

关于 ARIMA 数学模型的实际应用

Discussion on the Practical Application of ARIMA Mathematical Model

廖萌祺

Mengqi Liao

西南大学计算机与信息科学学院软件学院 中国·重庆 400715

College of Computer & Information Science, Southwest University, Chongqing, 400715, China

摘要: 近年来, 由于受到特殊时期及国际大环境的影响, 消费水平下降成为阻碍中国经济快速发展的重要因素。因此, 国家及重庆市推出了增强城镇和农村居民消费的相关政策, 其中成渝双城经济圈对重庆市的影响比较显著。论文通过利用 ARIMA 模型建模, 分析了影响城乡居民消费的特征和差异、关键指标, 预测了成渝双城经济圈的建立对五年后重庆居民消费的影响, 同时预测了居民消费可以恢复正常的时间。

Abstract: In recent years, due to the influence of the special period and the international environment, the insufficient consumer demand has become an important factor hindering the rapid development of China's economy. Therefore, the state and Chongqing municipality have launched the relevant policies to enhance the consumption of urban and rural residents, among which the Chengdu-Chongqing Shuangcheng Economic Circle has a relatively significant impact on Chongqing municipality. By learning the ARIMA model and modeling, this paper analyzes the characteristics, differences and key indicators affecting the consumption of urban and rural residents, predicts the impact of the establishment of Chengdu-Chongqing Shuangcheng Economic Circle on the consumption of Chongqing residents in five years, and predicts the time when residents' consumption can return to normal.

关键词: 最小二乘法时间序列分析; Eviews10 软件; ARIMA 模型

Keywords: least squares time series analysis; Eviews10 software; ARIMA model

DOI: 10.12346/sde.v4i9.7137

1 引言

论文收集了中国重庆市 1993—2021 年城镇居民和农村居民消费水平情况, 利用时间序列分析方法对其进行预测。首先, 运用 EViews10 程序, 对所采集的消费信息进行大数据分析处理; 其次, 再利用指数函数、差分方法去除了一些确定性趋势, 从而构建起居民消费价格序列与时间 T 的相互关联模型; 再次, 再通过最小二乘得到模型的各个系数; 最后, 再分别构建了重庆城镇居民与农村共同消费的 ARIMA 模型, 进而讨论成渝双城经济圈的建立对五年后重庆居民消费的影响。

2 ARIMA 模型的简要介绍

ARIMA 在 20 世纪 70 年代被提出, 全称为自回归

移动平均模型 (Autoregressive Integrated Moving Average Model, 简记 ARIMA)。在 70 年代初期, 博克斯 (Box) 和詹金斯 (Jenkins) 共同发明了 ARIMA 这一经典的时间序列预测方式, 所以此方法也叫做 Box-Jenkins 模型、博克斯-詹金斯法。其中 ARIMA (p, d, q) 称为差分自回归移动平均模型, AR 是自回归, p 为自回归项; MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数^[1]。

ARIMA 模型的基本思想是将消费情况随时间序列而形成的数据序列视为一个随机序列, 用这个序列对过去消费水平的变化和特性进行分析, 并对未来的消费情况做出预测。

论文基于当前时事粗浅地使用 ARIMA 模型对近几年的城镇、农村居民消费作简要的分析与讨论。

【作者简介】廖萌祺 (2002-), 女, 中国江西瑞金人, 在读本科生。

3 成渝双城经济圈的建立对五年后重庆居民消费的影响

3.1 数据收集

关于 1993 年至 2021 年重庆市城乡居民消费水平的数据收集见表 1。

3.2 研究思路

3.2.1 主要方法

论文主要使用应用时间序列分析方法^[2]对中国重庆市 1993—2021 年居民消费水平进行分析,其方法基本思想是:想要预测某一个现象的未来变化趋势时,会用过去现象的变化规律来预测未来,即以时间序列为主线,分析历史数据的变化规律,并将这种历史的变化规律沿着时间序列延伸到未来,从而对未来的现象做出预测。但必须注意的是,在实际中因为受许多因子的影响,时间序列通常是非平衡的,特别是某些长远性、关键性的因子往往导致时间序列出现了特定的规律性,甚至是往特定方向;还有部分因素则起着短期、长期非关键性的影响,从而导致时间序列的整体规律以及长期走势的局部规律出现了一定的无序化现象。为有效归纳出长期变化规律和作出预报,论文拟除去不能预测、长期不稳定的各种因素,以确定重庆市城乡居民生活状况随时间改变而呈现的变化趋势。

