

# 基于ARIMA模型研究分析我国居民消费价格水平

王晗琨

云南财经大学, 统计与数学学院, 山东 烟台

收稿日期: 2023年10月16日; 录用日期: 2023年11月28日; 发布日期: 2023年12月6日

## 摘要

居民消费价格指数(consumer price index, 简称CPI)是衡量总体价格水平的指标, 也是一个宏观经济指标, 反映日常生活中人们购买生活用品和消费服务时价格水平的趋势变化, 因此居民消费价格指数不仅同人们生活密切相关, 且在整个国民经济价格体系中也有一席之地。除此之外居民消费价格指数在分析和决策经济、调控和检测价格水平以及核算国民经济的重要指标。其变动率反映了一定程度的通货膨胀或紧缩。本文选取2001年至2020年全国居民消费价格指数数据, 建立ARIMA模型进行研究分析, 并预测未来5年居民消费价格指数的变化趋势。

## 关键词

居民消费价格指数, ARIMA模型, 国民经济, 通货膨胀, 预测

## Research and Analysis of Consumer Price Level of Chinese Residents Based on ARIMA Model

Hankun Wang

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Yantai Shandong

Received: Oct. 16<sup>th</sup>, 2023; accepted: Nov. 28<sup>th</sup>, 2023; published: Dec. 6<sup>th</sup>, 2023

## Abstract

The Consumer Price Index (CPI) is an indicator that measures the overall price level and is also a macroeconomic indicator that reflects the trend changes in price levels when people purchase

daily necessities and services in their daily lives. Therefore, the CPI is not only closely related to people's lives but also has a place in the entire national economic price system. In addition, the consumer price index of residents is an important indicator for analyzing and decision-making about the economy, regulating and testing price levels and accounting for the national economy. Its rate of change reflects a certain degree of inflation or contraction. This article selects the national consumer price index data from 2001 to 2020, establishes an ARIMA model for research and analysis, and predicts the trend of changes in the consumer price index in the next 5 years.

## Keywords

Consumer Price Index of Residents, ARIMA Model, National Economy, Inflation, 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. 引言

居民价格消费指数(CPI)与人们日常生活有着紧密的联系,不仅可以反映物价水平,还可以反映居民工资情况。居民价格消费指数还与通货膨胀有关,一般要是居民价格消费指数数值增长代表着商品价格上涨,随之而来的影响可能有通货膨胀、货币贬值等,反而对居民日常水平受到一定波动影响。除此之外,居民价格消费指数反映经济增长数量,对国民经济发展有一定影响。因此,为了更好地分析我国居民价格消费指数的发展趋势以及预测未来5年居民价格消费指数的发展趋势,需要建立模型进行分析研究。

### 1.1. 研究背景

根据国家统计局发布的数据来看,我国居民价格消费指数近几年不断波动发展,其发展已经影响到居民正常生活,那居民价格消费指数波动现象会如何影响未来发展?国家政府人们又将如何应对这个局面?未来几年居民价格消费指数的趋势又会如何变化?

居民消费价格指数衡量居民消费的商品和服务的价格随时间的变化趋势,每年都统计,且其变化程度可以反映出通货膨胀或紧缩的程度。一般来说当CPI变化率高于3%时,称为通货膨胀;当CPI变化率高于5%,称为严重的通货膨胀。如果CPI变化率过高,通货膨胀过于严重,会造成经济发展不稳定,带来其他风险。也就是说,CPI增长过高,生活成本过高随之而来的是货币价值下降,一张100的纸币只能买到价值低于100的商品和服务。除此之外CPI还可以衡量国内生产总值、资产、负债、消费以及收入等实际价值。

我国居民价格消费指数由国家统计局编制发布,各省、自治区、直辖市的CPI由国家统计局认定反馈,各调查总队公开。确定居民消费价格指数首先要被调查的商品和服务项目。因此我国统计部门选取一些能代表居民消费、对居民生活影响较大的商品和服务,并把这些固定数量的商品和服务统称为“商品篮子”。我国把CPI分为八大类:医疗保健、衣着、生活用品及服务、交通和通信、食品烟酒、居住、教育文化和娱乐、其他用品和服务,262个基本分类(功能、性质、结构相同或相近的产品)。

