

文章编号:1672-058X(2014)12-0034-04

GDP 时间序列的 ARIMA 模型研究

冯 瑞

(重庆工商大学 数学与统计学院,重庆 400067)

摘 要:在经典计量经济学建模过程中,通常假定经济时间序列是平稳的,而且主要以某种经济理论或对某种经济行为的认识来确定计量经济学的模型理论的关系式.然而在经济领域中,许多时间序列数据不是由平稳过程产生的,基于此,研究了国内生产总值 GDP 随时间位移而持续增长的特性,确定了模型的自回归阶数,建立了 ARIMA 模型,并对 ARIMA 模型进行了检验,确定了模型的平稳性与模型自回归影响的持久性.

关键词:非平稳;GDP;ARIMA 模型

中图分类号:O212

文献标志码:A

ARIMA 模型全称为差分自回归移动平均模型(Autoregressive Integrated Moving Average Model),简记 ARIMA,是由博克思(Box)和詹金斯(Jenkins)提出的著名时间序列预测方法,又称为 Box-Jenkins 模型、博克思-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.所谓 ARIMA 模型,是指将非平稳时间序列转化为平稳时间序列,然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型.ARIMA 模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程(MA)、自回归过程(AR)、自回归移动平均过程(ARMA)以及 ARIMA 过程^[1].

GDP 是国民经济核算的核心指标,也是衡量一个国家或地区经济状况和发展水平的重要指标.从价值形态看,它是所有常住单位在一定时期内所生产的全部货物和服务价值超过同期投入的全部非固定资产货物和服务价值的差额,即所有常住单位的增加值之和;从收入形态看,它是所有常住单位在一定时期内所创造并分配给常住单位和非常住单位的初次分配收入之和;从产品形态看,它是最终使用的货物和服务减去进口货物和服务.国内生产总值 GDP 是核算体系中一个重要的综合性统计指标,也是中国新国民经济核算体系中的核心指标.它反映一国(或地区)的经济实力和市场规模.一个国家或地区的经济究竟处于增长抑或衰退阶段,从这个数字的变化便可以观察到.当 GDP 的增长数字处于正数时,即显示该地区经济处于扩张阶段;反之,如果处于负数,即表示该地区的经济进入衰退时期了.国内生产总值是指一定时间内所生产的商品与劳务的总量乘以“货币价格”或“市价”而得到的数字,即名义国内生产总值,而名义国内生产总值增长率等于实际国内生产总值增长率与通货膨胀率之和.一般情况,GDP 会呈现强烈的趋势特征^[2].这些具有趋势特征的经济变量,当发生经济震荡或者冲击后,一般会出现两种情形,一是受到震荡后,经济变量逐渐又回到它们的长期趋势,二是这些经济变量没有回到原有轨迹,而呈现游走状态^[3].若研究的经济变量遵从一个非平稳过程,运用最小二乘法理论是不成立的.然而 GDP 常作为模型的重要变量,许多预测模型依赖于 GDP 数据,GDP 的预测有非常重要的意义.基于以上的原因,许多研究者致力于对 GDP 的估计.此处用 ARIMA 模型对全国的 GDP 进行估计和预测^[4],希望对关注中国经济发展的人们有所借鉴.

收稿日期:2014-04-05;修回日期:2014-05-30.

作者简介:冯瑞(1988-),男,重庆人,硕士研究生,从事经济与社会发展统计方法及应用研究.

1 ARIMA(p, d, q) 模型建立过程

1.1 数据来源与平稳性检验

通过查阅《中国统计年鉴》,得到 1979-2012 年,以当年价格表示的国内生产总值(GDP)序列,如表 1 所示.

1978 年我国实行的改革开放政策对国民经济发展产生了至关重要的影响,1979 年是改革之后的一年,改革的影响效果更能直接反应到 GDP 上,所以选择 1979-2012 年的数据,并对上述数据取对数得到 1979-2012 年 GDP 的对数值 $\log(\text{GDP})$,画出 $\log(\text{GDP})$ 的趋势如图 1.

从图 1 可以看出,对数 GDP 曲线呈直线上升的趋势;对数 GDP 序列具有一定的非线性趋势性.因此从图形可以初步判断该序列是非平稳的.对对数 GDP 的平稳性进行进一步 ADF 检验(表 2).

从表 2 中可以看出,在 1%,5%,10%3 个显著性水平下,单位根检验的临界值分别为 -3.679 32, -2.967 77, -2.622 99, T 统计量的值为 -1.270 35,大于相近临界值,从而无法拒绝原假设,表明对数 GDP 的差分序列存在单位根.因此,该时间序列是非平稳的时间序列,即使采用传统法拟合直线或指数曲线等趋势模型都存在自相关,预测结果与实际情况存在很大的差距.

为消除时间序列的非平稳性影响,将对数 GDP 的一阶差分记为 $D\log(\text{GDP})$,表 3 为对数 GDP 的一阶差分的 ADF 检验^[5].

