

基于ARIMA与协整模型的居民人均消费支出预测研究

林元书

上海工程技术大学管理学院, 上海

收稿日期: 2023年10月11日; 录用日期: 2023年12月8日; 发布日期: 2023年12月18日

摘要

居民人均消费支出是体现居民生活水平和质量的重要指标, 本研究利用1980~2021年居民人均消费支出数据, 构建ARIMA模型与协整模型, 对2022~2026年居民人均消费支出进行预测, 并将ARIMA模型与协整模型从模型拟合效果和预测精度两方面进行了对比分析, 选择了AIC值更小的ARIMA模型预测值{28568.50, 32124.36, 36123.18, 40619.36, 45675.62}作为本研究的预测结果。

关键词

ARIMA模型, 协整模型, 居民人均消费支出

Research on Predicting Per Capita Consumption Expenditure of Residents Based on ARIMA and Cointegration Model

Yuanshu Lin

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Oct. 11th, 2023; accepted: Dec. 8th, 2023; published: Dec. 18th, 2023

Abstract

Per capita consumption expenditure is an important indicator reflecting the living standards and quality of residents, this study uses the per capita consumption expenditure data of residents from 1980 to 2021 to construct ARIMA model and Cointegration model to predict the per capita consumption expenditure of residents from 2022 to 2026, compares and analyzes the ARIMA model

and Cointegration model from the aspects of model fitting effect and prediction accuracy, selects the predicted value of ARIMA model with smaller AIC value {28568.50, 32124.36, 36123.18, 40619.36, 45675.62} as the prediction results.

Keywords

ARIMA Model, Cointegration Model, Per Capita Consumption Expenditure of Residents

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

提高居民消费, 扩大有效需求, 是经济可持续发展的基础, 目前对居民消费的研究集中在影响研究上, 黄家林[1]基于中国家庭追踪调查数据, 研究大病医疗保险对居民消费的影响, 姚战琪[2]研究数字经济对居民人均消费支出的影响, 罗正媛[3]分析城市人口密度对居民消费的影响, 倪坤[4]以相关资料和理论为基础, 研究近代中国居民消费影响因素, 鲜有文献对居民人均消费支出预测进行研究。因此, 本研究基于 1980~2021 年我国居民人均消费支出数据, 构建 ARIMA 模型与协整模型, 对 2022~2026 年居民人均消费支出进行预测研究, 为居民消费研究提供一种新的思路。

2. 基于 ARIMA 模型对我国居民人均消费支出的预测

ARIMA 模型用于时间序列数据, 能反映动态数据变化规律, 对未来趋势进行预测[5]。本节采用 ARIMA 模型对 2022~2026 的居民人均消费支出进行预测。

2.1. 数据描述

研究对象为我国的居民人均消费支出{c}, 选用来自国家统计局 1980~2021 年我国居民人均消费支出数据, 数据连续完整, 可靠性强, 作为本节研究的原始数据, 如表 1 所示。

Table 1. Per capita consumption expenditure of Chinese residents from 1980 to 2021 (yuan)

表 1. 1980~2021 年我国的居民人均消费支出(元)

Time	c	Time	c	Time	c	Time	c	Time	c	Time	c
1980	211	1987	521	1994	1540	2001	3139	2008	7548	2015	15,712
1981	244	1988	639	1995	1957	2002	3548	2009	8377	2016	17,111
1982	273	1989	712	1996	2288	2003	3889	2010	9378	2017	18,322
1983	304	1990	768	1997	2437	2004	4395	2011	10,820	2018	19,853
1984	340	1991	844	1998	2516	2005	5035	2012	12,054	2019	21,559
1985	402	1992	937	1999	2658	2006	5634	2013	13,220	2020	21,210
1986	465	1993	1145	2000	2914	2007	6592	2014	14,491	2021	24,100

2.2. 平稳性及白噪声检验

对时间序列数据进行研究,首先要对数据的平稳性进行检验[6]。利用时序图对平稳性进行简单判别,考虑到该序列有呈指数上升的趋势,对序列{c}作对数变换得到序列{lnc},绘制出序列{c}与序列{lnc}时序图,如图1所示。