$CPI = (\text{一组固定商品当期价值} / \text{一组固定商品基期价值}) \times 100\%$ 。也就是说CPI反映出来的是当居民消费一组商品时,在当前时间点比以往某一时间点多支出多少。但是在日常生活中,更关注的是通货膨胀的程度,也就是一定时期内商品价格增长程度,通过通货膨胀率来反映,用CPI来计算。通货膨胀率 =

(报告期 CPI - 基期 CPI) ÷ 基期 CPI × 100%。所以说, 这个通货膨胀率就是不同时间段 CPI 变化的百分比。

时间序列分析, 是将周期、趋势、时期和不稳定因素这四个因素综合进行预测, 也需要根据要检测的对象的变化特点, 选择恰当的数学软件, 然后根据数据建立数学模型, 得出其相应的数学规律。对于本文来说, 选择什么样的数学软件, 建立哪种数学模型来分析预测居民价格消费指数的变化, 是至关重要的。统计学中, ARIMA 模型常用来进行数据分析好预测, 其短期预测的效果很好。

## 1.2. 研究目的及内容

2001 年到 2020 年, 我国居民价格消费指数变化波动较大, 尤其是 2009 年以后呈增长趋势, 也就是说我国面临通货膨胀的严峻形势, 随着居民日常生活消费成本不断增加, 分析 CPI 趋势发展也成为热门的研究话题。本文将从这个角度出发, 用 Eviews 建立 ARIMA 模型[1]对 2001 年到 2020 年居民价格消费指数进行研究分析并预测未来 5 年的发展趋势[2], 有助于我国的国民经济发展。

本文从统计学角度出发对居民价格消费指数进行研究分析, 利用软件建立 ARIMA 时间序列模型对国家统计局官网公布的 2001~2020 年居民价格消费指数数据进行分析研究, 并预测未来 5 年居民价格消费指数的发展趋势[3]。主要内容如下: ARIMA 模型的理论与方法介绍、利用 ARIMA 模型分析研究居民价格消费指数、预测未来 5 年居民价格消费指数的变化趋势[4]。

## 2. 方法介绍

### 2.1. 模型理论与方法介绍

#### 2.1.1. ARIMA 模型介绍

ARIMA 模型全称自回归移动平均模型(Autoregressive Integrated Moving Average Model), 由 Box 和 Jenkins 提出的一个时间序列预测方法, 运用数学模型来表现随机序列数据, 可以从已知的时间序列数据预测未来。

ARIMA 模型也叫差分整合移动平均自回归模型, 结合自回归模型(AR)、移动平均模型(MA)和差分法得到模型 ARIMA( $p, d, q$ ), 其中  $p$  是自回归项数,  $q$  是滑动平均项数,  $d$  是序列成为平稳序列所需要进行的差分的阶数, 也是至关重要的一步[5]。

自回归模型(AR):

AR 模型通过已知的不同时间段序列之间的相关关系, 利用回归方程来预测。也就是说用以前时间点的的数据做回归变量来预测未来时间点的的数据。

其模型表示为:

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + \varepsilon_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t \quad (1)$$

由该式子可以看出  $t$  时刻的  $X_t$  由  $X_{t-1} \cdots X_{t-p}$  的加权以及随机干扰项决定。该模型被称为 AR( $p$ ) 模型, 即  $p$  阶自回归模型, 其中  $a_1 \cdots a_p$  是常数项, 成为模型的自回归系数。

移动平均模型(MA):

MA 模型是由自回归模型中误差线性组合而成, 此误差服从正态分布且相互独立, 其通过误差的线性组合来观察其振动, 可以消除预测时的随机扰动, 也就是说, 当运用的数据受外在影响波动明显无法很好判断其发展趋势时, 可以利用此模型消除这些影响因素以此来分析, 进而预测其未来趋势发展[6]。

其模型表示为:

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (2)$$

此模型被称为MA( $q$ )模型,其描述的是 $t$ 时刻与 $t, t-1, \dots$ 时刻误差的相关关系。

自回归滑动平均模型(ARMA):

ARMA模型是由AR模型和MA模型结合而成,是时间序列模型中的一种。ARMA( $p, q$ )模型中包含了 $p$ 个自回归项和 $q$ 个移动平均项,可表示为:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (3)$$

其中 $p$ 和 $q$ 是模型的自回归阶数和移动平均阶数; $\varphi$ 和 $\theta$ 是不为零的待定系数; $\varepsilon_t$ 是独立的误差项; $X_t$ 是平稳、正态、零均值的时间序列。

