

SARIMA 模型的建模及其信贷预测分析

夏 天 程细玉

(华侨大学商学院, 福建 泉州 362021)

摘要 对自回归单整移动平均季节模型(SARIMA 模型)的原理,以及建模思想进行诠释.指出在经济数据中普遍存在的季节性,并在 ARIMA 模型基础上提出 SARIMA 模型.通过对中国人民银行的月度信贷总量资料的建模及预测分析,得到良好的效果. SARIMA(1, 1, 0)(1, 1, 1)¹²模型这一短期预测模型及其短期预测的结果,可为中国人民银行进行信贷政策的制订提供依据.

关键词 SARIMA 模型, 随机时间序列, 银行信贷总量, 预测

中图分类号 O 212 : F 830. 5

文献标识码 A

20 世纪 80 年代以来,以经济理论为基础建立经济计量模型的传统结构建模方法受到重大挑战.“数据驱动”式的建模方法以其预测的准确性,富于动态性而大受欢迎,其中如单位根检验,单整理论等不仅改变了传统经济计量学建模思想,并在经济学金融学中得到广泛的应用. SARIMA 模型作为这些先进统计计量思想的重要代表形式,更是在预测方面显示出其良好的应用性与合理性.中国人民银行在制定货币政策及金融运行控制的实际操作过程中,总要涉及到信贷预测的问题,信贷预测要受到许多经济因素的制约,而经济因素之间又保持着错综复杂的关系,表现为极大的不确定性^[1]. 本文运用建立在随机过程理论基础上的 SARIMA 模型,首次将该模型应用于中国人民银行月度信贷总量预测,并获得了良好的效果.

1 SARIMA 模型的建模原理

自回归单整移动平均季节模型(SARIMA 模型)^[2~4]实际是源于自回归单整移动平均模型(ARIMA 模型)^[5]. ARIMA 模型的具体表达形式是 ARIMA(p, d, q),其中,AR 指自回归过程,MA 指移动平均过程, p 是自回归项数, q 是模型的移动平均项数, I 是单整, d 为时间序列由非平稳成为平稳序列需要差分的次数.当 $d = 0$ 时,ARIMA 模型即变为 ARIMA(p, q). 在考虑了经济时间序列的大多数非平稳性的情况下,通过引入差分项 d 及滞后算子 I , ARIMA(p, d, q) 模型的一般表达式为

$$\left. \begin{aligned} \Phi(I)(1-I)^d y_t &= \theta(I)u_t, \\ u &\cong \text{IID}(0, \sigma^2). \end{aligned} \right\} \quad (1)$$

当模型的随机时间序列存在季节性的非平稳性问题时,就要应用自回归单整移动平均季节模型,即 SARIMA(p, d, q)(P, D, Q)^s. 其表达式为

$$\left. \begin{aligned} \Phi_p(I)\Phi_q(I^s)(1-I)^d(1-I^s)^D y_t &= \theta_q(I)\theta_p(I^s)u_t, \\ u &\cong \text{IID}(0, \sigma^2). \end{aligned} \right\} \quad (2)$$

在式(2)中, P 是季节自回归阶数, Q 是季节移动平均阶数,并且分别称 $\Phi(I^s)$ 与 $\theta_q(I^s)$ 为季节 P 阶自回归算子和季节 Q 阶移动平均算子.当差分平稳序列 $(1-I)^d(1-I^s)^D y_t$ 建模时,阶数 P 与 Q 可能为零.通过上面几个模型具体表达形式的分析,可以得到

收稿日期 2005-12-15

作者简介 夏 天(1979-),男,硕士研究生,主要从事金融数量分析与金融工程的研究;通信作者:程细玉(1964-),男,教授, E-mail: changxiyu@hqu.edu.cn