3.2.2 基本模型

① ARMA (p, q) 自回归移动平均模型。

如果时间序列是它的现在和过去的随机误差项以及过去值的线性函数^[3],那么式子为:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

则称该时间序列是自回归平均序列,上式为 (p, q)

阶的自回归移动平均模型,记为 ARMA (p, q)。其中, $\varphi_1, \varphi_2, \dots, \varphi_p$ 为自回归系数, $\theta_1, \theta_2, \dots, \theta_q$ 为移动平均系数,都是模型的待估系数。在此基础上,引入滞后算子 B,上述方程可简记为: $\Phi(B)X_t = \Psi(B)\varepsilon_t$ 。ARMA (p, q) 过程的平稳条件是滞后多项式 $\Phi(B)$ 的根均在单位圆外,可逆条件是 $\Phi(B)$ 的根都在单位圆外。

② ARIMA (p, d, q) 求和自回归移动平均模型。

ARIMA 模型的实质^[3]就是差分运算与 ARMA 模型的组合,其表达式为

$$\begin{cases} (B)^d \Phi(B)X_t = \Psi(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_s \varepsilon_t) = 0, s \neq t \\ EX_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (1)$$

式中 $\Phi(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$ 为平稳可逆 ARMA (p, q) 模型的自回归系数多项式; $\Psi(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为平稳可逆 ARMA (p, q) 模型的移动平滑系数多项式^[4]。

3.3 建立模型及预测

3.3.1 模型识别

利用 EViews10 程序画出 1993 年至 2021 年重庆市城镇和农村居民消费水平时序图,见图 1。

从图 1 重庆市城乡居民消费水平时序图看出序列呈现确定性上升趋势(确定性趋势就是时间序列在一个比较长的时期内,受某种或某几种确定性因素影响而表现出的某种持续上升或持续下降的趋势,可以通过适当的数学模型很好地拟合这种趋势)。

如果序列数据在经单位根检验后不拒绝原假设,就可以充分说明原始序列不平稳,且伴有明显的上升趋势。于是论文拟对该序列作对数变换的调整,并以此为依据画出对数变化后时间序列图,如图 2 所示。

表 1 1993 年至 2021 年重庆市城乡居民消费水平

年份	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011
重庆市城镇居民消费水平(元/人)	29850	26464	25786	24154	30101	28209	25794.5	24000	21681	19873.3	17972.65
重庆市农村居民消费水平(元/人)	16096	14140	13112	11977	10527	9433	8836.7	7577	6538	5740.5	4614.895
年份	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000
重庆市城镇居民消费水平(元/人)	12818	11709	10781	9596	9032	7959	9033	8447	7274	6766	6544
重庆市农村居民消费水平(元/人)	4359	3722	3369	2906	2341	2251	1827	1599	1532	1491	1380
年份	1999	1998	1997	1996	1995	1994	1993				
重庆市城镇居民消费水平(元/人)	6190	5711	5646	4366	3658	2925	2262				
重庆市农村居民消费水平(元/人)	1351	1363	1388	1268	993	830	657				

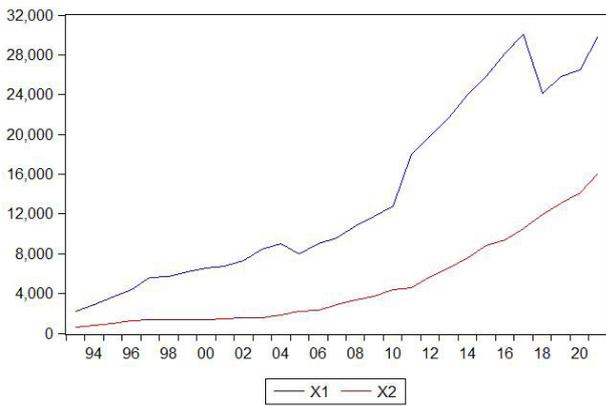


图 1 1993 年至 2021 年重庆市城镇和农村居民消费水平时序图
注: X1 为重庆市城镇居民消费水平(元/人), X2 为重庆市农村居民消费水平(元/人)