自回归移动平均模型(ARIMA):

ARIMA模型是ARMA模型的拓展[7],其建立模型分析一随机的时间序列数据,通过已知的数值预测未来的时间序列数值。ARIMA模型只能分析平稳的时间序列,非平稳时间序列必须先转化为平稳时间序列后再运动ARIMA模型,因此有差分的过程。

ARIMA( $p, d, q$ )模型一般可表示为:

$$(1-L)^d X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (4)$$

简记为:

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t \quad (5)$$

其中,AR是自回归模型, $p$ 为自回归项系数;MA为移动平均模型, $q$ 为移动平均项系数模型, $d$ 为时间序列成为平稳时所做的差分次数。

### 2.1.2. 非平稳时间序列

非平稳时间序列指时间序列随时间的变化而变化。随机游走是一较特殊的非平稳时间序列,其过程定义为: $y_t = y_{t-1} + \varepsilon_t$ ;其中 $\varepsilon$ 为均值为零的误差项。另一种非平稳时间序列是在随机游走中加一个常数项,被称为带漂移项的随机游走,其定义为 $y_t = c + y_{t-1} + \varepsilon_t$ ,其中 $c$ 为常数。漂移项使得这个非平稳时间序列拥有了长期趋势。

使用ARIMA模型进行预测时,首先通过差分得到平稳时间序列[8]。差分法是一种非常简便、有效的确定性信息提取法,运算的实质是使用自回归的方式提取确定性信息,即:

$$\nabla^d x_t = (1-B)^d x_t = \sum_{i=0}^d (-1)^i C_d^i x_{t-i} \quad (6)$$

其一阶差分为: $\nabla x_t = x_t - x_{t-1}$ ;

$p$ 阶差分为: $\nabla^p x_t = \nabla^{p-1} x_t - \nabla^{p-1} x_{t-1}$ ;

$k$ 步差分为: $\nabla_k x_t = x_t - x_{t-k}$ 。

## 2.2. ARIMA 模型建立过程

模型建立分为四步:处理数据、估计参数、模型建立和模型检验

### 2.2.1. 数据处理

ARIMA模型建立的必要条件是平稳时间序列,因此首先要检查数据的平稳性[9]。平稳性检验方法如下:

### 1) 图检验方法

#### a) 时序图检验

根据平稳时间序列均值、方差为常数的性质，平稳序列的时序图的序列始终在一个常数值附近随机波动，而且波动的范围有界、无明显趋势及周期特征。

#### b) 自相关图检验

平稳序列通常具有短期相关性。该性质用自相关系数来描述就是随着延迟期数的增加，平稳序列的自相关系数会很快地衰减向零。

### 2) 单位根检验(ADF 检验)

如果该时间序列数据是平稳的，那么该时间序列的所有特征根都应该在单位圆内。

#### a) ADF 检验原理：

假设序列为： $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \xi_t$ 。

如果序列平稳，它必须满足所有非零特征根都在单位圆内。假如有一个单位根存在，不妨假设  $\lambda_1 = 1$ ，则序列非平稳。

把  $\lambda_1 = 1$  代入特征方程，得到  $1 - \phi_1 - \phi_2 - \dots - \phi_p = 0 \Rightarrow \phi_1 + \phi_2 + \dots + \phi_p = 1$ 。

这意味着，如果序列非平稳，存在特征根，那么序列回归系数之和恰好等于 1。因而，对于序列的平稳性检验，可以通过检验它的回归系数之和的性质进行判断。

#### b) 检验假设条件

令  $\rho = \phi_1 + \phi_2 + \dots + \phi_p - 1$ ，则假设条件为： $H_0: \rho = 0 \leftrightarrow H_1: \rho < 0$ 。

#### c) 检验统计量

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})} \quad (7)$$

#### d) 检验结果判定

通过蒙特卡洛方法，可以得到 ADF 检验统计量的临界值表。当显著性水平取为  $\alpha$  时，记  $\tau_\alpha$  为 ADF 检验的  $\alpha$  分位点，则当  $\tau \leq \tau_\alpha$  时，拒绝原假设，认为序列平稳；当  $\tau > \tau_\alpha$  时，接受原假设，认为序列非平稳。

### 2.2.2. 参数估计

如果所用的时间序列数据为非平稳时间序列，则根据 ADF 检验得出的结果判断其为平稳序列时在几阶差分，确定  $d$  的值。然后通过自相关图和偏自相关图确定  $p$  与  $q$  的值，初步建立模型。