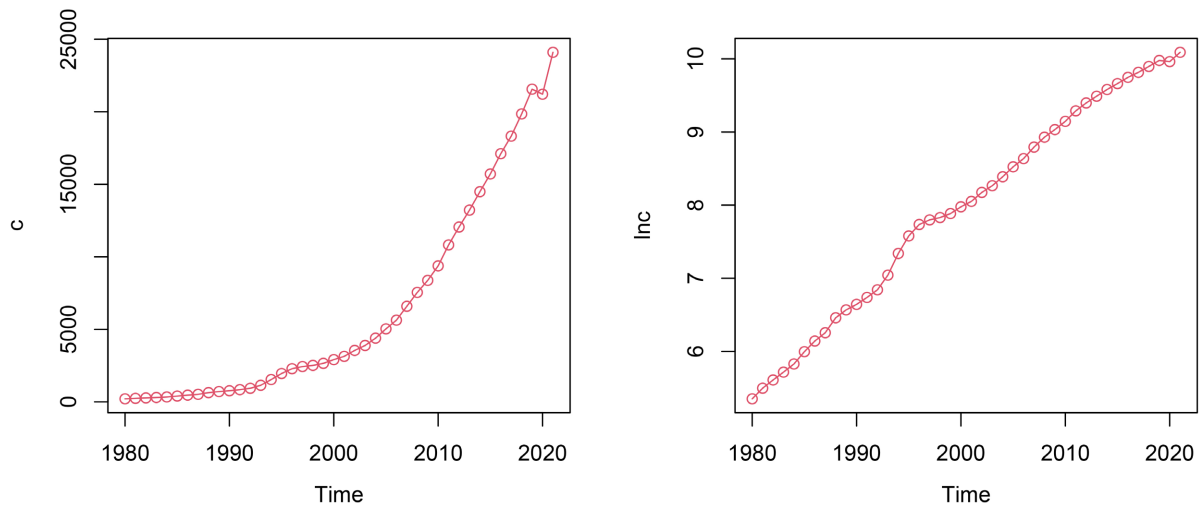


Figure 1. The timing diagram of the sequence {c} and the sequence {lnc}

图 1. 序列{c}和序列{lnc}的时序图

由图1可知,序列{c}与序列{lnc}都有明显递增趋势,均为非平稳序列,再对序列{lnc}进行一阶差分,得到序列{dlnc1},绘制出时序图,如图2所示。

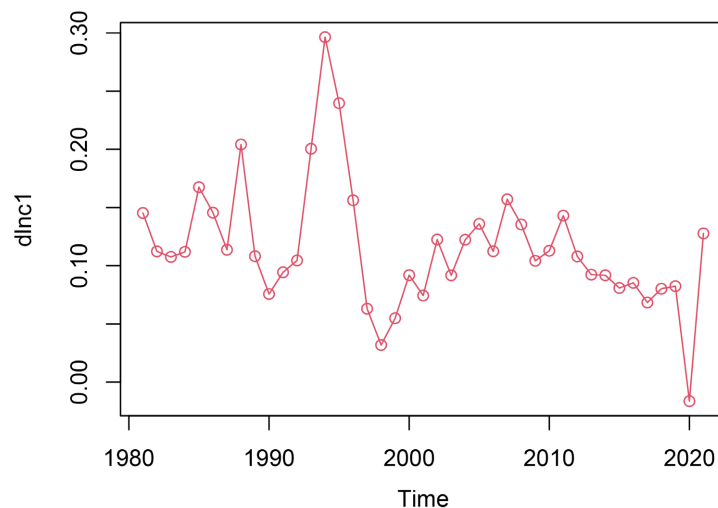


Figure 2. Timing diagram of the sequence {dlnc1}

图 2. 序列{dlnc1}的时序图

由图2可知,序列{dlnc1}在一定范围内来回波动,不能准确判定{dlnc1}是否为平稳序列,再进行ADF检验[7],如表2所示。

Table 2. ADF result
表 2. ADF 检验结果

不含漂移项和趋势项			只含漂移项			含漂移项和趋势项		
lag	ADF	p.value	lag	ADF	p.value	lag	ADF	p.value
0	-1.356	0.189	0	-3.39	0.0202	0	-3.59	0.0461
1	-1.161	0.258	1	-3.48	0.0167	1	-3.84	0.0269
2	-1.008	0.313	2	-3.42	0.0192	2	-3.98	0.0205
3	-0.927	0.342	3	-3.58	0.0127	3	-4.51	0.0100

由表 2 可知, 只含漂移项与含漂移项和趋势项两种类型的 p 值均小于显著性水平 0.05, 可认为 $\{dlnc1\}$ 序列显著平稳, 且 $\{dlnc1\}$ 的确定性部分可以用只含漂移项与含漂移项和趋势项等多种模型结构进行拟合。为判定序列 $\{dlnc1\}$ 是否具有建模的价值, 进行白噪声检验[8], 如表 3 所示。

Table 3. Sequence $\{dlnc1\}$ white noise test results
表 3. 序列 $\{dlnc1\}$ 白噪声检验结果

Ljung-Box			
lag	6	p.value	0.002442

由表 3 可知, 序列 $\{dlnc1\}$ 检验 p 值为 0.002442, 远小于显著性水平 0.05, 表明序列 $\{dlnc1\}$ 是平稳非白噪声序列, 可以进行后续建模。

