

文章编号: 1000-8152(2008)03-0511-06

非线性时间序列建模的均值异方差混合转移分布模型

王红军^{1,2}, 田 铮^{2,3}, 韩四儿²

(1. 西安电子科技大学 理学院 应用数学系, 陕西 西安 710071; 2. 西北工业大学 理学院 应用数学系, 陕西 西安 710072;
3. 中国科学院 自动化研究所 模式识别国家重点实验室, 北京 100080)

摘要: 进一步研究了由Berchtold提出的均值异方差混合转移分布(expectation heteroscedastic mixture transition distribution model, EHMTD)模型. 讨论并得到了EHMTD模型的平稳性条件和分布函数的尾部特征. 运用ECM(expectation conditional maximization)算法估计模型的参数. 条件分布的多样性使得该类模型能够对非对称、多峰、厚尾等非Gauss特征进行描述. 模拟及实例分析的结果表明EHMTD模型是一类易于建模, 并且有着广泛应用前景的非线性时间序列模型.

关键词: 平稳性; BIC准则; ECM算法; 非对称; 多峰; 厚尾; 条件异方差
中图分类号: TP391, N93 **文献标识码:** A

Expectation heteroscedastic mixture transition distribution model for modeling nonlinear time series

WANG Hong-jun^{1,2}, TIAN Zheng^{2,3}, HAN Si-er²

(1. Department of Applied Mathematics, School of Science, Xidian University, Xi'an Shaanxi 710071, China;
2. Department of Applied Mathematics, School of Science, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China;
3. National Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: The expectation heteroscedastic mixture transition distribution (EHMTD) model first introduced by Berchtold is further studied in this paper. First, the stationary conditions and tail behaviors of the model are derived. The estimation of parameters is easily performed via expectation conditional maximization (ECM) algorithm. The variety of conditional distributions of the EHMTD model makes the model capable of modeling time series with asymmetric multimodal or heavier tail distribution. The model is applied to simulate real data sets with satisfactory results. The EHMTD model is easy to model and potentially useful in modeling nonlinear time series.

Key words: stationarity; BIC; ECM algorithm; asymmetric; multimodal; heavier tail; conditional heteroscedasticity

1 引言(Introduction)

时间序列分析作为一种处理动态数据的有效方法被广泛应用于工程技术和金融领域, 但随着科学技术和工程应用的发展, 许多物理系统所展示出来的复杂行为不能被传统的时间序列模型准确描述. 例如工程和金融中有的变量具有非对称、厚尾及其它一些非Gauss特征. 若采用线性模型对这些时间序列描述, 往往达不到准确预报和有效控制的目的. 由于混合模型作为一种简单的、不必拘泥于某一种单一分布假设的方法, 具有可以有效逼近任何复杂分布形式的优势. 自从Pearson(1894)根据实际问题的需要提出混合模型, 该类模型一直受到不同领域研

究者的关注和应用^[1]. 计算工具的的不断更新和发展加速了混合模型的研究与应用范围. 将混合模型应用于时间序列分析是近年才展开的.

Le、Martin和Raftery(1996)^[2]首先提出Gauss混合转移分布模型(Gauss mixture transition distribution, GMTD)模型. 通过放松GMTD模型中的一些约束条件, Wong和Li(2000)^[3]将其推广得到混合自回归(mixture autoregressive, MAR)模型. Wong和Li(2001)进一步将其推广得到混合自回归条件异方差(mixture autoregressive conditional heteroscedastic, MAR-ARCH)模型^[4]和具有外生变量的对数混合自回归(logistic mixture autoregressive with ex-

收稿日期: 2006-07-04; 收修改稿日期: 2007-01-29.

基金项目: 国家自然科学基金资助项目(60375003); 教育部重点科研基金资助项目(03153059); 西北工业大学科技创新基金资助项目(2007KJ01033).

ogenous variables, LMARX) 模型^[5]. 考虑到残量信息, Wang和Tian (2004)^[6]将MAR模型推广得到混合ARMA (mixture autoregressive moving-average model, MARMA) 模型. Shao (2006)^[7]将该类模型应用于周期性时间序列分析, 提出混合周期自回归(mixture periodic autoregressive, MPAR)模型. 通过定义每个时刻各分量的标准差是以往某几个观测值的函数, Berchtold (2003)^[8]把异方差引进MAR模型, 从而将MAR模型推广得到异方差混合转移分布(heteroscedastic mixture transition distribution, HMTD) 模型. Wang和Tian (2005)在HMTD^[9]模型基础上, 着重研究了异方差混合双自回归(heteroscedastic mixture double-autoregressive, HMDAR) 模型. 以上各种混合时间序列模型在各分量的参量服从Gauss分布的假设下, 却可以在不同情况下有效描述非Gauss特征.