### 2.2.3. 模型建立

衡量模型的参数指标有调整 R2、AIC、SC 等。调整 R2 代表着模型拟合的整体优度，其值在 0 到 1 之间且调整 R2 值越大，表示模型拟合的效果越好[10]。AIC 也叫最小信息量准则，其统计量为：

$$AIC = -2\ln(\text{模型的极大似然函数值}) + 2(\text{未知参数个数}) \quad (8)$$

SC 准则和 AIC 准则一样，值越小表示模型拟合效果越好。因此分析初步建立的几个模型，比较其调整 R2、AIC、SC 等参数，并且进行检验，根据检验结果，确定出一个最优的模型。

### 2.2.4. 模型检验

在确定模型后要先对已经拥有的数据进行预测，比较得出的预测值和官网的实际值，计算出相对误差，并且比较其误差大小。当相对误差很小时，说明该模型的拟合效果较好，并可以运用该模型对实际问题进行比较分析和对未来数据进行预测[11]。

### 3. 结果

#### 3.1. 平稳性检验

本文的数据取自我国国家统计局官网公布的 2001 年至 2020 年的居民消费价格指数的年度数据(上年同期 = 100)。由国家统计局官网公布的数据可知从 2001 年以来我国 CPI 在不断波动变化, 初步判断其为不平稳时间序列数据, 但由于该趋势图有一定局限, 不能直接认定其不平稳, 无法由此图准确得知我国居民价格消费指数是否平稳, 因此进一步进行平稳性检验。

对该序列进行 ADF 检验, 结果如图 1:

Augmented Dickey-Fuller Unit Root Test on INDEX

Null Hypothesis: INDEX has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=4)				
	t-Statistic		Prob.*	
Augmented Dickey-Fuller test statistic	0.622075		0.9864	
Test critical values:	1% level	-3.831511		
	5% level	-3.029970		
	10% level	-2.655194		
*MacKinnon (1996) one-sided p-values.				
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 19				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(INDEX)				
Method: Least Squares				
Date: 10/12/23 Time: 15:20				
Sample (adjusted): 2002 2020				
Included observations: 19 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
INDEX(-1)	0.018796	0.030215	0.622075	0.5421
C	0.487438	3.759655	0.129650	0.8984
R-squared	0.022257	Mean dependent var	2.805263	
Adjusted R-squared	-0.035258	S.D. dependent var	2.151690	
S.E. of regression	2.189293	Akaike info criterion	4.504335	
Sum squared resid	81.48109	Schwarz criterion	4.603750	
Log likelihood	-40.79119	Hannan-Quinn criter.	4.521160	
F-statistic	0.386978	Durbin-Watson stat	1.822390	
Prob(F-statistic)	0.542146			

Figure 1. ADF test results

图 1. ADF 检验结果

从图中可以知道 P 值为 0.9864, 因此该序列有单位根, 接受原假设, 即该序列是非平稳时间序列。ARIMA 模型预测的必须是平稳时间序列, 因此要对序列进行差分处理, 一阶差分结果如图 2:

Augmented Dickey-Fuller Unit Root Test on D(INDEX)

Null Hypothesis: D(INDEX) has a unit root				
Exogenous: Constant				
Lag Length: 1 (Automatic - based on SIC, maxlag=4)				
	t-Statistic		Prob.*	
Augmented Dickey-Fuller test statistic	-4.478051		0.0031	
Test critical values:	1% level	-3.886751		
	5% level	-3.052169		
	10% level	-2.666593		
*MacKinnon (1996) one-sided p-values.				
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 17				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(INDEX,2)				
Method: Least Squares				
Date: 10/12/23 Time: 15:27				
Sample (adjusted): 2004 2020				
Included observations: 17 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(INDEX(-1))	-1.406582	0.314106	-4.478051	0.0005
D(INDEX(-1),2)	0.323049	0.217322	1.486498	0.1593
C	4.273320	1.005990	4.247877	0.0008
R-squared	0.632715	Mean dependent var	0.202941	
Adjusted R-squared	0.580246	S.D. dependent var	3.000129	
S.E. of regression	1.943737	Akaike info criterion	4.325887	
Sum squared resid	52.89359	Schwarz criterion	4.472925	
Log likelihood	-33.77004	Hannan-Quinn criter.	4.340503	
F-statistic	12.05877	Durbin-Watson stat	1.864831	
Prob(F-statistic)	0.000902			

