021 年第四季度的季度数据的时序图,通过时序图可以发现河北省居民消费性支出数据序列具有明显的线性递增趋势和季节性波动趋势,为典型的非平稳序列。从图 1 可以看出河北省居民消费性支出在每一年的第一季度达到谷底,在每一年的第四季度达到峰顶,且该序列每个季节的振幅相对稳定,根据该序列的季节性特征,本文以一年为季节波动周期,并选择加法函数构建因素分解模型。



**Figure 1.** Sequence diagram of consumer expenditure of residents in Hebei Province

图 1. 河北省居民消费性支出序列时序图

对序列进行 1 阶 4 步差分后  $\nabla X_i$  的时序图如图 2 所示,发现序列仍存在较为明显的周期性,初步判断一阶差分后的序列非平稳。

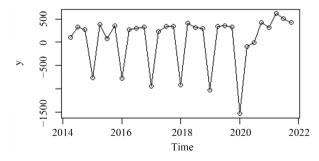


Figure 2. First difference time sequence diagram of consumer expenditure sequence of residents in Hebei Province 图 2. 河北省居民消费性支出序列一阶差分时序图

#### 5.1.2. 单位根检验

通过时序图初步判断原序列和一阶差分后的序列均是非平稳的,但通过时序图判断序列是否平稳具有一定的主观因素,为了更加客观地判断序列的平稳性,现利用 ADF 检验对 1 阶 4 步差分后的序列进行平稳性检验,检验结果如表 1。该表显示,无截距无趋势、有截距无趋势、有截距有趋势的三种情况的单位根检验的 P 值均小于显著性水平 0.05,接受序列中不含单位根的备择假设,认为一阶差分后的河北省居民消费性支出序列平稳。

**Table 1.** ADF value of differential sequence of consumer expenditure of residents in Hebei Province 表 1. 河北省居民消费性支出差分后序列的 ADF 值

无截距无趋势			有截距无趋势				有截距有趋势		
Lag	ADF	p.value	Lag	ADF	p.value	Lag	ADF	p.value	
0	-6.206	0.010	0	-6.150	0.010	0	-6.114	0.010	
1	-4.656	0.010	1	-4.610	0.010	1	-4.639	0.010	
2	-5.199	0.010	2	-5.130	0.010	2	-5.226	0.010	
	原假设: 序列非平稳 备择假设: 序列平稳								

#### 5.2. 随机性检验

根据上文的平稳性检验结果,原序列为非平稳序列,一阶差分后序列平稳,符合 SARIMA 模型序列 平稳的要求。序列平稳以后还需检验其差分后是否为随机序列,因此对差分后序列进行随机性检验,检验结果如表 2 所示。

 Table 2. Results of randomicity test

表 2. 随机性检验结果

	河北省居民消费性支出		序列	
指标名称	X-squared	p-value	指标名称	
延迟6阶	25.162	0.0003186	延迟6阶	
延迟 12 阶	57.554	6.278e-08	延迟 12 阶	

由随机性检验结果可知,河北省居民消费性支出差分后序列的 P 值小于显著性水平 0.05,说明一阶差分后的序列在 5%的显著性水平下为非纯随机序列,即差分后的序列为平稳非白噪声序列,具有研究意义,因此可以对其进行进一步的拟合研究。

#### 5.3. 模型选择和优化

根据前文的平稳性检验,可确定 SARIMA(P,D,Q)(p,d,q)模型中的 D=1,d=1,接下来通过分析河北省居民消费性支出差分后序列自相关系数和偏自相关系数来识别可能的 P、Q、p、q 值。

差分后序列的自相关图和偏自相关图如图 3 所示,自相关系数周期上 3 阶截尾或拖尾,周期内既不拖尾也不截尾;偏自相关系数周期上 1 阶截尾,周期内拖尾。由于可尝试拟合模型选择性较多,在此仅选择 P=1、Q=0或 3,p、q 的选择从低阶开始,逐渐升阶,对比各模型的 AIC 值,以确定最优模型。现尝试拟合模型 SARIMA  $(1,1,0)(0,1,1)_4$ 、 SARIMA  $(1,1,0)(1,1,0)_4$ 、 SARIMA  $(1,1,0)(1,1,1)_4$ 、

 $\begin{aligned} & \text{SARIMA} \big(1,1,3\big) \big(0,1,1\big)_4 \; \cdot \; \; & \text{SARIMA} \big(1,1,3\big) \big(1,1,0\big)_4 \; \cdot \; \; & \text{SARIMA} \big(1,1,3\big) \big(1,1,1\big)_4 \; \cdot \; \; & \text{SARIMA} \big(0,1,3\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,3\big) \big(1,1,0\big)_4 \; \cdot \; \; & \text{SARIMA} \big(0,1,3\big) \big(1,1,1\big)_4 \; \cdot \; \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(1,1,0\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \; & \text{SARIMA} \big(0,1,0\big) \big(0,1,1\big)_4 \; \cdot \\ & \text{SARIMA} \big(0,1,0\big)_4 \; \cdot \\ &$ 

