

ARIMA 模型在云南省固定资产投资预测中的应用

郭利京, 孙雪莲

(云南农业大学 经济贸易学院, 云南 昆明 650201)

摘要: 采用自回归求积移动平均 (ARIMA) 法, 对云南省历年固定资产投资额进行了分析。结果显示, ARIMA (4, 1, 4) 模型提供较准确的预测效果, 可用于未来的预测, 并为今后几年云南省全社会固定资产投资提供可靠依据。

关键词: ARIMA 模型; 固定资产投资额; 时间序列; 预测

中图分类号: F 127 **文献标识码:** A **文章编号:** 1004-390X (2007) 01-0063-05

The Application of ARIMA Model in Yunnan Province's Fixed Investments Forecasting

GUO Li-jing, SUN Xue-lian

(College of Economy Trade Academy, Yunnan Agriculture University, Kunming 650201, China)

Abstract: This paper applies ARIMA method to analyse data of the volume of investment in fixed assets of Yunnan Province. It shows the forecast of ARIMA (4, 1, 4) model is exact. It can be used to forecast, and provide the reliable basis for investing in fixed assets of Yunnan Province.

Key words: ARIMA; fixed assets amount of capital invested; time series; forecasting

云南省全社会固定资产投资往往受到许多因素的制约, 且这些因素之间又保持着错综复杂的联系, 因此, 运用结构性的因果模型分析和预测往往比较困难。本文从另一角度出发, 认为云南省全社会固定资产投资额是一时间序列, 因此可以根据过去的资料得出其变化规律, 并用此来预测未来的发展变化。但是, 为了准确预测云南省固定资产投资的发展趋势, 使所建模型既满足实际的要求, 也满足统计方法理论的要求, 有必要先对传统的时间序列分析方法 (简称为传统法) 和自回归求积移动平均法 (简记 ARIMA), 也普遍称它为 Box—Jenkins 法 (简称 B—J) 进行比较和选择。

一、传统法和 ARIMA 法的比较与选择

传统的时间序列分析在经济中的应用, 主要是确定性的时间序列分析方法, 包括指数平滑法、

滑动平均法、趋势预测法、趋势季节模型预测法、时间序列的分解等等。

1970 年, Box 和 Jenkins 提出了以随机理论为基础的时间序列分析方法, 其基本模型有三种: 自回归 (AR) 模型; 移动平均 (MA) 模型以及自回归求积移动平均 (ARIMA) 模型。自回归 (AR) 模型和移动平均 (MA) 模型实际上是自回归求积移动平均 (ARIMA) 模型的特例。

传统法和 ARIMA 法都属于时间序列分析法, 都是通过分析现象 (变量) 随时间而发展变化的特征, VAR (变量) 的历史资料建立时间序列模型 (或带回归项的时间序列模型) 的方法。两种方法的主要区别在于:

1. 分析的前题条件不同

传统法假定时间序列的资料存在着某个确定的模式, 随机变量相对来说并不显著。它事先假

定有一个由历史数据所表现的固有模式，除此之外还表现有某种偶然性。可以根据历史数据得出某种确定的函数，然后用此函数去分析事物的发展趋势。而 ARIMA 法假定数据序列是由某个随机过程产生的，它把事物在某一固定时刻的状态视为一个随机过程，利用随机过程去分析描述事物的发展趋势。

2. 适用环境不同

传统法与 AMIMA 法相比，原理简单易懂，适合于具有某种典型趋势特征变化的社会经济现象的预测。但是在实际应用中，由于一些社会经济现象的时间序列资料并不总是具备这种确定型趋势特征，一定程度上限制了传统法的应用。此外，这种方法产生的模型中的误差项可能不完全满足随机性质，从而影响模型的预测效果。