Figure 2. First order difference results

图 2. 一阶差分结果

由上图可知，一阶差分后 T 值为-4.478051，P 值为 0.0031，因此拒绝原假设，一阶差分后的序列为平稳的时间序列。

通过以上分析可知我国居民价格消费指数是一个不平稳的时间序列，一阶差分后变为平稳的时间序列数据。

### 3.2. 参数估计

通过上面的 ADF 检验和差分处理后得到一阶差分后的居民消费价格指数为平稳的时间序列，因此 ARIMA( $p, d, q$ )模型中参数  $d = 1$ ，为估计参数  $p, q$ ，分析一阶差分后的自相关图和偏自相关图，如图 3 所示。

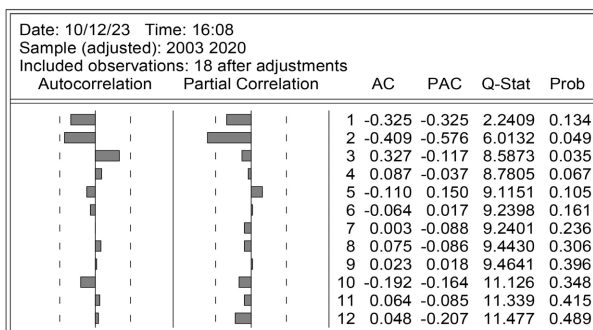


Figure 3. Autocorrelation and partial autocorrelation graphs  
图 3. 自相关图和偏自相关图

建立模型前要确认模型参数，通过一阶差分后时间序列的自相关图和偏自相关图来估计。如上图可知可选取  $p$  值  $q$  值为 1 或 2，因此可建立模型 ARIMA(1, 1, 1)、ARIMA(1, 1, 2)、ARIMA(2, 1, 1)、ARIMA(2, 1, 2)。

### 3.3. 模型建立

为了得到最优模型，需要对比不同模型的参数值，运用适当的比较准则选取参数，确定最优模型。对模型 ARIMA(1, 1, 1)、ARIMA(1, 1, 2)、ARIMA(2, 1, 1)、ARIMA(2, 1, 2)进行建模，结果如下：

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.000000	0.005448	183.5369	0.0000
MA(1)	-0.999069	0.001427	-699.9984	0.0000
SIGMASQ	4.634074	1.857810	2.494375	0.0239
R-squared	-0.056537	Mean dependent var		2.805263
Adjusted R-squared	-0.188604	S.D. dependent var		2.151690
S.E. of regression	2.345840	Akaike info criterion		4.869065
Sum squared resid	88.04741	Schwarz criterion		5.018187
Log likelihood	-43.25612	Hannan-Quinn criter.		4.894302
Durbin-Watson stat	1.622989			
Inverted AR Roots	1.00			
Inverted MA Roots	1.00			

Figure 4. ARIMA(1, 1, 1) model results  
图 4. ARIMA(1, 1, 1)模型结果

Dependent Variable: D(INDEX)				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/12/23 Time: 16:20				
Sample: 2002 2020				
Included observations: 19				
Convergence not achieved after 500 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.117443	0.000938	1191.701	0.0000
AR(2)	-0.117445	0.000135	-867.4650	0.0000
MA(1)	-0.997751	0.004044	-246.7531	0.0000
SIGMASQ	4.648343	2.016423	2.305242	0.0359
R-squared	-0.059790	Mean dependent var		2.805263
Adjusted R-squared	-0.271748	S.D. dependent var		2.151690
S.E. of regression	2.426500	Akaike info criterion		4.963623
Sum squared resid	88.31852	Schwarz criterion		5.162452
Log likelihood	-43.15442	Hannan-Quinn criter.		4.997273
Durbin-Watson stat	1.775774			
Inverted AR Roots	1.00	.12		
Inverted MA Roots	1.00			

