

doi:10.16055/j.issn.1672-058X.2015.0007.020

# 基于 ARIMA 建模的重庆市第三产业产值预测\*

徐 浩<sup>1</sup>, 刘宇琴<sup>2</sup>

(1.重庆工商大学 电子商务及供应链系统重庆市重点实验室,重庆 400067;2.重庆工商大学 管理学院,重庆 400067)

**摘 要:**随着经济的快速发展,第三产业的产值对经济的促进作用越发显得举足轻重,采取自回归求积移动平均法 (ARIMA) 模型在重庆市 2013 统计年鉴分析的基础上对第三产业的总值进行了 ARIMA 建模并进行预测,结果显示,ARIMA(1,2,2)模型对重庆第三产业产值提供了较为准确的预测结果,有较好的拟合程度,可用于未来的预测.

**关键词:**第三产业产值;ARIMA;预测

**中图分类号:**F224

**文献标识码:**A

**文章编号:**1672-058X(2015)07-0085-06

随着我国经济社会的快速发展,第三产业对于经济的发展越来越起到至关重要的核心作用.发达国家的 GDP 构成中,第三产业占到了 GDP 的 80%左右,而我国第三产业占国民经济的比重大约为 40%.重庆在 2013 年的统计中,第三产业占本市生产总值的比重只有 40.01%<sup>[1]</sup>,而天津、上海的第三产业占市生产总值的比重已经超过 60%,北京更是达到了 76%左右,大大超过重庆.《重庆十一五规划》明确提出,要加快发展以服务业为核心的第三产业,把“促进服务业全面发展和升级摆在突出位置,充分发挥其促进经济增长特别是促进就业的重要作用<sup>[2]</sup>”,因此研究第三产业产值的预测问题具有十分重要的意义.此处根据时间序列预测模型—ARIMA 建立了重庆市的第三产业预测模型,并有十分优异的拟合程度,为以后政府的决策提供了参考.

## 1 ARIMA 模型简介及建模

### 1.1 ARIMA 模型过程

ARIMA (autoregressive integrated moving average) 模型简称自回归求积移动平均模型,是 1970 年博克斯与詹金斯所提出.这种预测方法着重分析经济时间序列本身的概率或随机性质,而不在意构造单一方程或联立方程模型,其主要优点是对短期预测具有十分良好的效果<sup>[3]</sup>.张华初,林洪(2006)利用 ARIMA 模型对我国社会消费品零售额进行了预测,预测结果较好,并提出了扩大我国内需的建议<sup>[4]</sup>;池启水(2007)利用 ARIMA 模型对我国的石油消费量进行了预测,结果表明,ARIMA 模型很适合我国石油消费量的非平稳时间序列的特点,拟合较好<sup>[5]</sup>;龚国勇(2008)也利用 ARIMA 模型对深圳的 GDP 进行了预测,结果表明,ARIMA 模型可用于深圳国内生产总值短期预测,为深圳制定经济发展目标提供决策参考<sup>[6]</sup>;蒋艳(2010)运用 ARIMA 模型预测了广西固定资产投资,得出误差在 5%,可以接受<sup>[7]</sup>.此处选取 ARIMA 模型对第三产业产值

收稿日期:2014-10-22;修回日期:2014-11-25.

\* 基金项目:重庆市科委前沿与应用基础研究一般项目资助(cstc2014jcyjA00050).

作者简介:徐浩(1991-),男,重庆人,硕士研究生,从事经济及决策分析研究.