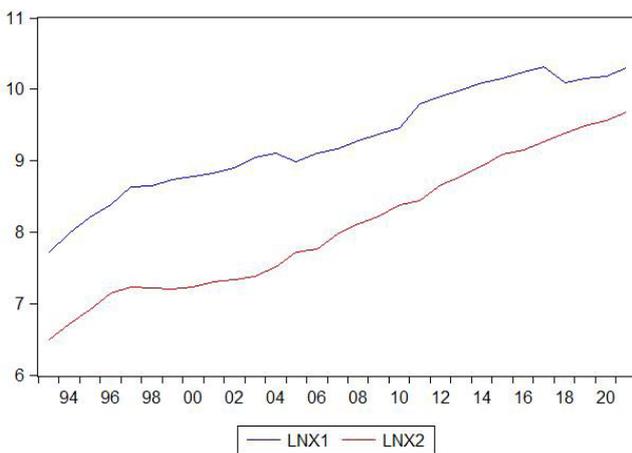


图 2 对数化后的时间序列图

注: LNX1 为重庆市城镇居民消费水平(元/人)的对数, LNX2 为重庆市农村居民消费水平(元/人)的对数。

LNX1 的自相关图如图 3 所示。

Date: 05/22/22 Time: 01:56
Sample: 1993 2021
Included observations: 29

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.873	0.873	24.477	0.000	
2	0.763	0.002	43.854	0.000	
3	0.661	-0.022	58.977	0.000	
4	0.569	-0.020	70.600	0.000	
5	0.477	-0.048	79.127	0.000	
6	0.379	-0.087	84.746	0.000	
7	0.287	-0.045	88.121	0.000	
8	0.199	-0.054	89.825	0.000	
9	0.112	-0.071	90.389	0.000	
10	0.028	-0.062	90.426	0.000	
11	-0.041	-0.017	90.512	0.000	
12	-0.089	0.024	90.929	0.000	

图 3 LNX1 的自相关图

可以看到原序列是含有截距和时间趋势, 根据 ADF 单位根检验, 可以得到原序列检验的 P 值均大于 0.05, 因此表明原序列是非平稳序列。

3.3.2 关于 ARIMA 模型的建模分析

可采用 ARIMA (p, d, q) 的模式进行拟合, 并通过对

数化、差分处理等使之更加稳定, 通过观察自相关与偏自相关图和单位根检验确定其均衡度。接着, 对已经均匀化了的数列使用 EViews10 软件进行其自相关曲线与偏自相关性图, 如若观测到负指数型的等差对比数列同时控制了自相关性和偏自相关性函数并且二者均不是截尾, 将迅速衰减。而对 ARMA (p, q) 模型来说, 论文由低到高对 p 和 q 的阶数进行尝试, 即设 (p, q) 为某个数值, 然后进行参数估计并固定模型, 验证这个模型是否通过检验, 如果没有就需要重新调整 (p, q) 值, 同理再次进行估计并进行检验, 直到模型被接受为止。

为了得到平稳可拟合模型的序列, 消除原序列 GDP 的趋势, 论文对序列 GDP 进行一阶自然对数差分。生成新的序列 R1、R2。然后绘制序列 R 的自相关图和 Q 统计量。

对 R1, 进行 ADF 单位根检验, 原序列的 ADF 检验的 P 值为 0.0016<0.05 通过检验, 在 1% 水平下的 t 值的绝对值大于 ADF 检验统计量。

对 R2, 进行 ADF 单位根检验, 原序列的 ADF 检验的 P 值为 0.0046<0.05 通过检验, 在 1% 水平下的 t 值的绝对值大于 ADF 检验统计量。