Figure 5. ARIMA(1, 1, 2) model results

图 5. ARIMA(1, 1, 2)模型结果

Dependent Variable: D(INDEX)				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/12/23 Time: 16:20				
Sample: 2002 2020				
Included observations: 19				
Failure to improve objective (non-zero gradients) after 14 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	1.000000	2.391283	0.418186	0.6817
MA(1)	-0.570137	0.019441	-29.32584	0.0000
MA(2)	-0.429858	0.065115	-6.601514	0.0000
SIGMASQ	4.603652	1.862740	2.471441	0.0259
R-squared	-0.049601	Mean dependent var		2.805263
Adjusted R-squared	-0.259521	S.D. dependent var		2.151690
S.E. of regression	2.414807	Akaike info criterion		4.942133
Sum squared resid	87.46939	Schwarz criterion		5.140962
Log likelihood	-42.95026	Hannan-Quinn criter.		4.975782
Durbin-Watson stat	2.182020			
Inverted AR Roots	1.00			
Inverted MA Roots	1.00	-.43		

Figure 6. ARIMA(2, 1, 1) model results

图 6. ARIMA(2, 1, 1)模型结果

Dependent Variable: D(INDEX)				
Method: ARMA Maximum Likelihood (OPG - BHHH)				
Date: 10/12/23 Time: 16:21				
Sample: 2002 2020				
Included observations: 19				
Failure to improve objective (non-zero gradients) after 169 iterations				
Coefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.448821	0.153206	2.929524	0.0110
AR(2)	0.551179	0.207861	2.651672	0.0190
MA(1)	-0.053766	4.98E-05	-1078.883	0.0000
MA(2)	-0.946208	0.000228	-4156.859	0.0000
SIGMASQ	3.747182	1.517568	2.469202	0.0270
R-squared	0.145669	Mean dependent var		2.805263
Adjusted R-squared	-0.098426	S.D. dependent var		2.151690
S.E. of regression	2.255097	Akaike info criterion		4.911636
Sum squared resid	71.19646	Schwarz criterion		5.160173
Log likelihood	-41.66054	Hannan-Quinn criter.		4.953698
Durbin-Watson stat	2.000236			
Inverted AR Roots	1.00	-.55		
Inverted MA Roots	1.00	-.95		

Figure 7. ARIMA(2, 1, 2) model results

图 7. ARIMA(2, 1, 2)模型结果



由图 4 知，模型方程为：

$$x = 4.634074 + x_{t-1} - 0.999069\varepsilon_{t-1}。$$

由图 5 知，模型方程为：

$$x = 4.648343 + 1.117443x_{t-1} - 0.117445x_{t-2} - 0.997751\varepsilon_{t-1}。$$

由图 6 知，模型方程为：

$$x = 4.603652 + x_{t-1} - 0.570137\varepsilon_{t-1} - 0.429858\varepsilon_{t-2}。$$

由图 7 知，模型方程为：

$$x = 3.747182 + 0.448821x_{t-1} + 0.551179x_{t-2} - 0.053766\varepsilon_{t-1} - 0.946208\varepsilon_{t-2}。$$

### 3.4. 模型建立

为了在建立的四个模型中确立最优模型，需要对比参数调整 R2、AIC 和 SC。根据上面模型图可知这三个参数值对比表如下：

**Table 1.** Model parameter comparison table

**表 1.** 模型参数对比表

模型	AIC	SC
AR = 1, MA = 1	4.869065	5.018187
AR = 1, MA = 2	4.963623	5.162452
AR = 2, MA = 1	4.942133	5.140962
AR = 2, MA = 2	4.911636	5.160173

由表 1 可以很容易得到，模型拟合效果更优的是 ARIMA(1, 1, 1)模型。这是因为对于 AIC 和 SC 来说，他们都表示信息准则，这两个值越小越好。因此综合比较，可以得知 ARIMA(1, 1, 1)模型对于拟合我国居民价格消费指数的趋势来说效果更优。

### 3.5. 模型检验

通过以上结果可知，我国居民价格消费指数的趋势的最优模型为 ARIMA(1, 1, 1)。下面将验证该模型对于数据预测的准确性，现利用该模型拟合我国 2016 年至 2020 年的居民价格消费指数，并与之前的数据进行比较，检测结果如表 2 所示。

**Table 2.** Predicted and actual values from 2016 to 2020

**表 2.** 2016 年至 2020 年预测值与实际值

年度	2016	2017	2018	2019	2020
预测数值	141.2569	142.3856	147.8267	150.9981	156.2549
真实数值	140.77	143.02	146.02	150.25	154.00
相对误差	0.345883%	0.443574%	1.237296%	0.497903%	1.464221%