进行短期预测,建模的主要过程包括:

1) 自回归过程(AR).若考虑时间序列  $Y_t - \delta = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + u_t$ , 其中,  $\delta$  是  $Y$  的均值,  $u_t$  是一个白噪声, 则表明  $Y_t$  是一个  $p$  阶自回归, 或  $AR(p)$  过程;

2) 移动平均过程(MA).假定一个时间序列  $Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q}$ , 其中,  $\mu$  是常数,  $u$  和前面一样, 是随机误差项, 则被称为一个  $MA(q)$  过程;

3) 自回归移动平均过程(ARMA).当然, 一个时间序列可能兼有 AR 和 MA 的特性, 那么兼有的模型比如  $Y_t$  可以写为  $Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1}$ , 其中有一个自回归和移动平均项, 所以就是  $ARMA(1, 1)$  过程. 其中,  $\theta$  为常数项;

4) 自回归求积移动平均过程(ARIMA).在一般的经济时间序列中, 原始的序列基本都不是平稳的过程, 假设一个时间序列经过一阶差分后变为平稳的时间序列, 则这个时间序列  $Y_t$  是一阶单积的, 即  $Y_t \sim I(1)$ . 因此, 如果必须将一个时间序列差分  $d$  次, 将它变为平稳的, 然后用  $ARMA(p, q)$  作它的模型, 则那个原始的时间序列是  $ARIMA(p, d, q)$  时间序列.  $p$  指回归项数,  $d$  指变为平稳的差分次数,  $q$  指移动平均次数.

## 1.2 ARIMA 模型建立的一般方法

1) 原时间序列的平稳性检验, 检验的方法一般是用单位根检验, 若序列不满足平稳性条件, 则可以通过差分来变平稳;

2) 通过计算能够描述序列特征的一些统计量, 如自相关(ACF)系数和偏自相关(PACF)系数来确定  $ARMA(p, q)$  模型的阶数  $p$  和  $q$ , 并根据一定的准则如 AIC 等综合考虑来确定模型的参数;

3) 估计模型的未知参数, 并通过参数的 T 统计量检验其显著性, 以及模型的合理性;

4) 通过诊断分析, 检验模型的拟合值和实际值的残差序列是否为一个白噪声序列.

## 2 重庆市第三产业产值序列 ARIMA 建模

### 2.1 数据的获取及平稳性处理

从重庆统计年鉴上统计出 1978-2011 年重庆市第三产业总值, 见表 1.

表 1 1978-2011 年重庆市第三产业总值

年份	总值/亿元	年份	总值/亿元
1978	12.43	1995	366.20
1979	13.98	1996	458.57
1980	15.69	1997	552.14
1981	17.19	1998	625.85
1982	20.32	1999	679.23
1983	24.01	2000	746.10
1984	30.35	2001	840.01
1985	37.10	2002	956.12
1986	43.16	2003	1 081.35
1987	53.27	2004	1 229.62

续表1

年份	总值/亿元	年份	总值/亿元
1988	68.66	2005	1 440.32
1989	85.92	2006	1 649.20
1990	91.73	2007	1 825.21
1991	110.69	2008	2 160.48
1992	149.64	2009	2 474.44
1993	194.37	2010	2 881.08
1994	260.66	2011	3 623.81

数据来源:重庆统计年鉴2013.

首先进行数据的平稳性检验,使用 Eviews 6.0 作为检验软件,如表 2,对原始数据进行单位根检验.

表2 原始序列的 ADF(单位根)检验

ADF 检验统计量	临界值 $\alpha=1\%$	临界值 $\alpha=5\%$	临界值 $\alpha=10\%$
0.923 626	-3.689 194	-2.971 853	-2.625 125

由表 2 可知,原始序列的  $\tau$  统计量为 0.923 626,远远大于在 1%显著性和 5%显著性以及 10%显著性下的临界值,即接受  $H_0: \delta=0$  的原假设,即存在单位根,所以原始时间序列是非平稳的.

为了消除原序列的不平稳,对原始序列取自然对数,实验发现取一阶差分后在 1%,5%和 10%显著性下仍然不能通过 ADF 检验,即序列仍不是平稳的,故取二阶差分,记为  $\Delta^2 \ln Y_t$ ,再对取二阶差分的序列进行 ADF 检验,可见时间序列趋势基本消除,可以认为是一个平稳的时间序列,如表 3.