SARIMA(0,1,0)(1,1,1)<sub>4</sub>。除 SARIMA(0,1,0)(1,1,1)<sub>4</sub> 模型含非平稳季节性的 AR 部分,其余每个模型对应的 AIC 值如表 3。

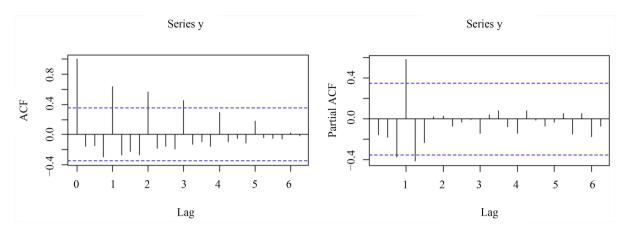


Figure 3. Sequential autocorrelation diagram and partial autocorrelation diagram after difference of consumer expenditure of residents in Hebei Province

图 3. 河北省居民消费性支出差分后序列自相关图和偏自相关图

Table 3. AIC value corresponding to SARIMA model 表 3. SARIMA 模型对应的 AIC 值

模型	AIC 值
$SARIMA(1,1,0)(0,1,1)_{4}$	472.02
$SARIMA(1,1,0)(1,1,0)_4$	463.52
$SARIMA(1,1,0)(1,1,1)_4$	458.62
$SARIMA(1,1,3)(0,1,1)_4$	472.17
$SARIMA(1,1,3)(1,1,0)_4$	464.37
$SARIMA(1,1,3)(1,1,1)_4$	456.71
$SARIMA(0,1,3)(0,1,1)_4$	470.64
$SARIMA(0,1,3)(1,1,0)_{4}$	466.42
$SARIMA(0,1,3)(1,1,1)_4$	461.79
$SARIMA(0,1,0)(0,1,1)_4$	470.87
$SARIMA(0,1,0)(1,1,0)_{4}$	464.70
SARIMA $(0,1,0)(1,1,1)_4$	含非平稳季节性的 AR 部分

从表 3 中可以发现, $SARIMA(1,1,3)(1,1,1)_4$ 模型的 AIC 值最小,故该模型能够较好地拟合河北省居民消费性支出序列。

# 5.4. 参数估计

 $SARIMA(1,1,3)(1,1,1)_4$ 模型参数拟合结果如表 4。在 5%的显著性水平下,大多参数为显著的,仅 MA(2)和 MA(3)未通过显著性检验。

变量	系数	标准差	t 统计量	P值
AR(1)	0.5733	0.2265	2.5311	0.0114
MA(1)	-0.4580	0.2183	-2.0980	0.0263
MA(2)	-0.2350	0.2554	-0.9201	0.3575
MA(3)	-0.2800	0.2129	-1.3152	0.1884
SAR(1)	0.9999	0.0006	1666.5000	0.0000
SMA(1)	-0.9732	0.1052	-9.2510	0.0000

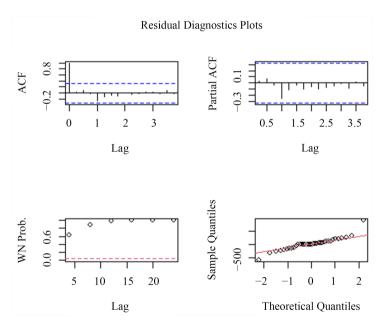
Table 4. AIC value corresponding to SARIMA model 表 4. SARIMA 模型对应的 AIC 值

模型展开式可写为:

$$(1 - 0.5733B)(1 - 0.9999B^4)(1 - B)(1 - B^4)X_t = (1 + 0.4580B + 0.2350B^2 + 0.2800B^3)(1 + 0.9732B^4)\varepsilon_t$$

## 5.5. 模型检验

图 4 中左下角为残差序列白噪声检验图,可以看到该序列所有延迟阶数的纯随机性检验的 P 值都在 0.05 显著性参考线之上,故可以认为该模型显著成立; 图 4 中右下角为残差序列正态性检验 QQ 图,可以发现正态分布的分位数与样本分位数构成的点大致密集地分布在对角线左右,可认为该序列近似服从正态分布。



**Figure 4.** Significance test of differential post-series fitting model of consumer expenditure in Hebei province

#### 图 4. 河北省居民消费性支出差分后序列拟合模型显著性检验

#### 5.6. 序列预测

利用得到的最优模型 SARIMA(1,1,3)(1,1,1)<sub>4</sub>模型预测 2022~2023 年的河北省居民消费性支出,可以得到预测支出如表 5,其中利用 2022 年第一季度至第三季度的消费性支出进行拟合预测,预测值与实际值的走势如图 5,其中相对误差为预测值和实际值的差与实际值之比。

