

基于SARIMA模型的山西省居民消费价格指数分析与预测

郑圆梦*, 孔舒, 孔梦瑶, 赵一帆, 周旭婧

河南科技大学数学与统计学院, 河南 洛阳

收稿日期: 2023年5月14日; 录用日期: 2023年6月4日; 发布日期: 2023年6月16日

摘要

居民消费价格指数(CPI)是一项反映城乡居民所购买生活消费品和服务项目的价格水平变动情况的宏观经济指标。本文选取山西省2016年1月至2022年12月的CPI月度数据, 利用R软件, 建立SARIMA模型进行时间序列分析。经检验, $ARIMA(0,1,1) \times (0,1,1)_{12}$ 模型的拟合效果较好。用该模型预测山西省未来一年的CPI, 发现数据走势平稳。

关键词

消费价格指数, 山西省, SARIMA模型, R软件

Analysis and Forecasting for Consumer Price Index of Shanxi Province Based on the SARIMA Model

Yuanmeng Zheng*, Shu Kong, Mengyao Kong, Yifan Zhao, Xujing Zhou

School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan

Received: May 14th, 2023; accepted: Jun. 4th, 2023; published: Jun. 16th, 2023

Abstract

The Consumer Price Index (CPI) is a macroeconomic index that reflects the price level changes of consumer goods and services purchased by urban and rural residents. In this paper, the monthly CPI data of Shanxi Province from January 2016 to December 2022 were selected, and the SARIMA

*通讯作者。

文章引用: 郑圆梦, 孔舒, 孔梦瑶, 赵一帆, 周旭婧. 基于 SARIMA 模型的山西省居民消费价格指数分析与预测[J]. 统计学与应用, 2023, 12(3): 697-705. DOI: 10.12677/sa.2023.123074

model was established by using R software for time series analysis. It is tested that the fitting effect of $ARIMA(0,1,1) \times (0,1,1)_{12}$ model is good. Using this model to forecast the CPI of Shanxi Province in the next year, it is found that the data trend is stable.

Keywords

Consumer Price Index, Shanxi Province, SARIMA Model, R Software

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

居民消费价格指数(简称 CPI)是反映一定时期内城乡居民所购买生活消费品和服务项目的价格变动趋势和程度的相对数,用来反映居民家庭购买消费商品及服务的价格水平的变动情况。CPI 是度量通货膨胀的一个重要指标, CPI 的高低可以在一定水平上说明通货膨胀的严重程度。当 CPI 的增长幅度超过 3%时,既可以认为发生了通货膨胀;而当 CPI 的增长幅度超过 5%时,认为发生了非常严重的通货膨胀。

我国学者基于时间序列模型对 CPI 进行的研究不在少数。袁芳(2013) [1]基于乘积季节模型对陕西省 2001 年 1 月至 2012 年 12 月 144 个月份的 CPI 数据进行实证分析,并利用该模型对陕西省 2013 年各月 CPI 进行预测。孙颖(2016 年) [2]根据我国 2010 年 1 月至 2015 年 6 月 CPI 月度数据建立 ARIMA 模型,对 2015 年下半年我国的 CPI 数据进行预测。石捡情和杨世娟(2017 年) [3]选取了 2009 年 1 月至 2017 年 4 月 CPI 时间序列数据,采用时间序列分析的方法,对数据进行处理及 ADF 检验等方法分析 CPI 序列的特征,选择 ARIMA 模型对我国的 CPI 数据进行建模预测,得出 2017 年 5 月 CPI 指数为 101.8%,预计通货膨胀温和。倪颖和年靖宇(2018 年) [4]选择 2012 年 1 月至 2017 年 2 月的月度同比数据,建立 ARIMA 模型对未来时期重庆市城市居民消费价格指数进行预测。袁志强和陈锐(2018 年) [5]利用 2012 年 1 月到 2017 年 8 月月度全国居民消费价格指数,探索运用 R 软件 forecast 程序包中的 auto.arima 函数进行了最优 ARIMA 模型构建、模型检验、及短期预测。陈希睿(2019) [6]介绍了 CPI 的几种数据的定义,分析了各自的优劣,并从中选取了 CPI 同比和 CPI 环比的数据做了季节项和趋势项的分析以及 ARIMA 模型的建模与预测,得到了良好的拟合和预测结果。侯甜甜和杨淑宁(2020 年) [7]利用 R 软件对 2001 年至 2017 年我国 CPI 的月度数据进行处理分析,通过建立 ARIMA 模型对 CPI 的未来走势进行预测。毛钰嘉(2020) [8]采用了数学中经典的最小二乘法与 MAE 权重系数法进行组合,提出组合权重系数法建立组合模型,并对 2019 年 4 月到 2020 年 1 月的江西省 CPI 序列值进行预测。蔡敏(2022) [9]以江苏省月度 CPI 同比数据为研究对象,基于 SARIMA 和 BP 神经网络建立不同模型进行比较并选取最优模型,最后,利用最优模型对 2022 年 1 月至 6 月的江苏省 CPI 数据进行了预测。王琳惠(2022) [10]为了对 CPI 做出精准的预测,将生产价格指数(PPI)作为外生变量,使用带外生变量的季节性差分自回归移动平均模型,并将 SARIMAX 模型与基于集成经验模式分解(EEMD)下的长短时记忆网络(LSTM)模型混合,提出 SARIAMX-EEMD-LSTM 混合模型来进行 CPI 的预测。