表3 二阶差分后的 ADF 检验结果

ADF 检验统计值	临界值 $\alpha=1\%$	临界值 $\alpha=5\%$	临界值 $\alpha=10\%$
-5.414 520	-3.670 170	-2.963 972	-2.627 007

表 3 中,经过二阶差分后序列的  $\tau$  值为 -5.414 520,均通过 1%显著水平、5%显著性水平以及 10%显著性水平的单位根检验,时间序列是平稳的,即序列为二阶单整序列,  $\Delta^2 \ln Y_t \sim I(2)$ . 此模型中的  $d=2$ ,即为二阶单整的时间序列.

## 2.2 模型识别

模型识别的主要工具是自相关函数 (autocorrelation function, ACF) 和偏自相关函数 (partial autocorrelation function, PACF) 以及由此可得的相关图(图 1).

由图 1 的偏自相关函数(PACF)可以看出,  $p=1$  或  $p=2$  都可以,由自相关函数(ACF)图可以看出,自相关图  $q=1$  或  $2$  均可.利用最佳准则函数(AIC)可知,在 ARMA(1,1), ARMA(1,2), ARMA(2,1) 和 ARMA(2,2) 中,ARMA(1,2) 的 AIC 和 SC 值最小,即赤池信息量和施瓦茨准则最小,故 ARMA(1,2) 为最优的模型,如表 4.

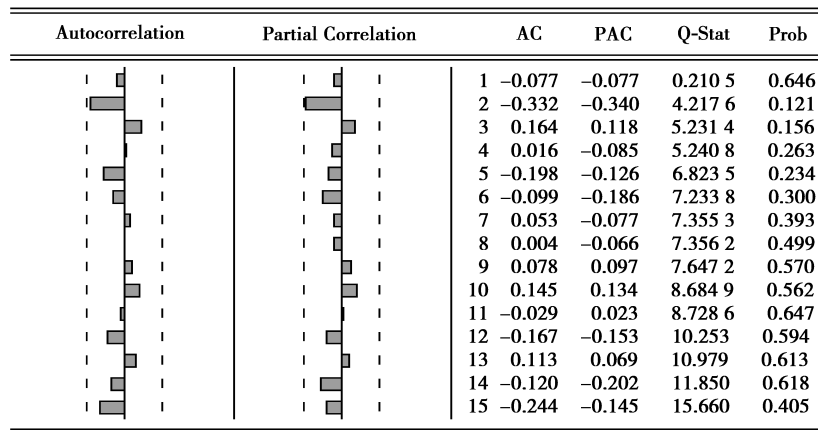


图 1  $\Delta^2 \ln Y_t$  的自相关及偏自相关图

表 4 4 种模型的拟合估计结果比较

模型	变量	估计的系数	T 值	AIC 值	SC 值	R <sup>2</sup>																																						
ARMA(1,1)	AR(1)	0.56	3.57	-2.76	-2.62	0.17																																						
	MA(1)	-0.94	-27.94				ARMA(1,2)	AR(1)	0.019	0.08	-2.83	-2.65	0.27	MA(1)	-0.21	-1.34	MA(2)	-0.72	-4.62	ARMA(2,1)	AR(1)	-0.64	-2.27	-2.69	-2.51	0.19	AR(2)	-0.34	-1.81	MA(1)	0.66	2.49	ARMA(2,2)	AR(1)	-0.64	-1.33	-2.62	-2.39	0.19	AR(2)	-0.34	-0.81	MA(1)	0.66
ARMA(1,2)	AR(1)	0.019	0.08	-2.83	-2.65	0.27																																						
	MA(1)	-0.21	-1.34																																									
	MA(2)	-0.72	-4.62				ARMA(2,1)	AR(1)	-0.64	-2.27	-2.69	-2.51	0.19	AR(2)	-0.34	-1.81	MA(1)	0.66	2.49	ARMA(2,2)	AR(1)	-0.64	-1.33	-2.62	-2.39	0.19	AR(2)	-0.34	-0.81	MA(1)	0.66	1.29		MA(2)	-0.000 1	-0.002								
ARMA(2,1)	AR(1)	-0.64	-2.27	-2.69	-2.51	0.19																																						
	AR(2)	-0.34	-1.81																																									
	MA(1)	0.66	2.49				ARMA(2,2)	AR(1)	-0.64	-1.33	-2.62	-2.39	0.19	AR(2)	-0.34	-0.81	MA(1)	0.66	1.29		MA(2)	-0.000 1	-0.002																					
ARMA(2,2)	AR(1)	-0.64	-1.33	-2.62	-2.39	0.19																																						
	AR(2)	-0.34	-0.81																																									
	MA(1)	0.66	1.29																																									
	MA(2)	-0.000 1	-0.002																																									