<b>Table 5.</b> Forecast results of consumer expenditure of residents in Hebei Province	:
表 5. 河北省居民消费性支出预测结果	

时间	预测值	实际值	相对误差(%)
2022 年第 1 季度	5565.22	5474	1.67
2022 年第 2 季度	10286.18	9743	5.58
2022 年第 3 季度	15075.03	14943	0.88
2022 年第 4 季度	21253.64		
2023 年第 1 季度	5987.14		
2023 年第 2 季度	10919.86		
2023 年第 3 季度	15934.44		
2023 年第 4 季度	22421.63		

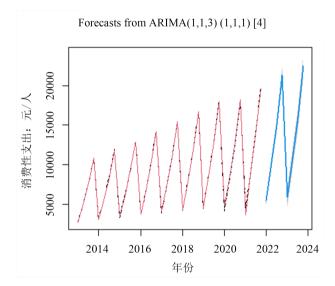


Figure 5. Effect Chart of Residential Consumption Expenditure Forecast in Hebei Province
图 5. 河北省居民消费性支出模型预测效果图

由表 5 和图 5 可以看出,SARIMA(1,1,3)(1,1,1)<sub>4</sub> 模型拟合 2022 年第 1 季度至第 3 季度数据的相对误差大多较小,只有第 2 季度的误差超过了 5%,拟合效果欠佳,其余季度的误差均控制在 2%以内,2022 年前三个季度的平均误差为 2.71%,预测值与真实值基本吻合,拟合效果较好,因此可以证明该模型用于拟合河北省居民消费性支出的合理性。图 5 所示的居民消费性支出预测效果图显示,预测值的走势与原序列基本一致,说明 SARIMA(1,1,3)(1,1,1)<sub>4</sub> 模型拟合效果较优。

# 6. 总结与展望

本文以 2013 年第 1 季度~2022 年第 3 季度河北省居民消费性支出序列作为分析,通过时序图发现该序列具有明显的周期性和上升趋势,对该序列进行 1 阶 4 步差分后,序列趋于平稳。依据自相关图和偏自相关图,初步确定预拟合模型,对比这些模型的 AIC 值,最终选择预测效果最优的 SARIMA (1,1,3)(1,1,1)<sub>4</sub>模型。继而用该模型对 2022 年前 3 个季度进行预测,将预测结果对比 2022 年第 1 季度、第 2 季度和第 3 季度的河北省居民消费性支出数据(分别为 5474、9743、14943),发现最大相对误差为 5.58%,最小相对误差为 0.88%,平均预测误差为 2.71%,表明该模型对河北省居民消费性支出有较好的预测效果。

本文在对居民消费性支出的预测中,采用的是河北省 2013 年~2022 年的季度数据,预测结果对河北省更具参考价值。但居民消费性支出发展趋势的特征在各省、直辖市、自治区以特别行政区具有一定共性,因此针对各个地域的具体数据,采用合适的模型进行预测,促使"供"、"需"联动,更好发挥消费的需求效应对经济增长的拉动作用,助力国内大循环向更加强劲可持续的方向发展。

# 参考文献

- [1] 封永刚, 郭薇, 张俊亮. 重庆农村居民消费支出的灰色预测及关联度分析[J]. 重庆理工大学学报(社会科学), 2012, 26(11): 40-46.
- [2] 夏法刚, 刘飞翔, 肖锦元. 基于灰色预测法的福建农村居民人均消费研究[J]. 科技和产业, 2010, 10(11): 107-113.
- [3] 陶晓婷. Lasso 类方法与 BP 神经网络在城镇居民消费支出预测中的应用[D]: [硕士学位论文]. 桂林: 桂林理工大学, 2017.
- [4] 郭倩, 王效俐. 基于 BP 神经网络的农村人均生活消费预测[J]. 商业经济, 2018(2): 80-83.
- [5] 李洪雄, 汪浩瀚. 向量自回归模型与向量误差修正模型预测功能的比较——基于我国国内生产总值和居民消费支出变量的实证研究[J]. 宁波大学学报(理工版), 2011, 24(2): 119-123.
- [6] 曹飞. ARIMA 模型在云南省农村居民人均消费预测中的应用[J]. 安徽农业科学, 2009, 37(30): 14923-14925.
- [7] 王妍婷. ARIMA 模型在湖北省城镇居民人均消费预测中的应用[J]. 湖北工业职业技术学院学报, 2017, 30(3): 67-70.
- [8] 石捡情, 杨世娟. 我国居民消费价格指数时间序列预测——基于 ARIMA 模型的分析[J]. 科技资讯, 2017, 15(33): 35-36+40.
- [9] 刘忠广, 刘德欣. 河南省城镇居民人均消费支出预测[J]. 北方经贸, 2020(3): 31-33.
- [10] 王燕. 时间序列分析——基于 R(第 2 版) [M]. 北京: 中国人民大学出版社, 2020.