山西省是我国重要的能源基地,对其 CPI 数据进行时间序列分析具有重要的实际价值。对于非平稳序列,可采用时间序列模型对其分析,ARIMA 模型具有很好的短期预测效果。经观察发现,CPI 数据序列具有明显周期性,综上建立季节性 ARIMA 模型对山西省 CPI 拟合和短期预测。本文选取山西省 2016

年 1 月至 2022 年 12 月的 CPI 月度数据, 利用 R 软件, 建立 SARIMA 模型进行分析与预测, 并给出相关政策建议。

2. 理论基础

2.1. 指标选取

居民消费价格指数是对城市居民消费价格指数和农村居民消费价格指数进行综合汇总计算的结果。通过该指数可以观察和分析消费品的零售价格和服务项目价格变动对城乡居民实际生活费支出的影响程度。其中计算公式为:

$$\text{CPI} = \frac{P_1}{P_2} \times 100\%$$

其中 P_1 : 一组固定商品按当期价格计算的价值; P_2 : 一组固定商品按基期价格计算的价值。

2.2. SARIMA 模型

2.2.1. 模型形式

SARIMA 模型的构造原理: 当序列具有短期相关性时, 通常可以使用低阶 ARIMA(p,q)模型提取; 当序列具有季节效应, 并且季节效应本身还具有相关性时, 这种季节相关性可使用以周期步长为单位的 ARIMA(P,D,Q)_s 模型提取。由于短期相关性和季节效应之间具有乘积关系, 所以拟合模型实际上为 ARIMA(p,d,q)×(P,D,Q)_s, 考拟合的乘积模型如下[11]:

$$\nabla^d \nabla_s^D x_t = \frac{\Theta(B)\Theta_s(B)}{\Phi(B)\Phi_s(B)} \varepsilon_t$$

式中:

$$\Theta_s(B) = 1 - \theta_1 B^S - \dots - \theta_Q B^{QS}, \quad \Phi_s(B) = 1 - \phi_1 B^S - \dots - \phi_P B^{PS}$$

在 ARIMA(p,d,q)×(P,D,Q)_s 中, AR 表示自回归; p 为自回归项数; MA 表示移动平均; d 为将时间序列变为平稳序列所需要做的差分次数, q 是移动平均项数[12] [13]。

2.2.2. 模型的显著性检验

原假设和备择假设分别为:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0 \quad \text{vs} \quad H_1: \text{至少存在某个 } \rho_k \neq 0 \quad \forall m \geq 1, k \leq m$$

模型检验统计量为 $LB(Ljung - Box)$:

$$LB = n(n+2) \sum \frac{\tilde{\rho}_k}{n-k} \sim \chi^2(m) \quad \forall m > 0$$

模型的显著性检验即对模型残差序列进行白噪声检验。检验时, 如果在各阶延迟下, 白噪声检验统计量的 p 值大多数都近似等于或小于 0.05, 可以认为这个拟合模型的残差序列属于非白噪声序列, 即该拟合模型显著; 反之, 该拟合模型不显著, 舍弃该模型。

