第29卷 第4期 中山大学学报(自然科学版) Vol.29 No.4
1990年 ACTA SCIENTIARUM NATURALIUM UNIVERSITATIS SUNYATSENI 1990

·研究简报·

门限自回归迟滞参数和门限参数的一种算法

马肖娥 叶向阳
(中山大学计算机科学系) (广东省统计局)

摘 要

根据非线性多元函数存在最大波动变元x_{n-d}的几何背景，由变量x_{n-d}的波峰和波谷点给出了非线性时间序列的迟滞参数和门限参数的估计。

关键词 门限自回归，门限参数，迟滞参数

门限自回归模型(Threshold Autoregressive Models)，简记为TAR模型，是1977年由汤家豪(Tong)首先提出的一种非线性模型[1]。近年来，这一模型的研究及其应用正在不断完善和深入发展。

设有时间序列$\{x_t\}$，本文讨论由此而建立的自激门限自回归模型

$$X_t=a_0^{(j)}+a_1^{(j)}X_{t-1}+\cdots+a_{m_j}^{(j)}X_{t-m_j}+\epsilon_t^{(j)} \tag{I}$$

$$\text{当}\quad X_{t-d}\in(r_{j-1},r_j]\text{时}$$

其中，$-\infty=r_0<r_1<r_2<\cdots<r_k<r_{k+1}=+\infty$

$\{\varepsilon_t^{(j)}\}$为服从$N(0,\sigma_j^2)$分布的独立随机序列，j的取值为1，2，…，$k+1$；相应的m_j称为TAR的阶数，r_j $j=1,2,\cdots,k$称为门限值，d称为迟滞参数。

1 建模的方法

多变元函数中各变元对函数波动之叠加是相当复杂的。但可以取其波动最大的变量x^*_{n-d}及相应变量波动之峰、谷点r^*_1，r^*_2，…，r^*_k作迟滞参数和门限值，并由此建立起门限自回归模型。基于上述思想，我们设想若非线性时间序列模型(I)(为方便起见记为$x_n=f(x_{n-1},\cdots,x_{n-l}),l=\max\limits_j\{m_j\}$)的第$d$个变元$x_{n-d}$能使$f(\cdot)$在所论范围内各波动幅度与相应波长比的累积最大，那么就可把d作为最佳时滞参数，$f(\cdot)$关于x_{n-d}的波峰点和波谷点$r_1,r_2,\cdots,r_k$就可作为最佳门限值。

当给定时间序列样本值$\{x_t\}_{t=1}^N$后，就可以用下面的方法确定$\hat{d}$与$\hat{r}_1$，$\hat{r}_2$，…，$\hat{r}_k$，用AIC准则确定m_j，并用最小二乘法确定系数$a_i^{(j)}$。

模型中变元$\{x_n\}$，用其均值$\{Ex_n\}$以消除随机干扰。这里的Ex_n，可通过非参数估

本文1988年9月20日收到

计方法求得$E(x_n|x_{n-1},\cdots,x_{n-L})$来作近似替代的。其计算公式如下：

$$E(x_n|x_{n-1},\cdots,x_{n-L})=\sum_{i=L+1}^{N}x_iW_{Ni}(x_n,x_{n-1},\cdots,x_{n-L})$$

其中核(Kernel)权函数为：

$$W_{Ni}(x_n,x_{n-1},\cdots,x_{n-L})=\left[\prod_{j=1}^{L}K(|x_{i-j}-x_{n-j}|/\lambda)\right]\Big/\left[\sum_{i=L+1}^{N}\prod_{j=1}^{L}K(|x_{i-j}-x_{n-j}|/\lambda)\right]$$

$$K(Z)=\begin{cases}1-|Z| & |Z|<1\\ 0 & |Z|\geqslant 1\end{cases}$$

这里，选取$\lambda=S$，其中S是样本的标准差，但这不是最佳的参数。非参数及核函数的估计参阅文[2]。

设有样本$\{x_n\}_{n=1}^{N}$，相应均值$\{Ex_n\}_{n=1}^{N}$，且有$Ex_n\leqslant Ex_{n+1}$，$n=1,2,\cdots,N-1$；（若条件不满足时，可对Ex_n加以重排，考虑Ex_n不够准确，在实际处理时，往往先作$Y_n=\frac{1}{2}(Ex_n+Ex_{n+1})$的平滑化，把不代表曲面变化趋势的波动磨平，以$\{Y_n\}$序列直接建模）。若令