从表 3 中可以看出,在 1%,5%,10%3 个显著性水平下,单位根检验的临界值分别为 -3.679 32, -2.967 77, -2.622 99, T 统计量的值为 -3.656 84,基本小于相近临界值,并且 P 值小于 0.05,从而拒绝原假设,表明对数 GDP 的差分序列不存在单位根.因此,该时间序列是平稳的时间序列.通过上述的检验表明对数 GDP 是一阶单整的,即对数 GDP 的一阶差分 $D\log\text{GDP} \sim I(1)$.

表 2 对数 GDP 的 ADF 检验

	T 统计量	P 值
扩大后的迪克福勒检定统计量	-1.270 35	0.629 4
检验的临界值	1%	-3.679 32
	5%	-2.967 77
	10%	-2.622 99

表 1 1979-2012 国民总收入 亿元

年份	国民总收入	年份	国民总收入
1979	4 062.579	1996	70 142.49
1980	4 545.624	1997	78 060.85
1981	4 889.461	1998	83 024.28
1982	5 330.451	1999	88 479.15
1983	5 985.552	2000	98 000.45
1984	7 243.752	2001	108 068.2
1985	9 040.737	2002	119 095.7
1986	10 274.38	2003	134 977
1987	12 050.62	2004	159 453.6
1988	15 036.82	2005	183 617.4
1989	17 000.92	2006	215 904.4
1990	18 718.32	2007	266 422
1991	21 826.2	2008	316 030.3
1992	26 937.28	2009	340 320
1993	35 260.02	2010	399 759.5
1994	48 108.46	2011	468 562.4
1995	59 810.53	2012	516 282.1

表 3 对数 GDP 一阶差分 ADF 检验

	T 统计量	P 值
扩大后的迪克福勒检定统计量	-3.656 84	0.010 6
检验的临界值	1%	-3.679 32
	5%	-2.967 77
	10%	-2.622 99

1.2 ARIMA 模型 p, q 值的确立

既然对数 GDP 的一阶差分是平稳的时间序列,则阶数 $d=1$,现在确定 ARIMA(p, d, q) 中的 p, q 值,利用 Eviews 画出对数 GDP 的自相关图(图 2):

从图 2 中可以看到,对数 GDP 的自相关图形与偏相关图形均呈现拖尾的势态(从图形上看,自相关与偏自相关的波动逐渐减小并没有突然收敛到临界值水平范围内;从数值上看,自相关与偏自相关的数值慢慢减少),观察自相关与偏自相关的波动图形可知对数 GDP 的自相关函数一阶都是显著的(超过图中虚线部分),并且从第二阶开始下降,数值也不太显著,因此先设定 q 值为 1.对数 GDP 的偏自相关函数也是一至四阶都较显著,从第五阶开始下降很大,此后一直收敛在临界值水平范围内.因此先设定 p 的值为 1~4.因为是一阶差分所以 d 取 1.

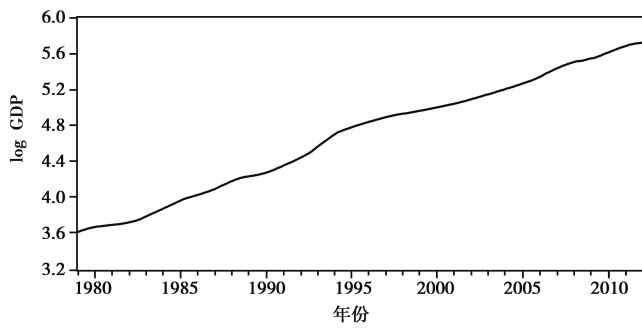


图 1 log(GDP) 的趋势图

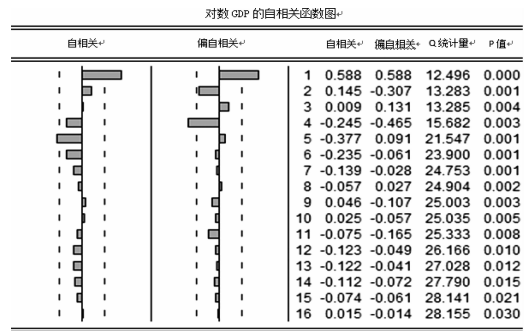


图 2 对数 GDP 的自相关函数图

模型参数估计一般分两步进行,首先找出参数的初步估计,又称矩估计,然后在初步估计的基础上,根据一定准则,求得参数在某种意义下的精确估计.通常用尤利-沃克方程作参数的矩估计,用最小二乘法和极大似然估计法作参数的精确估计.

