

乘积季节模型在储蓄存款预测中的应用

茹正亮,印凡成,张 燕

(河海大学理学院,江苏 南京 210098)

摘要 运用 SPSS 软件和 SAS 软件系统中的时间序列建模方法建立了我国城乡居民储蓄存款模型,并认为用最大似然估计法(ML)对结果进行短期预测,用无约束最小二乘估计法(ULS)对结果进行中长期预测,可得到较高的预测精度.

关键词 时间序列;自相关函数;偏自相关函数;Box-Jenkins 建模;储蓄存款预测

中图分类号 O213 **文献标识码** A **文章编号** 1000-1980(2006)06-0724-03

20 世纪 90 年代中期以来,城乡居民储蓄存款逐年稳定增长,占金融机构资金来源的 45% 以上,占金融机构全部存款的 50% 以上.城乡居民储蓄的快速增长有力地支持了工农业生产和经济建设,为社会投资提供有力的资金保证,也是社会公众对我国金融体系稳健运行充满信心的表现.稳定的储蓄资金保证了银行体系的正常经营和支付,对社会稳定和金融安全具有重要意义.

本文运用国际著名大型集成软件 SPSS 和 SAS 对 1993~1999 年城乡居民储蓄存款的月余额建立乘积季节模型并进行预测.分别用最大似然估计法(ML)、无约束最小二乘估计法(ULS)和条件最小二乘估计法(CLS)进行计算,发现对于短期预测来说,最大似然估计法(ML)结果较准确,4 个月的平均相对误差为 0.5563%,对于中长期预报来说,无约束最小二乘估计法(ULS)结果较准确,4 个月的平均相对误差为 0.5074%.

1 乘积季节模型

数学上纯 ARIMA(p, d, q)模型记作 $\Phi(B) \nabla^d X_t = \Theta(B) \varepsilon_t$ (1)

式中: t ——时间; X_t ——响应序列; B ——后移算子; $\nabla = 1 - B$; p, d, q ——自回归阶数、差分阶数和移动平均阶数; $\Phi(B)$ ——自回归算子,记作 $\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$; $\Theta(B)$ ——滑动平均算子,记作 $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$; ε_t ——独立扰动或随机误差.

乘积季节模型是随机季节模型与 ARIMA 模型的结合式,其阶数为 $(p, d, q) \times (P, D, Q)_S$,其形式为

$$\Phi(B) U(B^S) \nabla^d \nabla_S^D X_t = V(B^S) \Theta(B) \varepsilon_t \quad (2)$$

式中: $\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$; $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$; $U(B^S) = 1 - u_1 B^S - u_2 B^{2S} - \cdots - u_P B^{PS}$, S 值是一个季节循环中观测的个数; $V(B^S) = 1 - v_1 B^S - v_2 B^{2S} - \cdots - v_Q B^{QS}$. $\Phi(B) \nabla^d X_t$ 表示同一周期内不同周期点的相关关系, $U(B^S) \nabla_S^D$ 则描述了不同周期的同一周期点上的相关关系,二者结合起来便同时刻画了两个因数的作用^[1].

2 模型建立的步骤

基于全国月度数据^①,1993 年 1 月到 2000 年 12 月城乡居民储蓄存款的月余额是一个时间序列,应用 Box-Jenkins 的经典时序建模方法建立城乡居民储蓄存款模型(为便于比较选用 1993 年 1 月到 1999 年 11 月的数据).建模具体步骤为:(a)数据的预处理(平稳化);(b)模型的识别;(c)定阶和参数估计;(d)适应性检验;(e)模型的预测^[2-4].

2.1 数据的预处理(平稳化)