ARIMA 法由于不需要对时间序列的发展模式作先验的假设，同时方法的本身保证了可通过反复识别修改，直到获得满意的模型，因此适合于各种类型的时间序列数据，包括在辨别时间序列资料的典型特征十分困难和复杂情况下的预测。这种方法不仅考察预测变量的过去值与当前值，同时对模型同过去值拟合产生的误差也作为重要因素进入模型，有利于提高模型的精确度，是一种精确度相当高的短期预测方法。

3. 建模的基本思想不同

传统法建模的基本思想是：认为事物的变化是渐进式而不是跳跃式的，影响事物的因素在过去、现在和将来基本上是不变的，即使变化也不大，事物的发展具有稳定性和类推法。根据时间序列的历史资料可以采用定性和定量相结合的方法描述出这种确定型的趋势，并依此来预测将来的发展变化。

ARIMA 法建模的基本思想是：将预测对象随时间推移而形成的数据序列视为一个随机序列，即除去个别的因偶然原因引起的观测值外，时间序列是一组依赖于时间 t 的随机变量，构成该时序的单个序列值虽然具有不确定性，但整个序列的变化却有一定的规律性，可以用相应的数学模型近似描述。这组随机变量所具有的依存关系或自相关性表征了预测对象发展的延续性，而这种自相关性一旦被相应的数学模型描述出来，就可以从时间序列的过去值及现在值预测其未来的值。

从历年《云南统计年鉴》上搜集计算整理

出 1978—2004 年的年末全社会固定资产投资总额资料（见表 1）。

表 1 云南省全社会固定资产投资总额 单位：亿元

年份	x	年份	x	年份	x
1978	15.042 7	1987	54.379	1997	540.503
1979	16.778 4	1989	67.767 6	1998	672.541 4
1980	20.889 5	1990	67.798 5	2000	664
1981	18.434 4	1991	75.744 6	2001	684
1982	22.675 5	1992	98.315 5	2002	738.5
1983	22.808 8	1993	140.689 2	2003	814.6
1984	30.965 9	1994	251.400 4	2004	1 000.1
1985	46.276 9	1995	321.734 1	2005	
1986	15.042 7	1996	380.568 8		

数据来源：《新中国五十年统计资料汇编》和《2005 年云南统计年鉴》，本文所用云南省全社会固定资产投资总额数据未消除通货膨胀影响

表 1 中 x 表示全社会固定资产投资总额（亿元），先运用 Eviews5.0 软件中 View 下面的 Line graph 对 x 作出折线图（图 1），再运用 View 下面的 Unit Root Test 对 x 进行单位根检验。

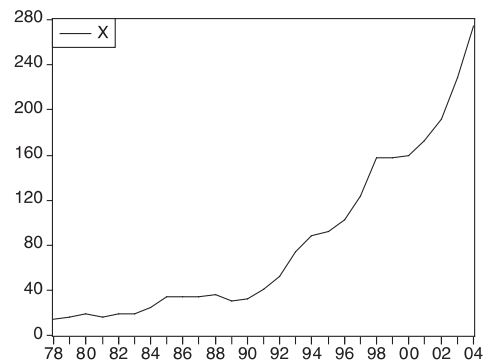


图1 序列 x 时序图

从图 1 可以看出，该折线图向右上方倾斜，并且其前后波动的幅度是不一致的，这说明此时间序列存在增长趋势，又存在异方差性，而且从表 2 的单位根检验的数据来看： X 没有通过扩充迪基·富勒（Augmented Diekey. Fuller，简称 ADF）的平稳性单位根检验，因此，该时间序列是非平稳的时间序列，很难识别云南省全社会固定资产投资总额变化的固定模式。即使采用传统法拟合直线或指数曲线等趋势模型都存在有自相关，预测结果与实际情况存在很大的差距，故选择 ARIMA 法进行预测。

表 2 x 序列单位根检验结果

ADF Test Statistic	2.93	1% Critical value	-3.74
		5% Critical value	-2.99
		10% Critical value	-2.64