Berchtold(2003)^[8]根据各分量的标准差与观测序列之间不同的函数关系, 给出了HMTD模型不同的定义形式. 本文主要研究了下面的模型(1). 为了同其它的定义形式相区别, 文中称其为均值异方差混合转移分布(expectation heteroscedastic mixture transition distribution, HMTD) 模型:

$$f(y_t | y_{t-p}^{t-1}) = \sum_{g=1}^K \frac{\alpha_g}{\sigma_{g,t}} \phi\left(\frac{y_t - \mu_{g,t}}{\sigma_{g,t}}\right), \quad (1)$$

$$\mu_{g,t} = \varphi_{g,0} + \sum_{i=1}^p g\varphi_{g,i}y_{t-i}, \quad p_g \geq 1;$$

$$\sigma_{g,t} = \sqrt{\theta_{g,0} + \sum_{j=1}^q g\theta_{g,j}(y_t - \mu_{g,t})^2}, \quad p_g \geq 1.$$

其中: $f(y_t | y_{t-p}^{t-1})$ 表示 y_t 的条件密度函数, $y_{t-p}^{t-1} = \sigma(y_{t-p}, \dots, y_t)$ 是 σ -代数, $\phi(\cdot)$ 是标准正态密度函数. α_g 是第 g 个分量的权重系数, $\sum_{g=1}^K \alpha_g = 1, \alpha_g > 0$. $\mu_{g,t}$ 和 $\sigma_{g,t}$ 是第 g 个分量在 t 时刻的条件期望和标准差. 模型(1)简记为EHMTD($K; p_1, \dots, p_K; q_1, \dots, q_K$).

EHMTD模型是具有均值异方差的各AR分量的加权混合. 文中给出了EHMTD模型在某些约束下的平稳性条件. 对于两个混合元的情况, 得到了EHMTD模型具有厚尾分布特征的条件. 关于EHMTD模型的建模, 本文运用BIC(Bayes information criterion) 准则来选择EHMTD模型, 采用ECM(expectation conditional maximization) 算法估计模型的参数. 通过仿真分析, 说明了ECM算法估计EHMTD模型参数的精度. 将模型应用于实例分析, 演示了EHMTD模型的建模方法及其具有描述多峰、厚尾(或薄尾)等非Gauss特征的优势.

2 EHMTD模型的平稳性和尾部特征(The stationarity and tail behaviors of the model)

EHMTD模型除了每个分量在不同时刻具有异方差, 同时也具有描述过程在不同时刻变化的条件方差的特征. y_t 的条件方差为

$$\text{Var}(y_t | y_{t-p}^{t-1}) = \sum_{g=1}^K \alpha_g (\mu_{g,t}^2 + \sigma_{g,t}^2) - \left(\sum_{g=1}^K \alpha_g \mu_{g,t}\right)^2.$$

定理2.1给出EHMTD($K; p_1, \dots, p_K; q_1, \dots, q_K$) 模型一阶平稳的充分必要条件. 定理2.2给出了EHMTD($K; 2, \dots, 2; 2, \dots, 2$) 模型二阶平稳的充分必要条件. 定理的证明运用了Le et al (1996)^[2]以及Wong和Li (2000)^[3]中所用的方法, 用到Benš(1967)^[10]中的结论. 方法类似, 这里略去证明过程.

定理 1 过程 $\{y_t, t \in T\}$ 一阶平稳的充分必要条件是下面差分方程的根 z_1, \dots, z_p :

$$1 - \sum_{i=1}^p \left(\sum_{g=1}^K \alpha_g \varphi_{g,i}\right) z^{-i} = 0,$$

都在单位圆内, 其中 $\varphi_{g,i} = 0 (i > p_K)$.

定理 2 假设 $\{y_t, t \in T\}$ 是一个由EHMTD($K; 2, \dots, 2; 2, \dots, 2$)模型所产生的一阶平稳过程, 则该过程二阶平稳的充分必要条件是

$$w_2 + w_1 < 1, |w_2| < 1.$$

其中:

$$w_1 = \sum_{g=1}^K \alpha_g [\theta_{g,1}(1 - \varphi_{g,1})^2 + (1 + \theta_{g,2})\varphi_{g,1}^2],$$

$$w_2 = \sum_{g=1}^K \alpha_g [\theta_{g,2}(1 - \varphi_{g,2})^2 + (1 + \theta_{g,1})\varphi_{g,2}^2] -$$

$$2 \sum_{g=1}^K \alpha_g [\theta_{g,1}(1 - \varphi_{g,1})\varphi_{g,2} - \varphi_{g,1}\varphi_{g,2} +$$

$$\theta_{g,2}(1 - \varphi_{g,2})\varphi_{g,1}] \left[\frac{\sum_{g=1}^K \alpha_g \varphi_{g,1}}{1 - \sum_{g=1}^K \alpha_g \varphi_{g,2}} \right].$$

随着对金融时间序列研究的深入, 人们逐渐认识到金融时间序列往往具有非对称、厚尾等非Gauss特征. 定理2.3给出了关于EHMTD($2; p_1, \dots, p_K; q_1, \dots, q_K$) 模型具有厚尾分布的一个结论.