2.3. ARIMA 模型的建立与检验

绘制出自相关图和偏自相关图, 进行模型定阶的尝试[9], 如图 3 所示。

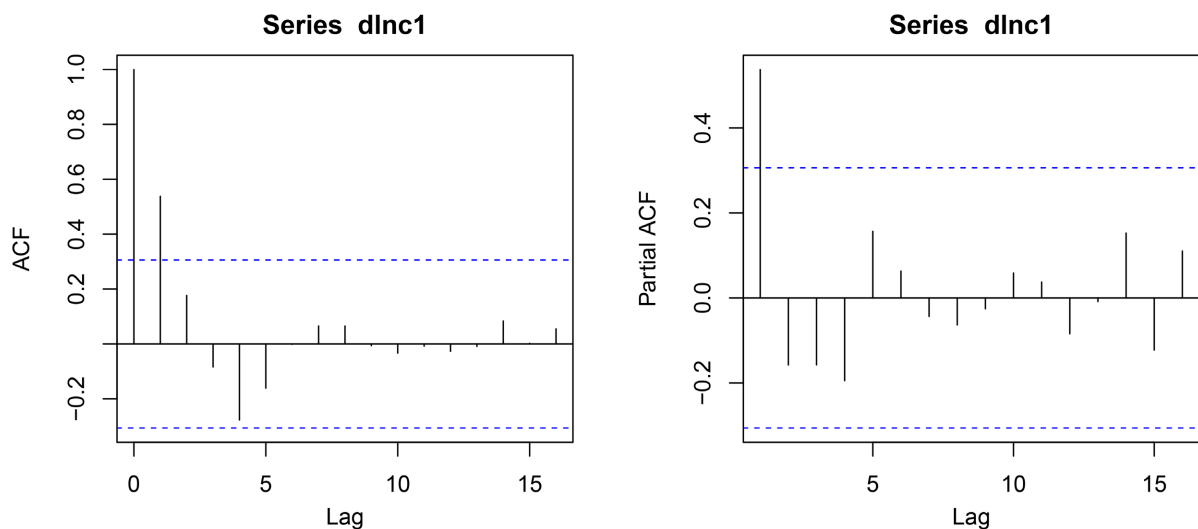


Figure 3. Autocorrelation plots and partial autocorrelation plots of the sequence $\{dlnc1\}$
图 3. 序列 $\{dlnc1\}$ 的自相关图和偏自相关图

由图 3 可知, $\{\ln c_1\}$ 自相关图 1 阶截尾, 偏自相关图拖尾, 再结合 ADF 检验结果, 使用带有漂移项的 ARIMA(0,1,1)模型拟合序列 $\{c\}$ 。为检验模型各参数的显著性, 进行参数估计[10], 如表 4 所示。

Table 4. Parameter estimation results

表 4. 参数估计结果

参数	Ma1	Drift
参数估计	0.6290	0.1173
标准差	0.1374	0.0112

由表 4 可知, 模型参数估计值均大于 2 倍标准差, 认为参数显著非零。为检验拟合模型是否显著, 对拟合模型残差序列进行白噪声检验, 若残差序列为白噪声序列, 则模型显著成立, 如表 5 所示。

Table 5. Results of the white noise test for the residual sequence

表 5. 残差序列白噪声检验结果

Ljung-Box			
lag	6	p.value	0.6546

由表 5 可知, 残差序列 p 值为 0.6546, 远大于显著性水平 0.05, 表明残差序列是白噪声序列, 因此, 构建带有漂移项的 ARIMA(0,1,1)模型显著成立, 可以进行后续预测。