如表 4, ARMA(1,2)的 AIC 值和 SC 值都最小,且 R<sup>2</sup> 在 4 个模型中最优,所以选择 ARMA(1,2)模型作为估计模型,且差分次数为 2,故最后的模型为 ARIMA(1,2,2).

### 2.3 ARIMA(1,2,2)模型的拟合和参数估计

建立 ARIMA(1,2,2)模型如下:

$$\Delta^2 \ln Y_t = \theta + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \alpha_1 \Delta^2 \ln Y_{t-1}$$

通过 Eviews 软件的最小二乘法估计出方程的参数:

$$\theta = -0.002\ 071, \alpha_1 = \text{AR}(1) = 0.019\ 526, \beta_1 = \text{MA}(1) = -0.201\ 365\ 9, \beta_2 = \text{MA}(2) = -0.720\ 009$$

即

$$\Delta^2 \ln Y_t = -0.002\ 071 + 0.019\ 526 \Delta^2 \ln Y_{t-1} - 0.201\ 365\ 9 u_{t-1} - 0.720\ 009 u_{t-2}$$

### 2.4 模型的诊断

图 2 为残差序列的前 12 阶自相关图(ACF)和偏自相关图(PACF).

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.137	-0.137	0.542 8	0.461
		2	-0.073	-0.094	0.705 7	0.703
		3	-0.135	0.164	1.285 3	0.733
		4	-0.117	-0.179	1.735 7	0.784
		5	-0.200	-0.305	3.118 8	0.682
		6	0.076	-0.106	3.329 0	0.767
		7	-0.133	-0.318	4.011 0	0.779
		8	0.253	0.021	6.608 7	0.579
		9	0.056	-0.057	6.742 5	0.664
		10	-0.009	-0.113	6.746 0	0.749
		11	0.046	0.033	6.848 0	0.811
		12	-0.102	-0.146	7.386 7	0.831
		13	0.053	0.151	7.541 2	0.872
		14	-0.014	-0.002	7.552 9	0.911
		15	-0.129	-0.061	8.647 6	0.895
		16	0.001	-0.019	8.647 7	0.927
		17	0.119	0.044	9.791 9	0.912
		18	-0.146	-0.107	11.738	0.860

图 2 残差序列的自相关图和偏自相关图

从图 2 可以看出,自相关图和偏自相关图的相关函数均落在置信区间内,可以认为白噪声过程,同时,对残差序列进行 ADF 检验.

表 5 残差序列的单位根检验

ADF 检验统计量	临界值 a=1%	临界值 a=5%	临界值 a=10%
-5.518 592	-3.727 04	-2.786 225	-2.632 604

由上述检验可知,通过模型的拟合程度较好,可以进行下一步的分析.

### 2.5 模型的预测

由 ARIMA(1,2,2)模型:

$$\Delta^2 \ln Y_t = -0.002 071 + 0.019 526 \Delta^2 \ln Y_{t-1} - 0.201 365 9u_{t-1} - 0.720 009u_{t-2}$$

又由  $\Delta^2 \ln Y_t = \ln Y_t - 2 \ln Y_{t-1} + \ln Y_{t-2}$ , 可得  $\ln Y_t$  的预测公式为

$$\ln Y_t = 2 \ln Y_{t-1} - \ln Y_{t-2} - 0.002 071 + 0.019 526 \Delta^2 \ln Y_{t-1} - 0.201 365 9u_{t-1} - 0.720 009u_{t-2}$$

最后预测的结果如表 6.