a. 如果一个序列的平均值和方差始终为常数,则称它为平稳的.在估计时间序列模型之前需把不平稳

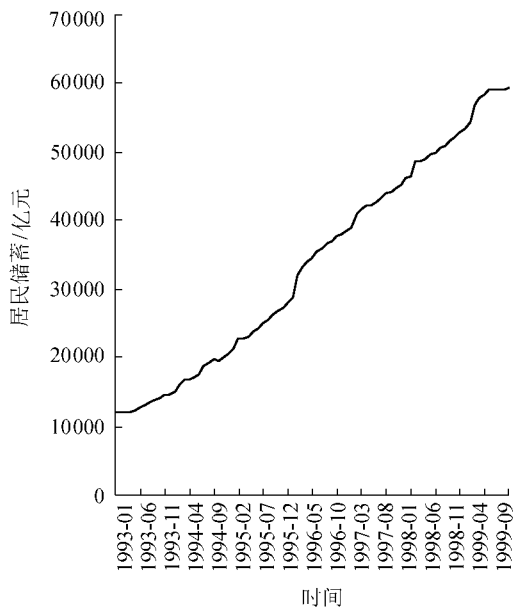


图 1 数据序列
Fig.1 Graph of data sequence

的时间序列转化为平稳序列.评估一个时间序列的平稳性可通过数据序列图和自相关函数及其图形进行.如果数据序列图呈现线性或二次趋势,则时间序列是不平稳的.如果自相关函数在前面少数几个值后下降为0,则序列是平稳的.如果在前几个值后,自相关函数没有下降为0,而是逐次减少,则序列不平稳.用 SPSS 软件对原数据作数据序列图(图 1),可以看出数据序列不平稳.

b. 若断定一个时间序列是不平稳的,通常可通过下列方法把它化为平稳的.(a)如果序列呈线性趋势,均值不平稳,利用一阶差分将产生一个平稳序列.(b)如果序列呈现二次趋势,均值不是常数,利用二阶差分可将其化为平稳序列.(c)如果随时间变化序列出现偏差,方差不是常数,则通常可利用自然对数将其转化为平稳序列.

对原始数据^①进行一阶自然对数逐期差分后,发现自相关图和偏自相关图存在周期为 12 个月的季节波动,所以对原始数据进行季节差分.从自相关图和偏自相关图可以看出季节波动已消除,通过计算其均值为 - 0.000 33,均值标准误差为

0.002 46,经检验均值与 0 没有显著差异^[5-6].

2.2 模型的识别、定阶和参数估计

在模型识别阶段,笔者用 SAS 软件中 IDENTIFY 程序对原始数据进行计算,发现序列经过一阶对数逐期差分后,其递增趋势基本消除,故 $d = 1$.经过一阶季节差分,季节性基本消除,故 $D = 1$,所以选用乘积季节模型 $(p, d, q) \times (P, D, Q)_s$.通过自相关图和偏自相关图可以看出 $p = (1\ 4)$ 或 $q = (1\ 4)$.当 $p = (1\ 4)$ 时,信息量准则统计量 AIC 为 - 386.133,贝叶斯准则统计量 SBC 为 - 379.387;当 $q = (1\ 4)$ 时,信息量准则统计量 AIC 为 - 383.226,贝叶斯准则统计量 SBC 为 - 376.481.由此可知,当 $p = (1\ 4)$ 时,其结果和实际值相差很小.

2.3 模型的适应性检验

模型的适应性检验,即模型的残差序列 ε_t 的独立性检验.若残差序列不是白噪声序列,说明残差序列中还存在有用的信息未被提取,需对原模型做进一步改进.本文用残差序列的卡方检验,公式为

$$\chi^2_m = n(n + 2) \sum_{k=1}^m \frac{r_k^2}{n - k} \tag{3}$$

其中

$$r_k^2 = \frac{\sum_{t=1}^{n-k} \varepsilon_t \varepsilon_{t+k}}{\sum_{t=1}^n \varepsilon_t^2}$$

经计算可以看出,残差自相关与 0 无明显差异,表明残差序列是相互独立的^[2].

2.4 模型的预测

用 SAS 软件中的 GPLOT 作图,分别作出实际值图、预测值图及置信上限和置信下限(图 2),预测值与实际观测值比较,两者基本吻合,这表示模型模拟效果较好.图 2 是用最大似然方法得出的原始数据拟合图和趋势图.图 2 中竖线左边为 1999 年 12 月之前城乡居民储蓄存款的拟合值,右边为 12 月之后城乡居民储蓄存款的预测值.

表 1 为用乘积季节模型预测的储蓄存款.从表 1 可见,最大似然方法的相对误差相对来说最小,无约束最小二乘法的相对误差其次,条件最小二乘法的相对误差最大,它们的均值分别是 0.507 4,0.556 3,0.908 0.