(相对误差 = |实际值 - 预测值| ÷ 实际值 × 100%)。

由上可知，模型拟合的数据预测值与实际值之间的误差相对较小，说明 ARIMA(1, 1, 1)的预测精度较高，能够预测未来的数据。因此我们用此来预测随后 5 年我国的 CPI 数据，结果如表 3：

**Table 3.** Predictions for 2021-2025  
**表 3.** 2021 年至 2025 年的预测值

年份	2021	2022	2023	2024	2025
预测值	157.2887	161.3425	163.2147	166.3826	168.5521

由预测值可以看出，未来我国居民价格消费指数处于小幅波动，但在合理范围内，总体较为稳定。

#### 4. 讨论

由以上分析可知 ARIMA 模型拟合效果还不错，可以用来分析预测我国居民消费价格指数。但是该模型缺陷时预测短期时间序列效果较好，时间越长与预测误差会逐渐变大。用该模型预测未来 5 年我国居民价格消费指数，发现其总体略微上升。

除此之外的建议措施如下：第一，我们可以加强物价监管的政策运行，面对通货膨胀时要有相对的措施应对；第二，完善生活必需品供给保障机制，确保市场有效供应；第三，推进全国各地物流体系建设，降低物流成本；第四，加强市场调控力度，整顿市场秩序；第五，合理控制价格改革节奏好力度，整顿市场秩序；第六，健全联动机制，系统解决物价上涨对低收入人群的影响。

#### 5. 结论

本文我们建立 ARIMA 模型，我国 2001 年至 2020 年的居民价格消费指数是一组非平稳的时间序列数据。本文通过运用 eviews 软件我国居民价格消费指数进行分析，建立 ARIMA(1, 1, 1)模型，并且用这个模型来预测我国未来 2021 年~2025 年居民价格消费指数的数值变化，通过对我国未来五年的居民价格消费指数进行分析比较，可以得知其相对误差相对较小，可以得出 ARIMA 模型的拟合程度相对于其他模型更好，该模型的预测准确度更高，因此我们可以利用 ARIMA(1, 1, 1)模型来预测我国未来五年内的居民价格消费指数的趋势。因此，便利用该模型对 2021 年至 2025 年的 CPI 进行了预测，结果显示，我国的居民价格消费指数呈略微上升的趋势。

但是影响我国居民价格消费指数的因素是多样且不确定的，而建立 ARIMA 模型时并未考虑其他因素，该模型只是靠统计的数据反映出的一些基本信息。在发生一些突然变故时(如：金融危机、投资行为、政治活动影响等)，模型预测效果可能不会这么好。因此在预测未来我国居民价格消费指数趋势发展时，有必要将这些因素考虑在内，结合当时的实际经济状况及国际上经济、政治的重大事件，才能得出更为合理准确的结果和经济政策。

#### 参考文献

- [1] 马慧慧, 郭庆然, 丁翠翠. Eviews 统计分析与应用[M]. 第 3 版. 北京: 电子工业出版社, 2016: 142-156, 293.
- [2] 李璇, 黄冬冬. 基于 ARIMA 模型的 CPI 实证分析及预测[J]. 沈阳大学学报(社会科学版), 2013, 15(3): 306-310.
- [3] 杨颖梅. 基于 ARIMA 模型的北京居民消费价格指数预测[J]. 统计与决策, 2015(4): 76-78.
- [4] 张云峰. 居民消费价格指数的分析统计与预测[J]. 河北企业, 2017(12): 40-41.
- [5] 初睿. 基于 ARIMA 模型的北京市 CPI 预测分析[J]. 价值工程, 2018, 37(5): 21-22.
- [6] 袁黎, 赵昌英, 杨忠丽, 姚娟. 基于 ARIMA 模型的贵州省 CPI 预测[J]. 科技经济前沿, 2019, 27(2): 21-22.
- [7] 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.
- [8] 刘锐. 基于 ARIMA 模型的西安市居民消费价格指数分析及预测[J]. 西部经济, 2017(7): 18-19.
- [9] 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.

- [10] 找伟霖, 旷永鑫, 郭兴方. 基于 ARMA 模型的南阳市 CPI 预测分析[J]. 河南教育学院学报(自然科学版), 2018, 27(3): 18-21.
- [11] 杜文晟. 基于时间序列方法的湖北省居民消费价格指数预测分析[J]. 湖北师范大学学报(哲学社会科学版), 2019, 39(6): 45-48.