3. 实证分析

3.1. 数据来源

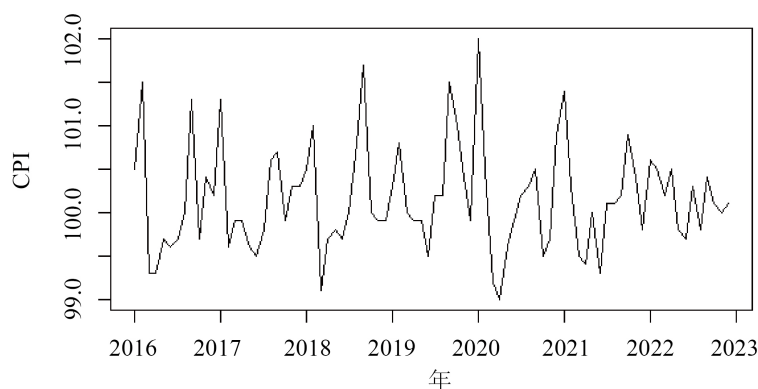
本文选取 2016 年 1 月~2022 年 12 月山西省 CPI (上月 = 100) 月度数据, 以下数据均来源于国家统计局, 如表 1 所示。

Table 1. Monthly CPI data of Shanxi Province from January 2016 to December 2022**表 1.** 2016 年 1 月~2022 年 12 月山西省 CPI 月度数据

年月	2016	2017	2018	2019	2020	2021	2022
1 月	100.5	101.3	100.5	100.3	102.0	101.4	100.6
2 月	101.5	99.6	101	100.8	100.5	100.3	100.5
3 月	99.3	99.9	99.1	100	99.2	99.5	100.2
4 月	99.3	99.9	99.7	99.9	99.0	99.4	100.5
5 月	99.7	99.6	99.8	99.9	99.6	100.0	99.8
6 月	99.6	99.5	99.7	99.5	99.9	99.3	99.7
7 月	99.7	99.8	100	100.2	100.2	100.1	100.3
8 月	100.0	100.6	100.8	100.2	100.3	100.1	99.8
9 月	101.3	100.7	101.7	101.5	100.5	100.2	100.4
10 月	99.7	99.9	100.0	101.0	99.5	100.9	100.1
11 月	100.4	100.3	99.9	100.4	99.7	100.4	100.0
12 月	100.2	100.3	99.9	99.9	100.9	99.8	100.1

3.2. 序列平稳化处理

将山西省 2016 年~2022 年 CPI 月度数据导入 R 软件，绘制出山西省 2016 年~2022 年月度 CPI 时序图，如图 1 所示。

**Figure 1.** CPI time series graph**图 1.** CPI 时间序列图

观察图 1，时序图具有趋势效应和季节效应。该数据为月度数据，周期步长为 12。为消除序列的趋势性和周期性，对该序列进行 1 阶差分提取趋势效应，12 步步长提取季节效应，1 阶 12 步差分后时序图，如图 2 所示。

3.3. 平稳性检验

经过趋势效应提取和季节效应提取之后初步观察序列呈现平稳化，但这种判断结果带有主观性。故需借助 ADF 统计量[12]，对 1 阶 12 步差分后的序列进行平稳性检验，检验结果如表 2 所示。

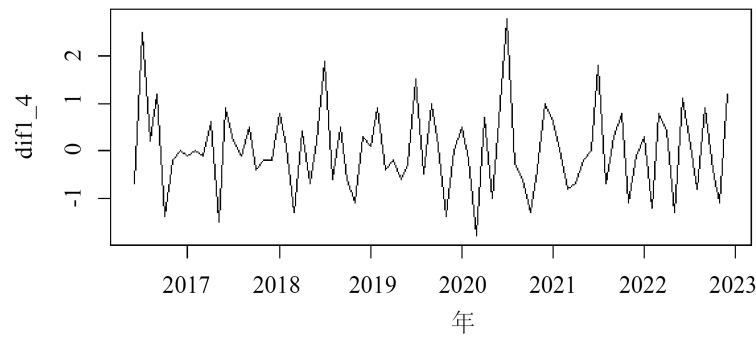


Figure 2. Monthly 1-order 12-step difference timing chart of CPI in Shanxi Province from 2016 to 2022

