逐步门限自回归模型及其建模方案

曹 杰 张万诚

(云南省红河州气象局)

摘要 针对门限自回归模型在模型识别方面的不足,提出了逐步门限自回归模型,并同时给出了该模型的一种建模方案。数值实例表明,逐步门限自回归模型在模拟和预报稳定上比一般门限自回归模型有一定程度的提高。

关键词 逐步,门限自回归,建模

门限自回归模型由英国曼彻斯特大学的 Tong H^[1]在 1987 年首次提出,由于该模型具备有效地描述非线性振动现象,解释各种稳定循环、反映出类似共振、跳跃的实变的功能,因而该模型在预测领域中得到了较为广泛的应用。但是,从目前所做的工作来看,门限自回归模型均是基于对各分段样本建立自回归模型而得,所建立的预报方程就可能混入一些对预报对象不起作用或影响极小的预报因子,而导致预报方程的稳定性减弱,最终造成预报失败。为此,本文把逐步算法引入到门限自回归的建模过程中^[2],提出并建立了逐步门限自回归模型。用该模型对二年独立样本进行试报,结果表明此模型具有一定的实用价值。

1 数学模型

与门限自回归模型相同,逐步门限自回归模型可写为

$$x_t = a_0^{(j)} + \sum_{i=1}^k a_i^{(j)} x_{t-i} + \epsilon_t^{(j)} \quad r_{j-1} \leq x_{t-d} < r_j \quad j = 1, 2, \dots, L \quad (1)$$

其中, x_t 是随机序列,且 $\{\epsilon_t^{(j)}\}$ 是 L 个相互独立的正态白噪声序列, $j=1, 2, \dots, L$; d 为时滞, x_{t-d} 为门限变元; $r_0 = -\infty, r_1 = +\infty, r_0 < r_1 < r_2 < \dots < r_L, r_1, r_2, \dots, r_{L-1}$ 为门限值。门限值将实轴分为 L 段,当 x_{t-d} 落入第 j 段时, x_t 采用上式中第 j 个自回归模型描述,阶数为 K_j 。

2 逐步门限自回归模型的建模步骤

逐步门限自回归模型中的待估参数有,各分段模型的系数、阶数、后延(时滞) d 和门限值 r 。为了清楚起见,以 $L=2$ 的情况为便说明建模步骤。

2.1 生成预报因子集

预先给定逐步门限自回归模型后延 d 的最大上限,考虑预报对象序列样本数的限制,后延 d 不宜取得太大。设后延 d 的最大上限为 R ; n 为样本长度,则生成的预报因子集为 A 。

$$A = \begin{bmatrix} x_k & x_{k+1} & \cdots & x_{n-1} \\ x_{k-1} & x_k & \cdots & x_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ x_1 & x_2 & \cdots & x_{n-k} \end{bmatrix}$$

简记 $A = \{x_{n-1}, x_{n-2}, \dots, x_{n-k}\}^T$ 。

2.2 最优分割法确定门限后延 d 和门限值 r

根据统计学,当预报系统中某个预报因子大于或小于某一临界值时,其所分成的两组子样本的组间方差有最大差异,并且在第一段样本内差异最小,两段间差异最大。据此可寻找使预报对象分组产生最大差异的预报因子值作为门限值,同时也确定了门限时滞。具体作法:

(1)在上述生成的预报因子集 A 中依次取因子 $x_{n-d}, d=1, 2, \dots, k$ 。 $x_{n-d} = (x_{n-i-k-1}, x_{n-i+1-k}, \dots, x_{n-i})$, $i=1, 2, \dots, k$;按其值从小到大顺序排列,预报对象 $x_n = (x_{k+1}, x_{k+2}, \dots, x_n)$ 随 x_{n-d} 序列的顺序相应重新排序,得新序列 x'_n 。

(2)对 x'_n 序列进行二分分割。按大于或小于等于 x_{n-d} 序列中的某个元素值分段。如: >174 或 ≤ 174 ,即可将 x'_n 序列分成两段,将两段记为: $x'_n(k+1, n), x'_n(m+1, n)$ 。于是得出 x'_n 序列的总方差为 $r^2 = \sum_{i=1}^n (x'_n(i) - \bar{x}'_n)^2$,两段的组内方差和为 $S^2 = \sum_{i=k+1}^m (x'_n(i) - \bar{x}'_n(k+1, m))^2 + \sum_{i=m+1}^n (x'_n(i) - \bar{x}'_n(m+1, n))^2$,组间方差为 $T^2 = r^2 - S^2$ 。