$E^*x_n\triangleq f(Ex_{n-1},Ex_{n-2},\cdots\cdots,Ex_{n-l})$。并且定义函数$op(d)$以描述分量$Ex_{n-d}$对函数$f(Ex_{n-1},\cdots,Ex_{n-l})$波动性的影响：

$$\begin{aligned}op(d)\triangleq&\sum_{n-d\in I_d}\left(\left|\frac{f(Ex_n,\cdots,Ex_{n-d+1},\cdots,Ex_{n-l+1})}{Ex_{n-d+1}-Ex_{n-d}}\right.\right.\\&-\left.\frac{f(Ex_{n-1},\cdots,Ex_{n-d},\cdots,Ex_{n-l})}{Ex_{n-d+1}-Ex_{n-d}}\right|+\left|\frac{f(Ex_{n-1},\cdots,Ex_{n-d},\cdots,Ex_{n-l})}{Ex_{n-d}-Ex_{n-d-1}}\right.\\&-\left.\left.\frac{f(Ex_{n-2},\cdots,Ex_{n-d},\cdots,Ex_{n-l-1})}{Ex_{n-d}-Ex_{n-d-1}}\right|\right)\\=&\sum_{n-d\in I_d}\left(\left|\frac{E^*x_{n+1}-E^*x_n}{Ex_{n-d+1}-Ex_{n-d}}\right|+\left|\frac{E^*x_n-E^*x_{n-1}}{Ex_{n-d}-Ex_{n-d-1}}\right|\right)\end{aligned}$$

$$令\ \overset{i=n-d}{=\!=\!=}\sum_{i\in I_d}\left(\left|\frac{E^*x_{i+1+d}-E^*x_{i+d}}{Ex_{i+1}-Ex_i}\right|+\left|\frac{E^*x_{i+d}-E^*x_{i-1+d}}{Ex_i-Ex_{i+1}}\right|\right)$$

其中 $I_d=\{i|(f(\cdots\underset{\substack{\uparrow\\ 第d个变量}}{Ex_{i+1}}\cdots)-f(\cdots\underset{\substack{\uparrow\\ 第d个变量}}{Ex_i}\cdots))(f(\cdots\underset{\substack{\uparrow\\ 第d个变量}}{Ex_i}\cdots)-f(\cdots\underset{\substack{\uparrow\\ 第d个变量}}{Ex_{i-1}}\cdots))<0\}$

最佳迟滞参数$\hat{d}$使$op(d)+op(d)/d$达到最大。因为事先不知道非线性模型的阶l，而在我们的计算过程利用了$Ex_n\approx E(x_n|x_{n-1},\cdots,x_{n-L})$，并且还让$E^*x_n\triangleq f(Ex_{n-l},\cdots,Ex_{n-l})$，当$d>l$时，便有$E^*x_n=f(Ex_{n-1},\cdots,Ex_{n-d},\cdots,Ex_{n-l})$，易于产生混乱，$op(d)$所反映的信息有可能失真，所以增加$op(d)/d$一项以调节$d$与$l$的关系，保证有$d<l$。

$\hat{d}$选出后，便可选取使

$$\left|\frac{Ex_{i+1+d}-Ex_{i+d}}{Ex_{i+1}-Ex_i}\right|+\left|\frac{Ex_{i+d}-Ex_{i-1+d}}{Ex_i-Ex_{i+1}}\right| \quad i\in I_d$$