表 4 不同 p, q 取值下的 ARIMA 模型

		T 统计量	P 值	AIC 值
	C	11.732 94	0	
ARIMA(4,1,1)	AR(4)	-0.729 02	0.472 5	-4.865 6
	MA(1)	5.325 286	0	
	C	9.911 892	0	
ARIMA(3,1,1)	AR(3)	0.165 098	0.870 1	-4.881 76
	MA(1)	5.584 689	0	
	C	9.379 398	0	
ARIMA(2,1,1)	AR(2)	0.418 934	0.678 5	-4.904 94
	MA(1)	5.714 066	0	
	C	8.209 676	0	
ARIMA(1,1,1)	AR(1)	1.089 81	0.284 8	-4.953 87
	MA(1)	3.881 995	0.000 6	
	C	10.667 71	0	
ARIMA(0,1,1)	MA(1)	6.243 559	0	-4.985 13
	C	17.617 37	0	
ARIMA(4,1,0)	AR(4)	-1.493 61	0.146 9	-4.412 42

由重新拟合的时间序列方程,可以看出AIC统计量数值为-4.985 13,较低,ARIMA(p, d, q)在 p 取 0, q 取 1 的时候较优,而且方程的 T 统计量也通过检验,可决系数 R 也基本满足要求,现进行白噪声检验,对残差进行白噪声检验(图 3).

由图 3 可知, Q 统计量的 P 值均是大于 0.05 的,所以时间序列残差项通过白噪声检验,因此最终建立的模型是 ARIMA(0,1,1).对数 GDP 的 ARIMA 模型如下:

$$D \log \text{GDP} = 0.063\ 573\ 429\ 782\ 3 + \varepsilon_t - 0.769\ 158\ 166\ 187 * \varepsilon_{t-1}$$

1.3 ARIMA 模型预测

通过模型进行预测,用 Eviews 测算出评价预测的一些标准,如表 5 所示.可以看到,希尔不相等系数为 0.19,表明模型的预测能力比较好,而它的方差比例较大,说明实际序列的波动较大,而模拟序列的波动较小,这可能是由于预测时间过长.

通过模型预测,预计 2013 年 GDP 为 550 404.4 亿元,2013 年实际 GDP 为 568 845.21 亿元,相差 3.24%,误差较小,模型拟合较好.

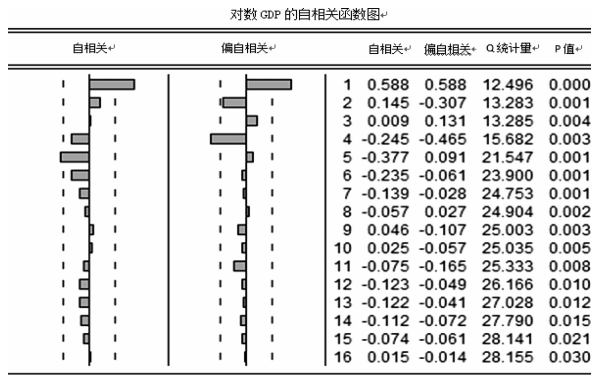


图 3 残差白噪声检验

表 5 ARIMA 模型预测效果

观测样本/年	1979-2013
观测值个数	33
均方根误差	0.025 704
平均绝对误差	0.020 911
平均绝对误差比例	39.001 68
希尔不等系数	0.194 255

2 结 语

经典计量经济学建模过程中,通常假定经济时间序列是平稳的,而且主要以某种经济理论或对某种经济行为的认识来确定计量经济学的模型理论的关系式,借用此形式进行数据收集.然而在经济领域中许多时间序列数据不是由平稳过程产生的,此处研究的国内生产总值 GDP 随时间的位移而持续增长,由于在实际中遇到的时间序列数据很可能是非平稳序列,而平稳性在计量经济学建模中又具有重要的地位,因此此处对对数 GDP 时间序列进行了单位根检验,确定了模型的平稳性,得到了对数 GDP 的差分阶数,由此建立了 ARIMA 模型,并根据对数 GDP 时间序列的自相关确定了 ARIMA 中 p, q 的值,最终得到对数 GDP 的 ARIMA(0, 1, 1) 模型,并对 ARIMA(0, 1, 1) 进行了白噪声检验,确定模型残差项的信息已提取完全,证实了模型调整后的平稳性与模型自回归影响的持久性.

参考文献:

- [1] 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(4): 53-57
- [2] 孙合超. ARMA 算法在 GDP 预测中的应用[J]. 金田, 2013(6): 389-390
- [3] 杨茜. 基于 ARMA 模型对我国 GDP 季度数据的建模[J]. 行政事业资产与财务, 2011(6): 105-106
- [4] 高铁梅. 计量经济分析方法与建模[M]. 北京: 清华大学出版社, 2006
- [5] 靳庭良. DF 单位根检验的势及检验式的选择[J]. 统计与决策, 2005(10): 13-17

Research on ARIMA Model of GDP Time Series

FENG Rui

(School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: In the process of classic econometric modeling, the correlated formula of econometrics model theory is determined by supposing that economic time series is stable and mainly by certain economic theory or by the understanding of certain economic behavior, however, Gross Domestic Product (GDP) studied in this paper is continuously growing with the time displacement, the auto-regression order number of the model is determined, based on this, ARIMA Model is set up, the ARIMA Model is tested then, and finally the stability of the model and the duration of the influence of the model auto-regression are determined.

Key words: non-stability; GDP; ARIMA Model

责任编辑:李翠薇