2.4. 模型的预测

以构建的 ARIMA 模型为依据, 对 1980~2021 年我国居民人均消费支出进行拟合, 如图 4 所示。

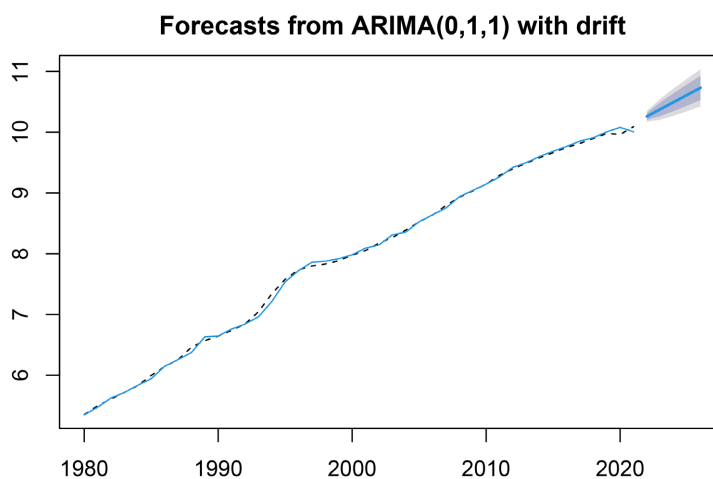


Figure 4. ARIMA(0,1,1) model fit plot with a drift term

图 4. 带有漂移项的 ARIMA(0,1,1)模型拟合曲线图

由图 4 可知, 最优拟合数据与实际数据较吻合, 能够描述出我国居民人均消费支出的增长趋势, 模型拟合效果较好。因此, 利用 ARIMA 模型预测 2022~2026 年的居民人均消费支出, 如表 6 所示。

Table 6. The forecast value of resident per capita consumption expenditure in 2022~2026
表 6. 2022~2026 的居民人均消费支出预测值

Time	Point Forecast	Lo80	Hi80	Lo95	Hi95
2022	28568.50	26946.27	48387.66	26125.05	31240.48
2023	32124.36	28728.07	48387.66	27078.08	38111.45
2024	36123.18	31190.54	48387.66	28858.22	45217.07
2025	40619.36	34098.55	48387.66	31081.56	53084.49
2026	45675.62	37424.84	48387.66	33678.67	61946.10

3. 基于协整模型对我国居民人均消费支出的预测

协整模型是有效描述非平稳变量关系的方法，可实现时间序列分析中短期和长期模型优势的有力结合[11]，本节采用协整模型对 2022~2026 的居民人均消费支出进行预测。

3.1. 数据的描述

相关研究指出，居民人均消费支出受居民人均可支配收入的影响最大，其次是人均 GDP [12]。因此，本节选用来自国家统计局 1980~2021 年我国的居民人均消费支出{c}、居民人均可支配收入{r}和人均 GDP{g}等数据，作为本节研究的原始数据，如表 7，表 8 所示。

Table 7. Per capita disposable income of Chinese residents from 1980 to 2021 (yuan)
表 7. 1980~2021 年我国居民人均可支配收入(元)

Time	r	Time	r	Time	r	Time	r	Time	r	Time	r
1980	247	1987	599	1994	1870	2001	4070	2008	9957	2015	21,966
1981	279	1988	709	1995	2363	2002	4532	2009	10,977	2016	23,821
1982	326	1989	804	1996	2814	2003	5007	2010	12,520	2017	25,974
1983	365	1990	904	1997	3070	2004	5661	2011	14,551	2018	28,228
1984	424	1991	976	1998	3254	2005	6385	2012	16,510	2019	30,733
1985	479	1992	1125	1999	3485	2006	7229	2013	18,311	2020	32,189
1986	541	1993	1385	2000	3721	2007	8584	2014	20,167	2021	35,128

Table 8. 1980~2021 China's per capita GDP (yuan)
表 8. 1980 年~2021 年我国人均 GDP(元)

Time	g	Time	g	Time	g	Time	g	Time	g	Time	g
1980	468	1987	1123	1994	4081	2001	8717	2008	24,100	2015	49,922
1981	497	1988	1378	1995	5091	2002	9506	2009	26,180	2016	53,783
1982	533	1989	1536	1996	5898	2003	10,666	2010	30,808	2017	59,592
1983	588	1990	1663	1997	6481	2004	12,487	2011	36,277	2018	65,534
1984	702	1991	1912	1998	6860	2005	14,368	2012	39,771	2019	70,078

Continued

1985	866	1992	2334	1999	7229	2006	16,738	2013	43,497	2020	71,828
1986	973	1993	3027	2000	7942	2007	20,494	2014	46,912	2021	80,976

3.2. 平稳性检验

对多元时间序列数据进行研究,首要考虑的是数据序列的平稳性。为消除数量级的影响,将原序列做对数变换,得到输入序列{lnr}和{lng},输出序列{lnc},并绘制出时序图,如图5所示。