(3)用 $F = T^2 \times (n-2) / S^2$ 检验两组间是否存在显著性差异。对序列 x'_n 共进行 $P = n - k - 2$ 次二分分割,计算 P 个 F 值,记为 $F_1(x_{n-d}), F_2(x_{n-d}), \dots, F_P(x_{n-d})$ 。选择 $F_{\max P^*}(x_{n-d}) = \max\{F_1(x_{n-d}), F_2(x_{n-d}), \dots, F_P(x_{n-d})\}$, P^* 是对应的分割点,即为第 i 个因子的最优分割点。

(4)对所有因子 $x_{n-d}, d=1, 2, \dots, k$;分别求出 $F_{\max P^*}(x_{n-1}), F_{\max P^*}(x_{n-2}), \dots, F_{\max P^*}(x_{n-k})$,再选择 $F_{\max P^*}(x_{n-d}^*) = \max\{F_{\max P^*}(x_{n-1}), F_{\max P^*}(x_{n-2}), \dots, F_{\max P^*}(x_{n-k})\}$ 。这样就确定了最优分割因子 x_{n-d}^* 和最优分割点 $\max P^*$ 。即确定了门限后延(时滞) d 和门限值 r 。

(5)将 x_{n-d}^* 按大于或等于 x_{n-d}^* 分为两段,预报因子集 A 中其余预报因子和预报对象相应地随 x_{n-d}^* 分为两段。

2.3 逐步门限自回归模型

将已分成的两组样本分别作逐步回归,记 $A' \subset A$,则得逐步门限自回归模型为

$$x_n = \begin{cases} \beta_0^{(1)} + \beta^{(1)} \cdot A^{(1)} + \epsilon_i^{(1)} & x_{n-d}^* \leq r \\ \beta_0^{(2)} + \beta^{(2)} \cdot A^{(2)} + \epsilon_i^{(2)} & x_{n-d}^* > r \end{cases} \quad (2)$$

与门限自回归模型的预报方法相同,当给出预报因子资料后,首先根据后延 d 和门限值 r ,判定其属于哪一段,然后用该段模型代入预报因子实时资料即可作出预报。

根据以上建模方法,不难推广到分为 L 段的门限回归模型,因此逐步门限回归模型最终为

$$x_n = \begin{cases} \delta_1(A^{(1)}) & x_j^* \leq r_1 \\ \delta_2(A^{(2)}) & r_1 < x_j^* < r_2 \\ \dots & \dots \\ \delta_L(A^{(L)}) & x_j^* > r_{L-1} \end{cases}$$

当给出预报资料后,根据后延和门限值 r ,判断其属于哪一段,再把因子时实资料代入,即可作出预报。

3 实例计算

以蒙自 1951—1990 年 3 月平均温度序列为模拟预报对象。每年 3 月是蒙自地区发生“倒春寒”天气频率最高的月份,故 3 月平均温度年际变化大,最高年与最低年差值达 5°C ,因而造成了一定的预报难度。给定后延 d 的最大上限为 6,按照前述建模步骤,建立逐步门限自回归模型

$$x_n = \begin{cases} -4.8315x_{n-1} + 0.4928x_{n-2} - 0.6169x_{n-3} - 0.4938x_{n-4} + 366.019 & x_{n-4} > 176 \\ 0.1989x_{n-2} - 0.2789x_{n-3} - 0.5463x_{n-4} + 0.4478x_{n-5} + 18.0484 & x_{n-4} \leq 176 \end{cases} \quad (3)$$

(3)式的总复相关系数 $R=0.72$,总拟合误差均方差 $S=9.28$,通过 0.01 显著性检验。而用一般门限自回归模型建立的 3 月气温预报模型的总复相关系数 $R=0.58$,均方差 $S=10.8$,只通过 0.05 显著性检验。