达到极大值（即波动最剧烈）的若干个$Ex_i(i=i_1,\ i_2,\ \cdots,\ i_k)$，作最佳门限值$\bar{r}_i$。

2　两种建模方法的比较

在Apple Ⅱ微机上，用BASIC语言分别对Tong和我们的建模算法编制程序计算，选取了加拿大山猫（1821～1920年）和太阳黑子数（1700～1920年）的两批数据进行拟合预测，算法结果表明，本算法有效地缩减了运算时间，且所得的预报结果和拟合残差平方及均值与Tong在〔1〕中的结果比较接近。

下面仅列举太阳黑子数建模的结果加以比较。若令

$$X'=(1,x_{N-1},x_{N-2},x_{N-3},x_{N-4},x_{N-5},x_{N-6},x_{N-7},x_{N-8},x_{N-9},x_{N-10},x_{N-11},x_{N-12},1)$$

则Tong与我们的模型分别为：

$$X_N=\begin{cases} X'\begin{pmatrix}10.544\\1.692\\-1.159\\0.237\\0.150\\0\\\vdots\\0\\\varepsilon_N^{(1)}\end{pmatrix} & \text{if } X_{N-3}<36.6\\[2ex] X'\begin{pmatrix}7.804\\0.743\\-0.041\\-0.202\\0.173\\-0.227\\0.019\\0.161\\-0.256\\0.319\\-0.389\\0.431\\-0.040\\\varepsilon_N^{(2)}\end{pmatrix} & \text{if } X_{N-3}>36.6\end{cases}
\qquad
X_N=\begin{cases} X'\begin{pmatrix}7.354\\1.515\\-0.223\\-1.651\\0.341\\0.503\\-0.364\\0.032\\0.083\\-0.187\\0.461\\-0.249\\0\\0\end{pmatrix} & \text{if } X_{N-3}\le 18.02\\[2ex] X'\begin{pmatrix}-2.780\\1.820\\-2.120\\1.711\\-1.057\\0.640\\0\\\vdots\\0\end{pmatrix} & \text{if } 18.02<X_{N-3}\le 32\\[2ex] X'\begin{pmatrix}12.401\\0.717\\-0.062\\-0.259\\0.182\\-0.196\\-0.032\\0.177\\-0.242\\0.252\\-0.236\\0.354\\0\\0\end{pmatrix} & \text{if } X_{N-3}>32\end{cases}$$

以我们所编的BASIC程序，在Apple机上建模，Tong的方法需时 2 小时20 分，我们的方法不到 1 小时。

两种方法进行一步向前预测的结果比较为:

时间(年)	1921	1922	1923 …	1928 …	1930 …	1950	1951	1952 …	1954 …
实测值	26.1	14.1	5.8	77.8	35.7	83.9	69.4	31.5	4.4
Tong发表值	29.182	10.236	5.294	75.358	55.433	97.849	57.312	33.665	2.767
本文预测值	28.201	12.312	5.661	77.288	56.422	94.196	55.481	31.277	3.202

两种算法的残差平方和均值都颇为接近，Tong的为153.71，而我们的为156.40。

参考文献

〔1〕 Tong H et al., *J.R.Statist. Soc.*, B.42(1980), 245～292

〔2〕 Stone C J, *Ann. Statist.*, 1977, 5, 595～645

An Algorithm for the Delay Parameter and Threshold Parameter of Threshold Autoregressive

*Ma Xiaoe** *Ye Xiangyang*

Abstract

According to the fact that in the nonlinear multivariabl case, there exists a maximum wave motion variable x_{n-d}, we are able to estimate the delay parameter and the threshold parameter of the nonlinear time series through points of the crest and the trough of the variable x_{n-d}.

Keywords threshold autoregressive, threshold parameter, delay parameter

* Department of Computer Science