从检验结果来看, 论文认为两序列是平稳的, 因此可以描绘出序列 R1、R2 的自相关图, 最终得到序列 R1 的 q=7, p=7, 序列 R2 的 q=6, p=12。

3.3.3 模型的检验

如果综合考虑 AIC 准则和 SC 准则, 一般认为 AIC 准则的值越小, 模型越好。但如果样本容量趋于无穷大, 就说明利用 AIC 准则选择出的模型可能所含的未知参数的个数需要比真实情况多。基于此, SC 准则为了弥补 AIC 准则的不足而被提出, 它对 AIC 的改进修正了其对于未知参数个数的惩罚权重。因此, 将 AIC 或 SC 函数达到最小值的模型作为考察范围内的最优模型。充分考虑以上情况, 利用 AIC 准则与 SC 准则对拟合模型进行选择, 最终决定建立对城镇建立 ARIMA (7, 1, 7) 模型, 对农村建立 ARIMA (6, 1, 12) 模型。

函数模型表达式分别如下:

$$\begin{cases} (B)(\ln X)_t = -(1-B)(\ln X)_{t-7} + \varepsilon_t + \varepsilon_{t-7} \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_s \varepsilon_t) = 0, s \neq t \\ EX_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (2)$$

$$\begin{cases} (B)(\ln X)_t = 0.533479(1-B)(\ln X)_{t-6} + \varepsilon_t + 0.140188\varepsilon_{t-12} \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_s \varepsilon_t) = 0, s \neq t \\ EX_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (3)$$

根据 R1、R2 方程估算结果, 可以得到模型拟合较好, 预测数据比较准确。验证该模型即为求得的正确居民消费水平模型。

3.4 AMIRA 模型预测结果

既然已知模型比较准确, 可以用来预测位置数据。最后进行序列拟合和预测见表 2。

表 2 LNX1, LNN2 的预测数据

	LNX1	LNX2
2022	10.18462	9.733269
2023	10.18802	9.784868
2024	10.2354	9.885048
2025	10.24911	9.936723
2026	10.26805	9.989995

由表 2 可得 X1、X2 的预测, 见表 3。

表 3 X1、X2 的预测

	X1	X2
2022	26492.58	16869.61
2023	26582.81	17762.91
2024	27872.62	19634.59
2025	28257.38	20675.88
2026	28797.68	21807.19

注: X1 为重庆市城镇居民消费水平(元/人), X2 为重庆市农村居民消费水平(元/人)

3.5 渝双城经济圈的建立对五年后重庆居民消费的影响

从 ARIMA 模型的预测结果来看, 在未来五年, 城镇居民消费水平逐步趋于平稳, 农村居民消费水平逐步提升, 所

以我们推测成渝双城经济圈会从多个方面刺激城镇和农村居民消费。但由于政策实施会在提高乡村旅游业发展、促进农业领域下对农村居民收入的提高作用更大, 从而对农村居民消费影响较城镇居民影响更大。

4 结语

ARIMA 是 GDP 预测中的典型代表, 是预测领域的研究热点。论文收集了近年来重庆城乡居民的消费情况, 利用 ARIMA 模型进行分析, 进而预测渝双城经济圈的建立对五年后重庆居民消费的影响, 模型经验证, 拟合较好, 因此可以推断此次预测较为准确。

参考文献

- [1] 顾昕雨. 基于 ARIMA-SVR 组合模型的卫星遥感数据预测研究[D]. 北京: 中国科学院大学(中国科学院国家空间科学中心), 2021.
- [2] 王黎明, 王连, 杨楠. 应用时间序列分析[M]. 上海: 复旦大学出版社, 2009.
- [3] 李洁. 基于 ARIMA 模型对安徽省居民消费水平分析建模与预测[J]. 现代商业, 2017(1): 195-196.
- [4] 党爽. 利用 GM 与 ARIMA 组合模型对于陕西省城镇人口消费水平预测[D]. 昆明: 云南大学, 2017.