基金项目 国务院侨务办公室科研基金资助项目(01QZR07)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

$$\text{SARIMA}(p, d, q)(P, D, Q)^s = \begin{cases} \text{AR}(p), & p \neq 0, \quad q = 0, \quad d = 0, \quad s = 0, \\ \text{MA}(q), & p = 0, \quad q \neq 0, \quad d = 0, \quad s = 0, \\ \text{ARIMA}(p, d, q), & s = 0, \\ \text{ARIMA}(p, q), & p \neq 0, \quad q \neq 0, \quad d = 0, \quad s = 0, \\ \text{白噪声}, & p = 0, \quad q = 0, \quad d = 0, \quad s = 0. \end{cases} \quad (3)$$

从式(3)中可以看到, SARIMA 模型与其他时间序列模型之间的关系. SARIMA 模型应该是表述最为全面的时间序列预测模型, 其他的模型都可以由它简化变形后得到. SARIMA 模型建模的思想更为一般的表述是: 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, 时间序列是一组依赖时间 t 的随机变量, 构成该时间序列的单个序列值虽然具有不确定性, 但整个序列的变化却是有一定的规律性, 可以用数学模型近似描述. 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, 而一旦这种自相关性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来预测其未来值^[6]. 当然, 其前提是该随机时间序列一定是平稳的, 它的趋势性与季节性都通过一次差分或季节差分的方法予以消除了.

2 数据的平稳性及建模

2.1 数据分析

本文运用 Eviews 4.0 软件进行建模. 数据选取了中国人民银行 1999 年 12 月至 2005 年 7 月的信贷额度(L)总量数据. 首先来对原始数据 x_t 进行观察(图 1), 然后对 x_t 进行扩充迪基-富勒(ADF)的平稳性单位根检验, 如表 1 所示. 从图 1 中可以看到, 该折线向右上方倾斜, 并且前后波动具有不一致性, 说明时间序列存在异方差性及增长趋势与季节趋势. 而由表 1 可以看到, x_t 没有能够通过平稳性单位根检验. 因此, 可以断定该时间序列 x_t 为非平稳的时间序列.

表 1 原始数据 ADF 检验表

| ADF 检验流计量 | 1% 关键值 | 5% 关键值 | 10% 关键值 |
|------------|------------|------------|------------|
| 2. 298 183 | - 2. 603 3 | - 1. 946 3 | - 1. 618 8 |

2.2 建模过程

在对原始数据 x_t 建模之前, 首先要对原始数据 x_t 的异方差性进行消除, 这就需要对原始数据做对数化处理. 然后应用自相关函数(ACF)与偏自相关函数(PACF)及扩充迪基-富勒(ADF)的平稳性单位根来检验 x_t 的趋势性及季节性规律, 即识别 x_t 的平稳性, 如图 2 所示. 从图 2 可以看到, 对于取完对数的原始数据的自相关函数(ACF)表现为缓慢的拖尾, 而它的偏自相关函数(PACF)则表现为截尾. 因而, 可以得到初步的结论是该时间序列是非平稳的. 进行差分化处理, 即对取过对数的原始数据进行一价差分, 此时得到 $d=1$, 如图 3, 4 所示. 从