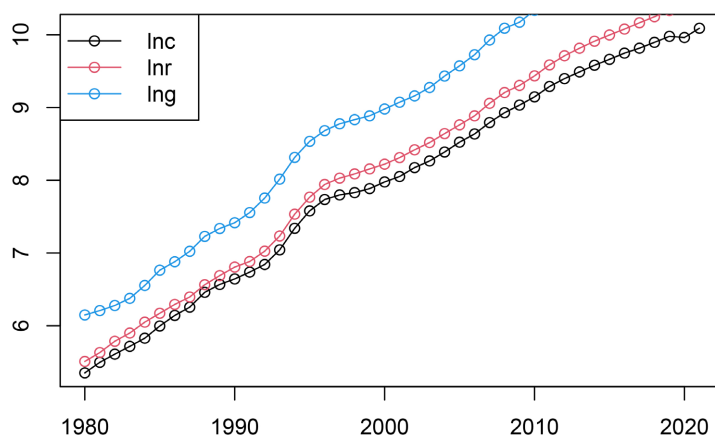


Figure 5. Time timing diagram of {lnc}, {lnr} and {lng}

图5. {lnc}、{lnr}和{lng}的时序图

由图5可知,序列{lnc}、{lnr}和{lng}呈现出一致的增长趋势,均为不平稳序列,再进行一阶差分,并对一阶差分后的序列{dlnc1}、{dlnr1}和{dlng1}进行具有漂移项的ADF检验,如表9所示。

Table 9. ADF result

表9. ADF 检验结果

Dlnc1			Dlnr1			Dlng1		
lag	p.value	结论	lag	p.value	结论	lag	p.value	结论
0	0.0202	平稳	0	0.0962	非平稳	0	0.0512	平稳
1	0.0167	平稳	1	0.0321	平稳	1	0.0101	平稳
2	0.0192	平稳	2	0.0102	平稳	2	0.1375	非平稳

在滞后阶数为1时,序列{dlnc1}、{dlnr1}和{dlng1}均达到了平稳。原序列为非平稳序列,一阶差分后达到平稳,可认为原序列是同阶单整的,再采用EG两步法检验序列{lnc}、{lnr}和{lng}间是否存在协整关系,若存在,则可构建协整模型[13]。

3.3. 协整模型的构建与检验

首先,需要建立响应序列与输入序列之间的回归模型。本研究有多个输入序列,考虑到不同的输入序列对输出序列的影响可能会不同,绘制出{lnc}分别与{lnr}、{lng}的互相关图,如图6所示。

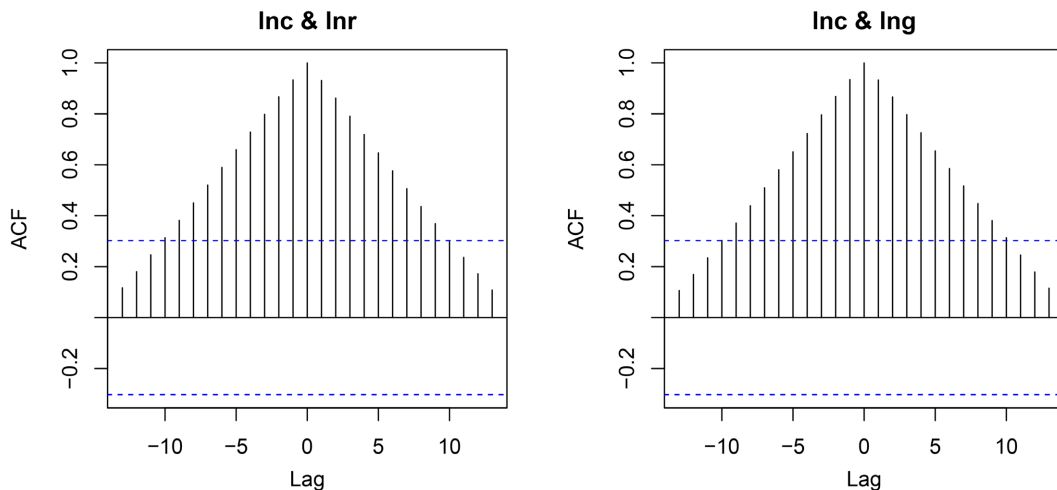


Figure 6. Cross-correlation plot of the input and output sequences
图 6. 输入序列与输出序列的互相关图

由图 6 可知，输入序列{lnr}、{lng}对输出序列{inc}的影响均在当期达到最大，利用当期序列建立如下回归模型：

$$\ln c_t = \beta_1 \ln r_t + \beta_2 \ln g_t + \varepsilon_t$$

利用 R 进行拟合后得到以下模型：

$$\ln c_t = 0.4853 \ln r_t + 0.4443 \ln g_t + \varepsilon_t$$

为检验模型各参数的显著性，进行参数估计，结果如表 10 所示。

Table 10. Parameter estimation results
表 10. 参数估计结果

参数	Lnc	Lnr
参数估计	0.4853	0.4443
标准差	0.0652	0.0598

由表 10 可知，参数估计值均大于 2 倍标准差，认为参数显著非零。

其次，对回归残差序列进行平稳性检验，如表 11 所示。

Table 11. Results of the regression residual sequence stationarity test
表 11. 回归残差序列平稳性检验结果