图 2. 山西省 2016~2022 年 CPI 月度数据 1 阶 12 步差分时序图

Table 2. Stationary test of sequence after 1 order 12 step difference

表 2. 1 阶 12 步差分后序列的平稳性检验

	无漂移项自回归结构			有漂移项自回归结构			带趋势回归结构		
	lag	ADF	p 值	lag	ADF	p 值	lag	ADF	p 值
1	0	-13.61	0.01	0	-13.51	0.01	0	-13.42	0.01
2	1	-9.90	0.01	1	-9.82	0.01	1	-9.73	0.01
3	2	-8.28	0.01	2	-8.22	0.01	2	-8.14	0.01
4	3	-7.16	0.01	3	-7.11	0.01	3	-7.05	0.01

从表 2 的检验结果可以看出，三种类型的单位根检验对应的 p 值均小于 0.05，因此拒绝原假设，说明该时序列为平稳序列。

3.4. 模型定阶

由平稳性检验可知，在即将建立的 SARIMA 模型中 $d = 1$ ，为了确定自回归项数 p 和移动平均回归项数 q ，在 R 软件绘制出 1 阶 12 步差分后序列的自相关图和偏自相关图，如图 3 所示。

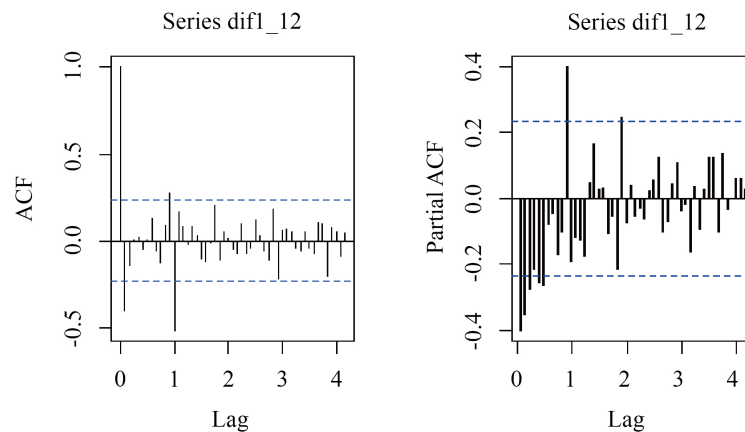


Figure 3. ACF and PACF of the sequence after 1 order 12 step difference

图 3. 1 阶 12 步后序列的自相关图和偏自相关图

观察 1 阶 12 步差分后的自相关函数和偏自相关函数, 自相关图 1 阶后拖尾, 偏自相关图 2 阶后拖尾, 故 p 的值可以为 0、1、2, q 的值可以为 0、1。初步拟合筛选, 选定参数 $p = 0$ 、 $d = 1$ 、 $q = 1$ 、 $P = 0$ 、 $D = 1$ 、 $Q = 1$ 或 $p = 0$ 、 $d = 1$ 、 $q = 1$ 、 $P = 2$ 、 $D = 0$ 、 $Q = 0$ 。最后根据最小信息准则判断选择得到的模型: $ARIMA(0,1,1) \times (0,1,1)_{12}$ 和 $ARIMA(0,1,1) \times (2,0,0)_{12}$ 。

3.5. 参数估计

模型定阶完成后, 利用 R 中 `arima` 函数对未知参数估计。模型 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 和 $ARIMA(0,1,1) \times (2,0,0)_{12}$ 的参数估计结果如下表 3 和表 4 所示。

Table 3. $ARIMA(0,1,1) \times (0,1,1)_{12}$ parameter estimation results

表 3. $ARIMA(0,1,1) \times (0,1,1)_{12}$ 的参数估计结果

	mal	smal
参数估计值	-1.0000	-0.9991
标准差	0.0844	0.3816

Table 4. $ARIMA(0,1,1) \times (2,0,0)_{12}$ parameter estimation results

表 4. $ARIMA(0,1,1) \times (2,0,0)_{12}$ 的参数估计结果

	mal	sar1	sar2
参数估计值	-1.000	0.2551	0.3812
标准差	0.023	0.0992	0.1091

3.6. 模型诊断

3.6.1. 模型的显著性检验

模型 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 和 $ARIMA(0,1,1) \times (2,0,0)_{12}$ 的残差序列的白噪声检验结果分别如下图 4、图 5 所示。