二、ARIMA 模型的建模思想

ARIMA 模型 (p, d, q) 又称为自回归求积移动平均模型。其中 AR 指自回归, p 为模型的自回归项数; MA 为移动平均, q 为模型的移动平均项数; I 指积分, d 为时间序列成为平稳之间必须取其差分的次数。其一般的表达式为:

$$y_t = a_0 + a_1 Y_{t-1} - l + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + \beta_0 u_t + \beta_1 u_{t-1} + \dots + \beta_q + \mu_{t-q}$$

其建模基本思想可归纳如下:

1. 根据时间序列的散点图, 自相关函数 (ACF) 图和偏自相关函数 (PACF) 图, 以及 ADF 单位根检验, 观察其方差、趋势及其季节性变化规律, 识别该序列的平稳性;

2. 数据进行平稳化处理。如果数据序列是非平稳的, 如存在一定的增长或下降趋势等则需对数据进行差分处理; 如果数据序列存在异方差性, 则需对数据进行对数转换或者开方处理, 直到处理后的数据的自相关函数值和偏相关函数值无显著地异于零;

3. 根据时间序列模型的识别规则, 建立相应的模型: (1) 若平稳时间序列的偏相关函数是截

尾的, 而自相关函数是拖尾的, 则可断定此序列适合 AR 模型; (2) 若平稳时间序列的偏相关函数是拖尾的, 而自相关函数是截尾的, 则可断定此序列适合 MA 模型; (3) 若平稳时间序列的偏相关函数和自相关函数均是拖尾的, 则此序列适合 ARMA 模型;

4. 进行参数估计, 估计暂定的模型参数, 检验是否具有统计意义;

5. 进行假设检验, 诊断白噪声。检验假设模型残差的 ACF 值和 PACF 值在早期或季节性

延迟点处不得大于置信区间, 同时残差应理想化为 0 均值。可观察残差的 ACF 图和 PACF 图, 并辅以 D—W 值、t 值等检验法。

三、ARIMA 模型的应用

1. 取 d。从图 1 和表二的单位根检验可以看出, 该时间序列是非平稳的时间序列, 因此, 需要对数据进行平稳化处理, 经过逐次尝试, 最后确定先对数据取自然对数, 然后进行一阶差分处理。处理得到的数据见表 3。

所以取 d = 1。

2. 下面对 P 和 q 进行筛选。做出时间序列 {z} 的折线图 (图 2), 直到滞后 12 期的 ACF 值和 PACF 值 (表 4), 以及 ADF 单位根检验 (表 5)

表 3 对表 1 数据取自然对数后进行一阶差分处理结果

年份	X _t	Y _t	Z _t	U _t	年份	X _t	Y _t	Z _t	μ _t
1978	15.04	2.710893			1993	251.40	5.527047	0.580494	0.0359
1979	16.77	2.820092	0.1092		1994	321.73	5.773725	0.246679	-0.014
1980	20.88	3.039247	0.219154	0.086	1995	380.57	5.941667	0.167942	0.0018
1981	18.43	2.914218	-0.12503	-0.076	1996	448.02	6.104846	0.163179	-0.044
1982	22.67	3.121285	0.207067	-0.025	1997	540.50	6.2925	0.187654	-0.014
1983	22.80	3.127146	0.005861	-0.288	1998	672.54	6.511064	0.218563	0.0112
1984	30.96	3.432887	0.30574	0.1243	1999	664.00	6.498282	-0.01278	-0.102
1985	46.27	3.834643	0.401756	0.0733	2000	684.00	6.527958	0.029676	-0.03
1986	49.91	3.910394	0.075751	-0.008	2001	738.5	6.604621	0.076663	-0.050
1987	54.37	3.995978	0.085584	-0.004	2002	814.6	6.702697	0.098076	0.0243
1988	67.76	4.216084	0.220106	0.001	2003	1000.1	6.907855	0.205158	0.0676
1989	67.79	4.21654	0.000456	0.1434	2004	1244.3	7.126328	0.218473	-0.010
1990	75.74	4.327367	0.110827	0.0613	2005	1743.0	7.463363	0.337035	0.1031
1991	98.31	4.588182	0.260815	0.0154	2006	2220.5	7.705488	0.242125	0.048
1992	140.69	4.946553	0.358372	0.0956	2007				