不含漂移项和趋势项			只含漂移项			含漂移项和趋势项		
lag	ADF	p.value	lag	ADF	p.value	lag	ADF	p.value
0	-3.70	0.0100	0	-3.71	0.0100	0	-4.46	0.0100
1	-3.03	0.0100	1	-3.00	0.0461	1	-3.50	0.0547
2	-2.93	0.0100	2	-2.91	0.0569	2	-3.53	0.0502
3	-2.55	0.0133	3	-2.51	0.1394	3	-3.18	0.1072

由表 11 可知, 不含漂移项和趋势项的 p 值小于显著性水平 0.05, 认为该序列显著平稳, 再对拟合模型残差序列进行白噪声检验, 如表 12 所示。

Table 12. Results of the regression residual sequence white noise test
表 12. 回归残差序列白噪声检验结果

Ljung-Box			
lag	6	p.value	0.01277

由表 12 可知, 残差序列检验 p 值为 0.01277, 远小于显著性水平 0.05, 表明残差序列为平稳非白噪声序列。通过残差序列的自相关图和偏自相关图, 进一步提取残差序列中的信息, 如图 7 所示。

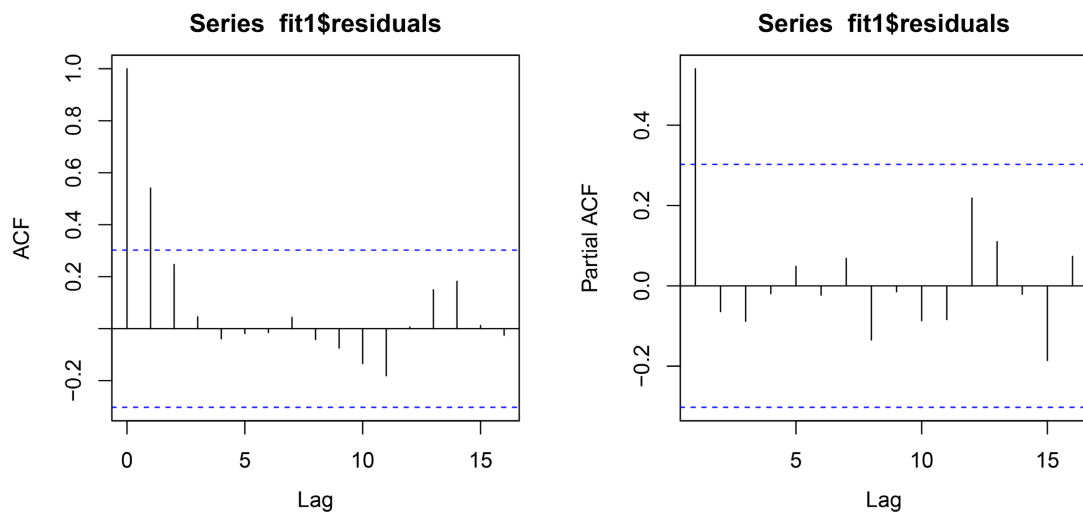


Figure 7. Autocorrelation plots and partial autocorrelation plots of the residue sequence
图 7. 残差序列的自相关图和偏自相关图

由图 7 可知, 自相关图拖尾, 偏自相关图 1 阶截尾, 采用 AR(1)模型拟合残差序列, 最终拟合的协整模型如下:

$$\ln c_t = 0.6378 \ln r_t + 0.3041 \ln g_t + \frac{\varepsilon_t}{1 - 0.6752B}$$