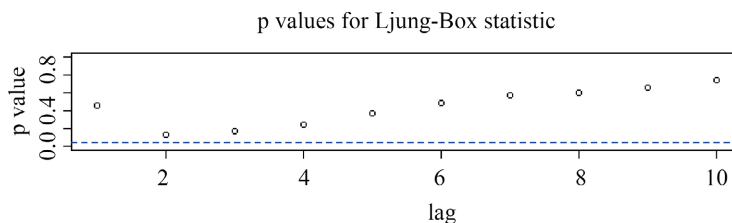


Figure 4. Diagnostics of the residuals from $ARIMA(0,1,1) \times (0,1,1)_{12}$

图 4. $ARIMA(0,1,1) \times (0,1,1)_{12}$ 的残差的诊断

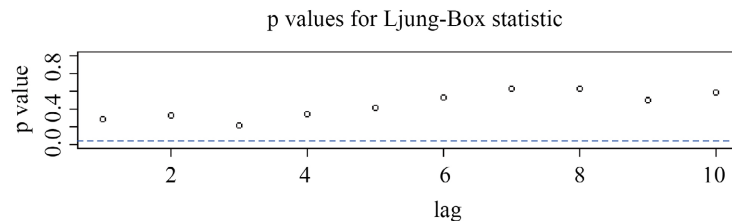


Figure 5. Diagnostics of the residuals from $ARIMA(0,1,1) \times (2,0,0)_{12}$

图 5. $ARIMA(0,1,1) \times (2,0,0)_{12}$ 的残差的诊断

从图 4、图 5 可以看出各阶延迟下模型 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 和 $ARIMA(0,1,1) \times (2,0,0)_{12}$ 残差序列的白噪声检验统计量的 p 值均大于 0.05, 接受原假设, 可以认为这两个拟合模型的残差序列属于白噪声序列, 模型拟合效果较好。

3.6.2. 参数的显著性检验

由上述的模型显著性检验筛选出的两个模型中进行参数的显著性检验, 根据参数估计的两个拟合模型中所列出来的各模型参数估计值均大于它的二倍标准差, 因此以上两个拟合模型的参数均显著非零, 不需要进行剔除。

3.7. 模型优化

经过模型的显著性检验以及参数的显著性检验最终确定的两个模型分别为 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 和 $ARIMA(0,1,1) \times (2,0,0)_{12}$, 通过 AIC 准则对两个模型做最终的选择。相关数据如下表 5 所示。

Table 5. AIC comparison of the model

表 5. 模型的 AIC 比较

	$ARIMA(0,1,1) \times (0,1,1)_{12}$	$ARIMA(0,1,1) \times (2,0,0)_{12}$
AIC	120.7	139.63

根据 AIC 准则, 选择函数值较小的那个模型作为最终的拟合模型, 由表 5 可知模型 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 的 AIC 函数值要优于模型另一个模型, 因此最终我们选择 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 作为预测模型。

拟合的 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 模型为:

$$x_t = x_{t-1} + x_{t-12} + x_{t-13} + \varepsilon_t - \varepsilon_{t-1} - 0.9991\varepsilon_{t-12} + 0.9991\varepsilon_{t-13}$$

3.8. 模型预测

运用较优的 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 模型, 对山西省 2023 年 1~3 月月度 CPI 数据进行预测, 并通过预测值与真实值做比较, 得出预测准确率。结果如表 6 所示。从预测准确率来看, 模型预测精度高, 预测效果好。

Table 6. Comparison of the predicted and true monthly CPI in Shanxi Province from January 2023 to March 2023

表 6. 2023 年 1 月~2023 年 3 月年山西省月度 CPI 预测值与真实值对比

时间	预测值	真实值	预测准确率
2023.01	100.96195	100.6	99.64150%
2023.02	100.61910	99.4	98.78840%
2023.03	99.61910	99.7	99.91879%

对山西省月度 CPI 数据走势进行预测, 预测区间为 2023 年 4~12 月, 结果如表 7 所示。

Table 7. Forecasting for monthly CPI of Shanxi Province from April to December 2019