数据来源:《新中国五十年统计资料汇编》和《2007 年云南统计年鉴》及计算而得

注: xt 为云南省全社会固定资产投资总额, yt = logxt, zt = yt - yt - 1, μt 为残差项

表 4 序列 z 的自相关和偏自相关值

ACF	0.148	-0.021	-0.226	-0.401	-0.095	0.158	0.052	0.313	0.112	-0.086	-0.036	0.006
PACF	0.148	-0.044	-0.222	-0.362	-0.027	0.142	-0.146	0.196	0.113	-0.028	0.054	0.248

表 5 序列 z 的单位根检验结果

ADF Test Statistic	-3.88	1% Critical value	-3.77
		5% Critical value	-3.00
		10% Critical value	-2.64

从时间序列 {z} 的折线图 (图 2), 看出序列 z 的趋势消除; 又从 ACF 值和 PACF 值 (表 4) 看出样本的自相关值和偏相关值很快地落入置信区间, 故序列的趋势已基本消除; 还有从表 5 得到时间序列通过扩充 ADF 单位根检验因此, 时间序列 {z} 是平稳的。从序列 {z} 窗口中 views 下的相关性检验, 可以看出样本的自相关图和偏自相关图都是拖尾的, 故选取 ARMA 模型, 经过反复筛选, 取 P=4, q=4。

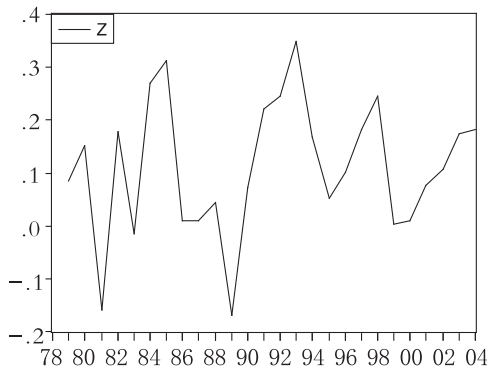


图 2 序列 z 折线图

3. 模型的估计

经过多次反复拟合得到时间序列 {z} 较理想的模型, 如下:

$$Z_t = 0.12 - 0.53 * Z_{t-4} + 0.91 * \mu_{t-4}$$

[0.02] [0.15]
[0.13]

(4.84) (-3.54)
(7.16)

4. 模型的诊断

根据时间序列 {z} 的模型, 我们可以对其进行回归拟合, 并得到拟合值和残差图 (图 3)。对残差进行检验, 得到残差扩充的 ADF 单位根检验结果。

由图 3 我们可以看出, 时间序列 {z} 的实际值和拟合值总的来说比较接近, 残差值较小。对方程残差进行相关性检验可知其自相关值和偏

相关值基本上在置信区间内, 与零无显著差异。从表 6 得到残差通过扩充 ADF 单位根检验。所以残差通过白噪声检验。

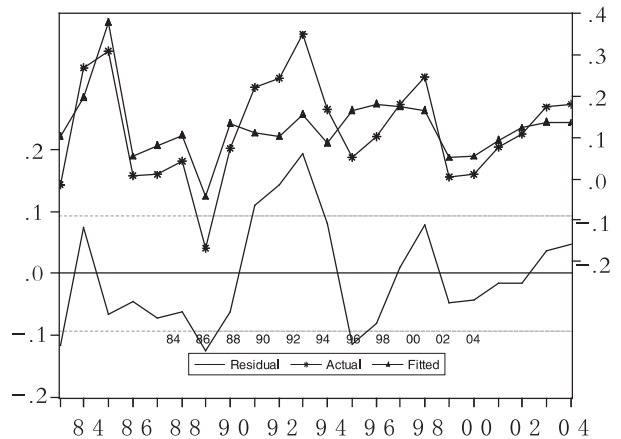


图 3 序列 z 的实际值、拟合值和残差

表 6 残差序列相关性检验

ADF Test Statistic	-5.059262	1% Critical	-2.664853
		5% Critical	-1.955681
		10% Critical	-1.608793