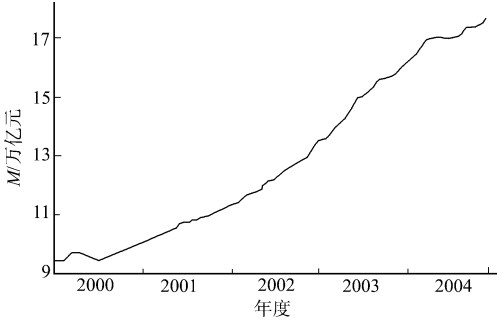


图 1 信贷总量趋势图

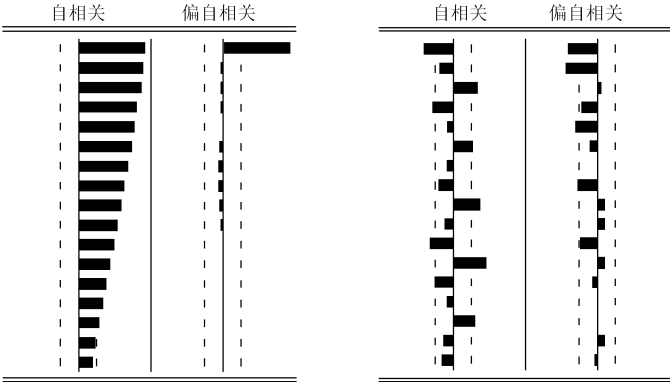


图 2 差分前的 ACF 与 PACF 图

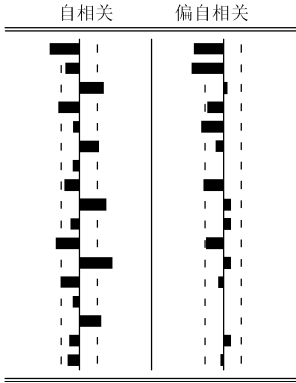


图 3 差分后的 ACF 与 PACF 图

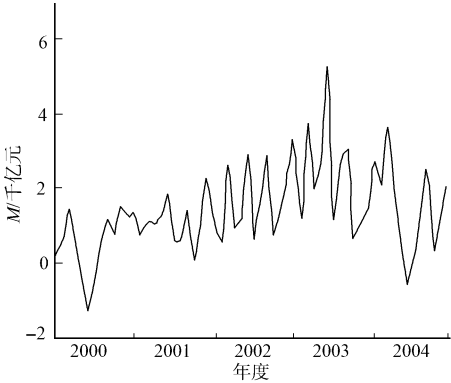


图 4 信贷差分图

图 3, 4 可以看到, 通过一价差分处理以后的 x_t 已经消除了时间序列的趋势性, 通过对其进行扩充迪基-

富勒(ADF)的平稳性单位根检验,结果如表 2 所. 从表 2 可以发现,在 10% 的水平下该时间序列可以通过单位根的检验,而从图 3,4 中可以看到,经过一阶差分后的时间序列 x_t 依然保存着明显的季节性. 这是因为在图 4 中的每年的时间间隔中总会有一个大的波峰出现;而图 3 在 $k=12$ 时表现为 ACF 显著不为零,说明时间序列 x_t 存在较为明显的季节性. 这其实在经济问题的月度数据、季节数据中是普遍存在的. 在一次差分的基础上进行季节性差分的处理方法($s=12$),结果如图 5,表 3 所示. 从图 5 中可以看到,该时间序列 x_t 通过季节性差分后已经消除了季节性,通过观察最后平稳的自相关函数(ACF)与偏自相关函数(PACF)图可以判定时间序列已经是平稳的了. 对季节性差分时间序列的单位根检验来看,在 5% 的水平下就可以通过单位根检验了. 说明经过上面的差分处理已经实现了时间序列的平稳化.

表 2 一阶差分后的 ADF 检验表

| ADF 检验流计量 | 1% 关键值 | 5% 关键值 | 10% 关键值 |
|------------|-----------|-----------|-----------|
| - 1.802 59 | - 2.602 6 | - 1.946 2 | - 1.618 7 |

表 3 季节性差分后的 ACF 检验表

| ADF 检验流计量 | 1% 关键值 | 5% 关键值 | 10% 关键值 |
|------------|-----------|-----------|-----------|
| - 3.546 27 | - 4.178 1 | - 3.513 6 | - 3.186 8 |

2.3 模型的确立

在时间序列实现了平稳化以后,就要建立基于中国人民银行信贷数据的 $SARIMA(p,d,q)(P,D,Q)^s$ 模型. 这里就要有个判定 p 与 q 的问题. 在前面时间序列平稳化的过程中得出,进行了一阶差分与季节性差分可以有 $d=1,s=12,D=1$. 由于在图 5 中的 $k=12$ 处自相关系数与偏自相关系数不显著,从而得出 $P=1,Q=$