构建出协整模型后, 对模型进行参数估计和显著性检验, 如表 13 和图 8 所示。

Table 13. Parameter estimation results
表 13. 参数估计结果

参数	Ar1	Lnr	Lng
参数估计	0.6752	0.6378	0.3041
标准差	0.1277	0.1000	0.0918

由表 13 可知, 参数估计值均大于 2 倍标准差, 认为参数显著非零。

Residual Diagnostics Plots

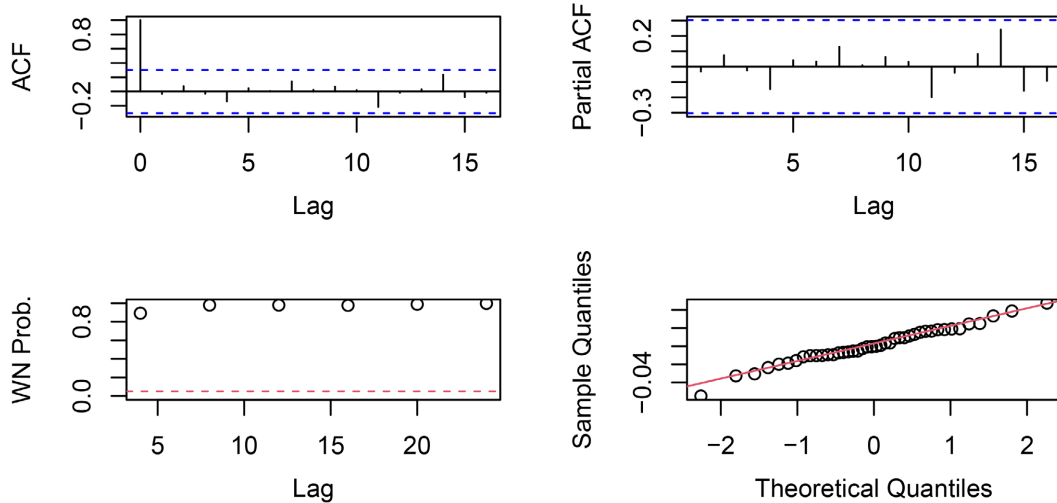


Figure 8. Model significance test
图 8. 模型显著性检验

由图 8 可知，p 值均显著大于 0.05，说明残差序列为白噪声序列，即该拟合模型显著有效。

3.4. 模型预测

首先，通过单变量拟合方法获得序列{lnr}和{lng}的预测值。采用 auto.arima 函数对序列{lnr}和{lng}进行拟合，得到序列{lnr}拟合模型为带有漂移项的 ARIMA(2,1,2)，{lng}拟合模型为带有漂移项的 ARIMA(0,1,2)，两种拟合模型的参数估计结果如表 14，表 15 所示。

Table 14. ARIMA (2,1,2) with drift parameter estimation results
表 14. ARIMA(2,1,2) with drift 参数估计结果

参数	ar1	ar2	ma1	ma2	drift
参数估计	-0.5037	0.3934	1.5398	0.7348	0.1208
标准差	0.1743	0.1867	0.1915	0.2025	0.0133

Table 15. ARIMA(0,1,2) with drift parameter estimation results
表 15. ARIMA(0,1,2) with drift 参数估计结果

参数	ma1	ma2	drift
参数估计	1.1990	0.4069	0.1267
标准差	0.2217	0.2269	0.0155

由表 14，表 15 可知，参数估计值均大于 2 倍标准差，认为参数显著非零。协整模型由{lnr}、{lng}和残差序列三项构成，且各项的系数已确定，只需获取 2022~2026 年序列{lnr}、{lng}和残差序列的值，再利用协整模型即可得到序列{c}的预测值，如表 16 所示。

Table 16. Forecast value of household per capita consumption expenditure from 2022 to 2026**表 16.** 2022~2026 年的居民人均消费支出预测值

Time	Point Forecast	Lo80	Hi80	Lo95	Hi95
2022	27914.64	26077.65	29878.07	25156.38	30973.29
2023	31928.82	28108.46	36268.54	26274.75	38799.46
2024	35704.70	29878.47	42667.15	27189.45	46886.62
2025	40338.28	32480.56	50097.10	28960.92	56185.26
2026	45086.55	35084.17	57940.95	30721.46	66169.07

4. 基于 ARIMA 模型与协整模型比较分析

本节将从模型拟合效果和预测精度等两方面来对 ARIMA 模型和协整模型进行比较分析，筛选出较优的模型，确定出 2022~2026 年居民人均消费支出预测值。

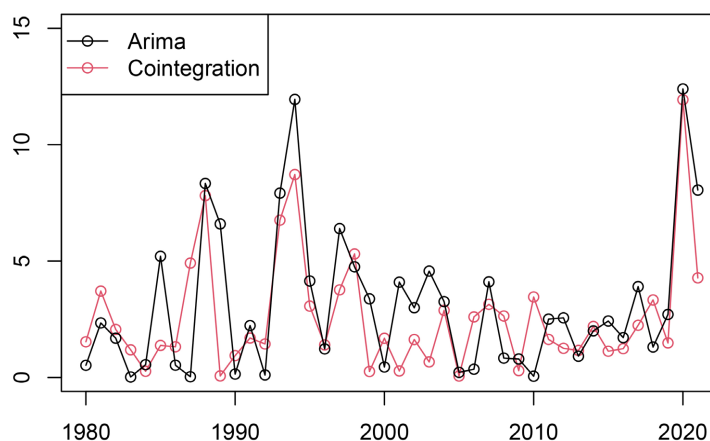
首先，从模型拟合效果的角度分析，比较 ARIMA 模型和协整模型的 AIC 值，如表 17 所示。

Table 17. AIC values of the two models**表 17.** 两个模型 AIC 值对比

模型	ARIMA 模型	协整模型
AIC	-132.38	-201.94

由表 17 可知，与协整模型相比，ARIMA 模型的 AIC 值相对较小。ARIMA 模型是根据居民人均消费支出随时间的变化而构建的单变量模型，不受其他变量的干扰，而协整模型是考虑其他变量而构建的多变量模型，考虑因素更多，更容易受到干扰，会在一定程度上增大模型拟合的误差。