定理 3 若EHMTD模型是由两个混合元构成的, 即

$$f(y_t | y_{t-p}^{t-1}) = \frac{\alpha_1}{\sigma_{1,t}} \phi\left(\frac{y_t - \mu_{1,t}}{\sigma_{1,t}}\right) + \frac{\alpha_2}{\sigma_{2,t}} \phi\left(\frac{y_t - \mu_{2,t}}{\sigma_{2,t}}\right),$$

则当 $\frac{3 + \sqrt{3}}{6} < \alpha_1 < 1$, 且 $0 < \sigma_{1,t} < \sigma_{2,t}$ 时, y_t 对

应的分布具有重尾特征.

3 EHMTD模型的参数估计及模型的选择(The parameter estimation of the EHMTD model and the model selection)

EHMTD模型对应的某些参数似然方程结构复杂, 运用EM算法得不到好的估计结果. 因此文中采用了GEM算法中的ECM算法来估计模型的参数. 假定 $Y = (y_1, \dots, y_n)$ 是EHMTD模型对应的一组观测值. $Z = (z_1, \dots, z_n)$ 是不可观测的添加随机变量序列, $z_t(t = 1, \dots, n)$ 是 K 维随机变量, $\sum_{g=1}^K z_{g,t} = 1$. 若 y_t 来自(1)中第 g 个分量, 则 $z_{g,t} = 1$; 否则:

$$\begin{aligned} z_{g,t} &= 0, \\ \alpha &= (\alpha_1, \dots, \alpha_{K-1})^T, \\ \varphi_g &= (\varphi_{g,0}, \varphi_{g,1}, \dots, \varphi_{g,p_g})^T, \\ \theta_g &= (\theta_{g,0}, \dots, \theta_{g,q_g})^T, \\ \Psi &= (\varphi_1^T, \dots, \varphi_K^T; \theta_1^T, \dots, \theta_K^T)^T, \\ \Theta &= (\alpha^T, \varphi_1^T, \dots, \varphi_K^T; \theta_1^T, \dots, \theta_K^T)^T, \\ g &= 1, \dots, K. \end{aligned}$$

令 $r = \max\{p_1, \dots, p_K; q_1, \dots, q_K\}$, 引入添加数据 Z 后, 完整对数似然函数为

$$\begin{aligned} LL_c &= \sum_{g=1}^K \log(\alpha_g) \sum_{t=r+1}^n z_{g,t} - \sum_{t=r+1}^n \sum_{g=1}^K z_{g,t} \log(\sigma_{g,t}) - \\ &\sum_{t=r+1}^n \sum_{g=1}^K z_{g,t} \frac{\epsilon_{g,t}^2}{2\sigma_{g,t}^2} - \frac{n-r}{2} \log(2\pi). \end{aligned} \quad (2)$$

EM算法由下面E步和M步构成:

E步 模型的参数假定已知, 不可观测变量的条件期望为

$$\tau_{g,t} = \frac{\frac{\alpha_g}{\sigma_{g,t}} \phi\left(\frac{\epsilon_{g,t}}{\sigma_{g,t}}\right)}{\sum_{g=1}^K \frac{\alpha_g}{\sigma_{g,t}} \phi\left(\frac{\epsilon_{g,t}}{\sigma_{g,t}}\right)}, \quad g = 1, \dots, K.$$

M步 把 $\tau_{g,t}$ 代入式(2), 通过极大化(2)得到参数的估计值

$$\hat{\alpha}_g = \frac{\sum_{t=r+1}^n \tau_{g,t}}{n-r}.$$

参数 $\varphi = (\varphi_1^T, \dots, \varphi_K^T)^T$ 满足下列方程

$$\sum_{t=r+1}^n \frac{\tau_{g,t} y_t w_i}{\sigma_{g,t}^2} = \sum_{s=0}^{p_g} \hat{\varphi}_{g,s} \sum_{t=r+1}^n \frac{\tau_{g,t} w_s w_i}{\sigma_{g,t}^2},$$