表 6 ARIMA(1,2,2)模型预测值

年份	2008	2009	2010	2011	2012
实际值/亿元	2160.48	2474.44	2881.08	3623.81	4494.41
预测值/亿元	2101.371	2530.986	2812.526	3487.614	4389.26
预测误差	-2.73%	2.28%	-2.37%	-3.75%	-2.34%

通过模型预测,可以发现除了 2011 年的预测误差百分比在-3.75%外,其余的误差百分比都小于 3%,预测的精度很好,预测精度在 98%左右,表明了 ARIMA(1,2,2)模型对于重庆市第三产业的产值具有良好的预测效果.

### 3 结束语

通过对 1978-2011 年重庆市第三产业产值序列进行分析,运用 ARIMA(1,2,2)模型对第三产业的产值进行了预测,不仅预测结果的准确性较高,而且预测结果的稳定性很好,可用为重庆市制定第三产业发展战略规划提供决策参考.在今后的研究中,可以将更多的预测技术作为方法,形成组合预测法,如 ARIMA 模型与灰色系统理论的结合,ARIMA 模型与神经网络的结合<sup>[5]</sup>,ARIMA 模型与 GARCH 法结合等,运用更为先进和现代的预测方法对其进行预测.

#### 参考文献:

- [1] 曾国平,刘娟.重庆第三产业发展水平差异分析[J].商业研究,2009(11):4-7
- [2] 达摩达尔·古扎拉蒂.计量经济学基础[M].北京:中国人民大学出版社,2009
- [3] 张华初,林洪.我国社会消费品零售额 ARIMA 预测模型[J].统计研究,2006(7):58-60
- [4] 池启水.中国石油消费量增长趋势分析—基于 ARIMA 模型的预测与分析[J].资源科学,2007,29(5):69-73
- [5] 龚国勇.ARIMA 模型在深圳 GDP 预测中的应用[J].数学的实践与认识,2008,28(4):53-57
- [6] 蒋燕.ARIMA 模型在广西全社会固定资产投资预测中的应用[J].数理统计管理,2006,25(5):588-592
- [7] 吕一清,何跃.基于灰色神经网络的第三产业发展趋势的预测模型[J].统计与决策,2011(4):157-159
- [8] 刘勇,汪旭晖.ARIMA 模型在我国能源消费预测中的应用[J].经济经纬,2007(5):11-13
- [9] 赵廷.ARMA 在我国 GDP 预测中的应用[J].中国市场,2011(1):60-62
- [10] 张蔚,张彦琦.实践序列资料 ARIMA 季节乘积模型及其应用[J].第三军医大学学报,2002,24(8):955-957
- [11] 万丽娟,徐孝勇.西部大都市区县第三产业发展水平差异分析—以重庆市为例[J].重庆大学学报:社会科学版,2010,22(16):6-10
- [12] 郎茂祥.预测理论与方法[M].北京:清华大学出版社,2011

## The Forecasting for the Output of the Tertiary industry in Chongqing based on ARIMA Modeling

**XU Hao<sup>1</sup>, LIU Yu-qin<sup>2</sup>**

(1.ECSC, Chongqing Technology And Business University, Chongqing 400067, China;

2. School of Management, Chongqing Technology And Business University, Chongqing 400067, China)

**Abstract:** With the rapid development of the economy, the role of the tertiary industry's output becomes more and more important to the development of economy. With ARIMA modeling, this paper forecasts the output of the tertiary industry of Chongqing, based on the 2013 Chongqing statistical yearbook. The results implies that ARIMA(1,2,2) provides the accurate result and has good fitting degree, so it can be used to predict the future. Therefore, it can provide reliable theoretical support for the prediction for the output of the tertiary industry of Chongqing.

**Key words:** output of the tertiary industry ; ARIMA; forecast