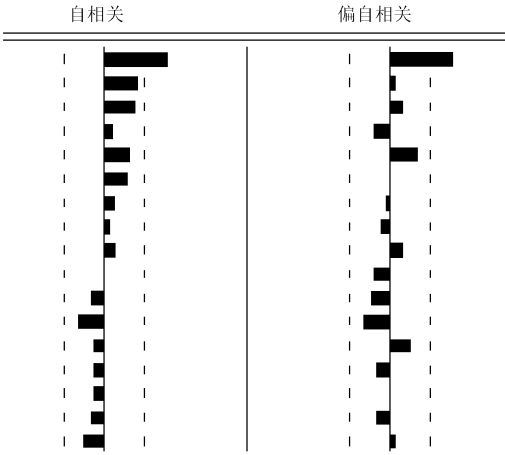


图 5 季节性差分后的 ACF 与 PACF 图

1. 再根据图 3,5 基本可以选定 (p,q) 的选择范围是 $(1,0),(2,0),(3,0),(3,1)$ 这 4 个值. 分别对这 4 个选择取值代入到模型中,结果如表 4 所示,从中比较可得最佳选择. 首先,通过对表 4 中的 4 个模型的

表 4 模型判别表

| 序号 | (p,q) | R^2 | AIC | SIC | LM |
|----|---------|-----------|-----------|-----------|---------|
| 1 | (2,0) | 0.557 630 | - 7.082 6 | - 6.903 0 | 0.592 0 |
| 2 | (3,1) | 0.552 144 | - 7.033 1 | - 6.761 0 | 0.732 7 |
| 3 | (3,0) | 0.554 108 | - 7.061 7 | - 6.835 0 | 0.445 2 |
| 4 | (1,0) | 0.555 099 | - 7.133 0 | - 6.999 7 | 0.894 1 |

可调整 R^2 , AIC, SIC 的值的大小比较,观察得出 $(1,0)$ 应该是最好的选择. 依据计量 AIC 的值及 SIC 的值却是最小的. 因此,可以基本认为 $(1,0)$ 是最好的选择,这也跟前面图 2 的结论向吻合. 再从 LM 值来看,作为 SARIMA 模型中自相关问题的最好检验方法. 以上 4 个模型的 LM 值都是大于置信度 0.05 的,从而判定 4 个模型都不存在自相关的问题,而 $(1,0)$ 在这方面的值是最大的. 因此,它还是最佳的选择. 综上所述,可以得出对于中国人民银行的月度信贷总量数据来说, $SARIMA(1,1,0)(1,1,1)^{12}$ 模型是最好的预测模型,即

$$(1+0.3362l^{12})(1-0.4347l)(1-l)(1-l^{12})\log(L)=(1+0.8342l^{12})u_t.$$
 (4)

2.4 模型的检验

建立模型后,还必须对模型进行检验. 对 SARIMA 模型,可应用 Box-Pierce 假设检验法^[7]进行检验. 其原理是检验该模型的残差是否是白噪声,Box-Pierce 假设检验的检验统计量 Q 值也可以通过对残差序的 Correlogram Q -Statistics 中得到. 在零假设下, Q 服从 $\chi^2_{(m-p-q)}$ 分布,给定置信度 $1-\alpha$ (α 一般取 0.05),有

$$Q \leq \chi^2_{(m-p-q)}.$$
 (5)

如果式(5)不能拒绝残差序列相互独立的原假设,可以得出此次检验通过;否则,就不能通过检验. 通过求得 $Q=6.9837<\chi^2_{0.05}(m-p-q)=26.296$,可知式(5)是成立的. 从而可以肯定该 SARIMA 模型通过

了检验.