而且, 模型的检验效果比较好, 其中 Akaike 信息量准则统计量 AIC = -1.88, 贝叶斯准则统计量 SCB = -1.734, 均方差估计 V. B = 0.172, 标准误差 S. E = 0.089 等, 所以可以诊断该模型是可行的, 可用于预测。

四、预测和分析

根据时间序列 {z} 的 ARIMA (4, 1, 4) 模型:

$Z_t = 0.12 - 0.53 * Z_{t-4} + 0.91 * \mu_{t-4}$, 我们可以推出时间序列

{Y} 的 ARIMA (4, 1, 4) 的预测公式为:

$$y_t = y_{t-1} + z_t = y_{t-1} + 0.12 - 0.53 * Z_{t-4} + 0.91 * \mu_{t-4} \tag{2}$$

进而推出时间序列 {x_t} 的 ARIMA (4, 1,

4) 预测公式为:

$$x_t = e^{y_{t-1} + 0.12 - 0.53 * z_{t-4} + 0.91 \mu_{t-4}} \tag{3}$$

对 {x_t} 作出 2002 - 2006 年的预测值 (亿

元), 与实际值的比较见表 7, 如下:

表 7 2002-2006 年的固定资产预测值和实际值

年份	2004	2005	2006	2007	2008
预测值	1251.2	1767.3	2169.7	2753.42	3304.1
实际值	1244.3	1743.0	2220.5		

1. 从表 7 可以看出预测和实际值的差异较小, 说明模型预测的效果较好, 可以用于预测。从模型的公式我们也可以看到, 云南省全社会固定资产投资额与其第 4 期的滞后值、随机扰动项的第 4 期滞后密切相关。从参数估计值来看, 与固定资产投资额的第 4 期滞后值负相关, 与随机扰动项的第 4 期滞后正相关。因此, 政府在引导投资时要注意这一点, 以免社会投资不当, 影响经济健康发展。

2. 该模型在短期内预测比较准确, 随着预测的延长, 预测误差会逐渐增大。这也是

ARIMA 模型的一个缺陷。但尽管如此, 与其它的预测方法相比, 其预测的准确度还是比较高的, 尤其在短期预测方面。

五、结 论

固定资产投资是经济增长的原动力, 它既能增加供给又能增加需求, 对经济增长具有乘数效

应。同时, 由于加速数原理, 产出的增加又会引致投资的扩大, 两者在经济增长中相互促进、相互制约, 总体上呈现同步发展趋势。并且, 从较长一段时期的社会经济发展过程来看, 投资增长与经济增长之间存在相互联系、相互作用的内在本质与要求。

本文根据 1978 年至 2006 年云南省全社会固定资产投资的数据, 运用自回归求积移动平均法 (ARIMA), 分析显示 ARIAM (4, 1, 4) 模型提供较准确的预测效果, 可用于未来的预测。通过预测得知, 在今后几年中云南省全社会固定资产投资应该保持持续增加的势头, 才能为全省经济持续、快速、稳定增长, 提供可靠的保证。

[参考文献]

- [1] 易丹辉. 数据分析与 Eviews 的应用 [M]. 北京: 中国统计出版社, 1994.
- [2] 李子奈. 计量经济学 [M]. 北京: 高等教育出版社, 2000.
- [3] 张晓峒. 计量经济学软件 Eviews 使用指南 [M] (第二版). 天津: 南开大学出版社, 2004.
- [4] 高铁梅. 计量经济分析方法与建模 Eviews 应用及实例 [M]. 北京: 清华大学出版社, 2006.
- [5] 蒋燕. ARIMAM 模型在广西全社会固定资产投资预测中的应用 [J]. 数理统计与管理, 2006 (5).

(责任编辑: 许 敏)