其次，从预测精度的角度分析，将 ARIMA 模型与协整模型对 1980~2021 年我国居民人均消费支出的最优拟合值导出，并与实际值进行比较。为方便说明，用实际值与最优拟合值之间的相对误差率来说明预测模型的精度，如图 9 所示。

**Figure 9.** The relative error rate comparison in Fig**图 9.** 相对误差率对比图

由图 9 可知, 两者的相对误差率均在 0%~15% 内波动, 协整模型相对误差率波动比 ARIMA 模型小, 但整体趋势接近, 难以区分哪种模型预测精度会更高。因此, 在预测精度大体相同时, 选择 AIC 值更小的 ARIMA 模型预测值作为本研究的预测结果, 如表 18 所示。

Table 18. Forecast value of household per capita consumption expenditure from 2022 to 2026

表 18. 2022~2026 年的居民人均消费支出预测值

Time	Point Forecast	Lo80	Hi80	Lo95	Hi95
2022	28568.50	26946.27	48387.66	26125.05	31240.48
2023	32124.36	28728.07	48387.66	27078.08	38111.45
2024	36123.18	31190.54	48387.66	28858.22	45217.07
2025	40619.36	34098.55	48387.66	31081.56	53084.49
2026	45675.62	37424.84	48387.66	33678.67	61946.10

由表 18 可知, 未来五年居民人均消费支出仍呈增长趋势, 并且将在 2025 年突破 4 万元。

5. 总结

本研究以 1980~2021 年居民人均消费支出数据为基础, 构建出 ARIMA 模型与协整模型, 进行模型的对比分析, 得到 2022~2026 年居民人均消费支出预测值。本研究巧妙地将 ARIMA 模型与协整模型运用到居民消费预测上, 为该研究领域提供了一种新的思路, 而由于数据收集的困难, 本研究只选择了 42 年的数据, 若增加数据量, 获取更长期的变动趋势, 预测结果将会更加精准。

参考文献

- [1] 黄家林, 傅虹桥, 宋泽. 补充医疗保险对居民消费的影响——来自城乡居民大病保险的证据[J]. 金融研究, 2022(10): 58-76.
- [2] 姚战琪. 数字经济对我国居民人均消费支出的影响研究[J]. 贵州社会科学, 2022(9): 111-120.
- [3] 罗正媛, 赵世秀. 城市化背景下人口密度对居民消费的影响——以成渝城市群为例[J]. 商业经济研究, 2022(14): 64-67.
- [4] 倪坤, 刘巍. 近代中国居民消费影响因素研究(1910-1936) [J]. 四川大学学报(哲学社会科学版), 2022(3): 182-192.
- [5] 郑明贵, 于明, 范秋蓉, 等. 中国 2025-2035 年碳酸锂需求预测——基于灰色关联分析和 ARIMA-GM-BP 神经网络的组合模型[J]. 地球科学进展, 2023, 38(4): 377-387.
- [6] 孟祥煜. 我国国债期货价格波动率预测研究——基于 HAR-GARCH-GRU 混合模型的分析[J]. 价格理论与实践, 2022(7): 84-88.
- [7] 陈治霖, 胡韬韬, 边迎迎. 新冠疫情下基于 SARIMA 模型的上海港集装箱吞吐量预测[J/OL]. 工业工程与管理: 1-16. <http://kns.cnki.net/kcms/detail/31.1738.T.20230914.1716.024.html>, 2023-10-10.
- [8] 龙宇, 许浩然, 余华云, 等. 基于 ARIMA-LSTM-XGBoost 组合模型的铁路货运量预测[J]. 科学技术与工程, 2023, 23(25): 10879-10886.
- [9] 杨美涛, 王彦丁, 李志强, 等. ARIMA-SVM 组合模型在肺结核发病趋势预测中的应用[J]. 现代预防医学, 2023, 50(11): 1921-1926.
- [10] 韩宇浩, 吴茜, 彭金碧, 等. 广东省职业性尘肺病 ARIMA 模型预测[J]. 中国职业医学, 2023, 50(2): 150-154.
- [11] 王晓兵, 郑海文, 孔祥玉. 基于协整分析与改进潜结构投影的质量相关故障检测[J]. 2023, 30(10): 108-113.
- [12] 吴旭. 我国居民消费水平的影响因素和现状分析[J]. 统计与管理, 2021, 36(10): 4-10.
- [13] 白仲林, 孙艳华. 一种协整时间序列的动态因果效应估计与推断方法[J]. 统计研究, 2021, 38(10): 134-150.