表 7. 2019 年 4~12 月山西省月度 CPI 预测值

月份	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
预测值	99.69053	99.79053	99.61910	100.06195	100.27624	100.91909	100.17624	100.17624	100.17622

图6为山西省2023年月度CPI走势图,可以看出2023年山西省CPI时序指数走势总体较为平稳,波动不大。

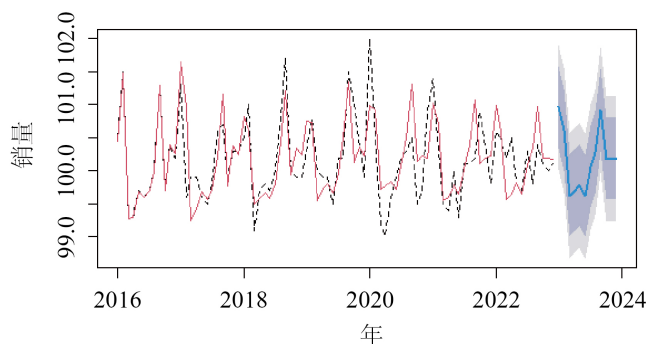


Figure 6. Twelve month forecast using the $ARIMA(0,1,1) \times (0,1,1)_{12}$ model on the dataset of CPI in Shanxi Province

图6. 使用 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 模型对山西省CPI数据的12个月预测

4. 总结

本文选取了山西省从2016年1月至2022年12月月度CPI数据,建立得到 $ARIMA(0,1,1) \times (0,1,1)_{12}$ 模型来拟合该省月度CPI数据的变动情况。接着,基于此模型对2023年1月到2023年3月山西省的月度CPI数据进行预测,预测效果良好。通过此模型预测出了山西省2023年的CPI的未来走势。

由预测结果可得,山西省未来一年内的CPI变动总体较为平稳,波动不大。该趋势表明山西省物价总体持续保持平稳,经济发展稳定。同时也进一步意味着政府现应将重点放在推动山西省经济发展上去。因此山西省有关部门应当根据预测的经济状况采取一定政策对经济进行宏观调整,既要保证现有经济稳中求进,又要防止刺激过度带来通货膨胀。

基金项目

河南科技大学SRTP基金项目(2021180)。

参考文献

- [1] 袁芳. 基于乘积季节模型的陕西省CPI的分析与预测[J]. 金融经济, 2013(12): 163-164.
- [2] 孙颖. 基于ARIMA模型的消费者价格指数预测[J]. 统计与决策, 2016(11): 83-85.
- [3] 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.
- [4] 倪颖, 年靖宇. 基于ARIMA模型的居民消费价格指数走势实证分析与预测——以重庆市为例[J]. 贵州商学院学报, 2018, 31(2): 14-23.
- [5] 袁志强, 陈锐. 基于ARMA模型的CPI短期预测研究[J]. 中国集体经济, 2018(3): 64-65.
- [6] 陈希睿. 基于时间序列分析的CPI预测[D]: [硕士学位论文]. 北京: 清华大学, 2019.
- [7] 侯甜甜, 杨淑宁. 我国消费价格指数的走势预测及影响因素分析[J]. 统计理论与实践, 2020(4): 34-41.
- [8] 毛钰嘉. 基于时间序列的组合预测模型研究——以江西省CPI为例[D]: [硕士学位论文]. 南昌: 江西财经大学, 2020.
- [9] 蔡敏. 基于SARIMA和BP神经网络模型的江苏省CPI预测分析[D]: [硕士学位论文]. 桂林: 广西师范大学, 2022.

-
- [10] 王琳惠. 基于 SARIMAX-EEMD-LSTM 模型的 CPI 指数预测[D]: [硕士学位论文]. 济南: 山东大学, 2022.
- [11] 王燕. 时间序列分析——基于 R [M]. 第 2 版. 北京: 中国人民大学出版社, 2020.
- [12] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Jung, G.M.L. (2015) Time Series Analysis: Forecasting and Control. 5th Edition, Wiley, Hoboken.
- [13] Shumway, R.H. and Stoffer, D.S. (2017) Time Series Analysis and Its Applications with R Examples. 4th Edition, Springer, New York. <https://doi.org/10.1007/978-3-319-52452-8>