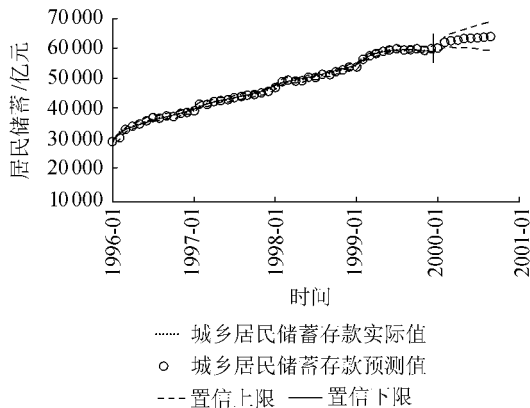


图 2 拟合图和趋势图

Fig.2 Curve fitting for deposit and its variation trend

① 北京经济信息网. <http://www2.beinet.cn/dpcnm.asp>.

表 1 用乘积季节模型预测的储蓄存款
Table 1 Predicted result of deposit by multiply season method

时间	实际值/亿元	ML 预测值/ 亿元	ML 相对 误差/%	ULS 预测值/ 亿元	ULS 相对 误差/%	CLS 预测值/ 亿元	CLS 相对 误差/%
2000-01	60 242.00	60 263.24	0.035 3	60 300.95	0.097 9	60 328.30	0.143 3
2000-02	62 270.00	62 596.64	0.524 6	62 628.52	0.575 8	62 743.88	0.761 0
2000-03	62 492.00	62 901.67	0.655 6	62 822.16	0.528 3	63 190.80	1.118 2
2000-04	62 536.00	63 167.42	1.009 7	63 053.48	0.827 5	63 542.50	1.609 5

3 结 论

本文所建模型能够对城乡居民储蓄存款余额进行短期和中长期预测,能够为金融机构把握未来城乡居民储蓄存款余额走势提供参考.

对于短期预报来说,采用最大似然方法计算出的结果最为精确,其模型为
$$(1 - 0.305\,92B - 0.335\,96B^4)(1 + 0.509\,78B^{12})(1 - B)(1 - B^{12})\ln X_t = \varepsilon_t$$

对于中长期预报来说,无约束最小二乘法的计算结果最精确,其模型为
$$(1 - 0.316\,93B - 0.332\,94B^4)(1 + 0.641\,91B^{12})(1 - B)(1 - B^{12})\ln X_t = \varepsilon_t$$

本文通过指定 P 或 Q 选项为括号内的一系列时间间隔,可以控制让某些时间间隔上具有参数.像这样只有一些在时间间隔上带有参数的模型被称为子集模型.大量数据模拟表明,子集模型比传统的方法精度高,更符合实际情况.本文如采用传统的定阶方法,则 AIC 为 -388.758 ,SBC 为 -377.516 ;用本文子集模型的定阶方法,则 AIC 为 -386.133 ,SBC 为 -379.387 ,可见本文的方法精确较高.

参考文献:

[1] 高惠璇. SAS 系统 SAS/ETS 软件使用手册 [K]. 北京: 中国统计出版社,1988 35-116.
[2] 王振龙. 时间序列分析 [M]. 北京: 中国统计出版社,2000 118-180.
[3] 陈平,达庆利. 运用 SAS 软件系统对我国农作物受灾面积的预测分析 [J]. 系统工程理论与实践,2001 (11) 360-663.
[4] 顾岚. 时间序列分析在经济中的应用 [M]. 北京: 中国统计出版社,1994 78-184.
[5] WILLIAM W W. Time series analysis :univariate and multivariate method [M]. Redwood :Addison-Wesley,1989 35-162.
[6] PETER J B, RICHARD A D. 时间序列的理论与方法 [M]. 田铮,译. 北京: 高等教育出版社,2001 39-153.

Application of multiply season model to deposit prediction

RU Zheng-liang, YIN Fan-cheng, ZHANG Yan
(College of Sciences, Hohai University, Nanjing 210098, China)

Abstract The time series method in SPSS and SAS software system was used for modeling residents' deposits in urban and rural areas in China. With the maximal likelihood method adopted for short-term prediction and unrestricted least square method for middle- and long-term prediction, high precision of prediction was achieved.

Key words time series; auto-correlation function; partial auto-correlation function, Box-Jenkins modeling; deposit prediction