表 1 为两种方法的模拟结果,其中 Y 为实测值, Y' 为模拟值;趋势中,“+”为高 $T \geq 183^{\circ}\text{C}$,”-”为低 $T \leq 173^{\circ}\text{C}$,”0”为正常 $173^{\circ}\text{C} < T < 183^{\circ}\text{C}$;趋势评定中“√”为正确,“×”为错误。

表 1 两种方法的预报模拟结果

年	Y	Y'		趋势		趋势评定		年	Y	Y'		趋势		趋势评定	
		逐步	一般	逐步	一般	逐步	一般			逐步	一般	逐步	一般		
1957	174	182	176	0	0	√	√	1974	157	165	176	-	0	√	×
1958	199	191	179	+	0	√	×	1975	195	191	182	+	0	√	×
1959	159	174	173	0	0	×	×	1976	173	176	174	0	0	√	√
1960	194	182	194	0	+	×	√	1977	171	172	179	-	0	√	×
1961	179	179	174	0	0	√	√	1978	188	191	179	+	0	√	×
1962	163	164	170	-	-	√	√	1979	194	182	175	0	0	×	×
1963	170	186	181	+	0	×	×	1980	202	185	194	+	+	√	√
1964	178	183	179	0	0	√	√	1981	189	186	188	+	+	√	√
1965	171	172	178	-	0	√	×	1982	197	176	198	0	+	×	√
1966	192	191	179	+	0	√	×	1983	156	166	162	-	-	√	√
1967	171	172	175	-	0	√	×	1984	179	179	180	0	0	√	√
1968	175	186	158	+	-	×	×	1985	152	152	177	-	0	√	×
1969	186	172	178	-	0	×	×	1986	176	188	183	+	0	×	√
1970	184	177	176	0	0	×	×	1987	188	191	178	+	0	√	×
1971	176	188	176	+	0	×	√	1988	163	170	175	-	0	√	×
1972	174	182	178	0	0	√	√	1989	159	174	172	0	0	×	×
1973	187	181	179	0	0	×	×	1990	163	164	182	-	0	√	×

从表中可看出,逐步门限自回归模型的模拟趋势准确率为 22/34;一般门限自回归的模拟趋势准确率则为 14/34。

将逐步门限自回归模型投入业务应用,于 1991 年 2 月作 1991 年 3 月平均温度预报,得 $x_{11}=19.6^{\circ}\text{C}$,趋势为特高。1991 年 3 月温度实况值为 20.2°C ,用一般门限回归模型得 $x_{11}=18.1^{\circ}\text{C}$ 。1992 年 2 月作 3 月平均温度预报,得 $x_{12}=18.6^{\circ}\text{C}$,趋势为正常至偏高;一般门限自回归模型得 $x_{12}=19.2^{\circ}\text{C}$,趋势为特高;实况是 18.3°C 。两年试用效果较好。可见,在长期预报中逐步门限自回归模型更优。文中温度值均扩大至 10 倍。

4 讨 论

本文建立的蒙自 3 月平均温度逐步门限自回归模型的模拟和预报效果均比一般门限自回归模型有一定程度的提高。但效果不很理想,其原因可能与样本长度短有关;另外,除温度年际变化的内在规律外,外界因子,特别是大气环流形势、海温等对温度的变化也起着重要作用。因此,改进模拟和预报稳定性的途径之一就是引进外部强迫因子,建立混合门限回归模型。

参 考 文 献

- 1 Tong H, Lin KS. *J of Royal statistical Society*, 1980, 42(B), 245-292
- 2 么枕生, 丁裕国. 气候统计(修订本). 北京:气象出版社, 1990
- 3 项静恬等. 动态和静态数据处理——时间序列和数理统计分析. 北京:气象出版社, 1991

DESIGNING OF THE STEPWISE THRESHOLD AUTOREGRESSIVE MODEL

Cao Jie Zhang Wancheng

(Honghe Prefecture Meteorological Bureau of Yunnan Province)

Abstract The designing of a stepwise threshold autoregressive model is presented in this paper to improve the ordinary threshold autoregressive model that is imperfect in recognition. Numerical experiments show that the designed model is better to some extent.

Key words stepwise, threshold autoregression, model-designing