2.5 模型的预测

对时间段 2005 年 1 月~ 2005 年 7 月进行动态预测, 具体的预测模拟效果, 如表 5 所示. 表 5 给出了最后用于预测的 7 个月的真实值(R)与预测值(F)的实际数值, 可以用模型评价指标进一步来进行评估, 其中最常用评估指标是 Theil 系数与 MAPE 值.

表 5 SARIMA 模型预测结果表

| 时间 | 2005 年 1 月 | 2005 年 2 月 | 2005 年 3 月 | 2005 年 4 月 | 2005 年 5 月 | 2005 年 6 月 | 2005 年 7 月 |
|-----|------------|------------|------------|------------|------------|------------|------------|
| R | 181 083.0 | 182 042.3 | 185 461.3 | 186 889.1 | 186 274.1 | 186 178.7 | 185 853.1 |
| F | 180 109.4 | 181 302.3 | 185 001.6 | 186 974.4 | 189 134.1 | 183 094.4 | 193 959.5 |

通过计算出 4 个模型的 Theil 系数, 其中 SARIMA(1, 1, 0)(1, 1, 1)¹²模型的 Theil 系数是 0.011 844. 它是前面所列的 4 个模型 Theil 系数最低且最接近于零的, 而它的 MAPE 值是 1.621 517, 这也是 4 个模型中最小的一个值. 从预测评估指标的结果上也可以断定, SARIMA(1, 1, 0)(1, 1, 1)¹²模型对中国人民银行月度信贷预测可以达到最佳的预测结果. 当然, 该模型也有其不足的地方, 即 SARIMA 模型应用在预测的时候, 对短期预测有着比较好的预测效果, 但随着时间的延长, 它呈现出较差的预测效果^[8].

3 结束语

本文介绍了一种先进的经济计量预测模型——SARIMA 模型, 并对该模型的原理, 建模的思想与过程进行了较为全面的阐述. 同时, 将 SARIMA 模型应用于中国人民银行的月度信贷总量数据, 得到了 SARIMA(1, 1, 0)(1, 1, 1)¹²这一预测模型形式及良好的短期预测结果.

参 考 文 献

1 戴根有. 走向货币政策间接调控——中国的实践与国外的经验[M]. 北京: 中国金融出版社, 1999. 43~ 44

2 Ghysels E, Lee H S, Noh J. Testing for unit roots in seasonal time series[J]. Journal of Econometrics, 1994, 62(3): 415~ 424

3 Hylleberg S. Modeling seasonal variation, in Hargreaves, C P. (ed.) Nonstationary time series analysis and cointegration[M]. Oxford: Oxford University Press, 1994. 54~ 66

4 Miron J A. The Economics of seasonal cycles[M]. Massachusetts: MIT Press, 1996. 176~ 189

5 Box G E P, Jenkins G M. Time series analysis forecasting and control[M]. San Francisco: Holden Day, 1970. 112~ 176

6 易丹辉. 数据分析与 Eview s 应用[M]. 北京: 中国统计出版社, 2002. 112~ 113

7 张世英, 樊 智. 协整理论与波动模型——金融时间序列分析与应用[M]. 北京: 清华大学出版社, 2004. 136~ 141

8 约翰斯顿 J. 计量经济学方法[M]. 唐齐鸣, 等译. 北京: 中国经济出版社, 2002. 119~ 120

The Study of SARIMA Model and the Application
in Loan Forecast

Xia Tian Cheng Xiyu

(College of Commerce, Huaqiao University, 362021, Quanzhou, China)

Abstract This article explain the principle and the modeling idea of the seasonal autoregressive integration moving average(SARIMA), and point out that there exists the seasonal problem in the economic data universally. Based on this ARIMA model, SARIMA model have been proposed. This short term forecast model and the outcome of the SARIMA(1, 1, 0)(1, 1, 1)¹² model, good in forecasting the monthly credit data, will provide the support for the credit policy of the People's Bank of China.

Keywords SARIMA model, random time series, the volume of loan, forecast