其中 $i = 0, \dots, p_g$, 及

$$w_i = \begin{cases} 1, & i = 0; \\ y_{t-i}, & i \neq 0. \end{cases}$$

参数 $\theta = (\theta_1^T, \dots, \theta_K^T)^T$ 满足下列方程

$$\sum_{t=r+1}^n \left[\frac{\tau_{g,t} u_j}{\sum_{s=0}^{q_g} \hat{\theta}_{g,s} u_s} \right] = \sum_{t=r+1}^n \left[\frac{\tau_{g,t} (y_t - \mu_{g,t})^2 u_j}{\left(\sum_{s=0}^{q_g} \hat{\theta}_{g,s} u_s \right)^2} \right],$$

其中 $j = 0, \dots, q_g$, 及

$$u_j = \begin{cases} 1, & j = 0; \\ y_{t-j}^2, & j \neq 0. \end{cases}$$

ECM算法保留了EM算法的简单性和稳定性, 它将原先EM算法中的M步分解为 $m = 2K + \sum_{g=1}^K (p_g + q_g)$ 次条件极大化: 在第 $i + 1$ 次迭代中, 记 $\Psi^{(i)} = (\varphi_1^{(i)}, \dots, \varphi_m^{(i)})$, 在得到 $Q(\Psi|\Psi^{(i)}, Y)$ 后, 首先在 $\varphi_2^{(i)}, \dots, \varphi_m^{(i)}$ 保持不变的条件下求使 $Q(\Psi|\Psi^{(i)}, Y)$ 达到最大的 $\varphi_1^{(i+1)}$, 然后在 $\varphi_1 = \varphi_1^{(i+1)}, \varphi_j = \varphi_j^{(i)}, j = 3, \dots, m$ 的条件下求使 $Q(\Psi|\Psi^{(i)}, Y)$ 达到最大的 $\varphi_2^{(i+1)}$, 如此继续, 经过 m 次条件极大化, 就得到 $\Psi^{(i+1)} = (\varphi_1^{(i+1)}, \dots, \varphi_m^{(i+1)})$, 完成一次迭代.

将E步和M步反复迭代直到收敛, 便可得到参数 Θ 的估计值.

关于EHMTD模型的选择, 一方面是确定模型的混合元个数 K , 另一方面是确定每个混合元的均值函数的回归阶数 p_g 和方差函数的阶数 q_g . 文中运用BIC 准则选择模型. HMDAR模型(1)中混合元数目 K 及模型阶数 $p_1, \dots, p_K; q_1, \dots, q_K$ 可由使如下BIC 值达到最小而得到

$$BIC = -2L + [2K - 1 + \sum_{g=1}^K (p_g + q_g)] \log(n - r).$$

其中

$$\begin{aligned} L &= \sum_{t=r+1}^n \log \left[\sum_{g=1}^K \frac{\alpha_g}{\sigma_{g,t}} \exp\left\{-\frac{(y_t - \mu_{g,t})^2}{2\sigma_{g,t}^2}\right\} \right] - \\ &\frac{n-r}{2} \log(2\pi). \end{aligned}$$

4 仿真模拟(Simulation)

GMTD, MAR, MARMA模型均有一个有趣的性质, 即非平稳分量与平稳分量经混合可能产生一个平稳过程. EHMTD模型仍然具有该特征. 模型(3)是由两个混合元构成的一个EHMTD(2; 2, 2; 1, 0) 模型. 它的第2个分量是非平稳的, 但与第1个分量混合后模拟产生的序列(如图1)经检验却是平稳的.

$$\begin{aligned} f(y_t|y_{t-2}^{t-1}) &= \frac{0.7}{\sigma_{1,t}} \phi\left(\frac{y_t - \mu_{1,t}}{\sigma_{1,t}}\right) + \frac{0.3}{\sigma_{2,t}} \phi\left(\frac{y_t - \mu_{2,t}}{\sigma_{2,t}}\right), \\ \mu_{1,t} &= 0.3y_{t-1} + 0.2y_{t-2}, \\ \mu_{2,t} &= 1.1y_{t-1} - 0.4y_{t-2}; \end{aligned}$$

$$\begin{cases} \sigma_{1,t} = \sqrt{0.5 + 0.3(y_{t-1} - \mu_{1,t})^2}, \\ \sigma_{2,t} = \sqrt{5}. \end{cases} \quad (3)$$

先由模型(3)模拟产生数据, 然后用ECM算法估计模型的参数. 实验共分两组进行, 每组产生500个样本轨道, 其中一组每个轨道上有250个样本, 另一组每个轨道上有500个样本. 表1给出模型(3)模拟产生数据的估计结果. 由表中估计结果可明显看出, ECM算法给出了EHMTD模型参数较好的估计. 同时, 随着样本量的增大, 参数估计的均值越来越接近真值.

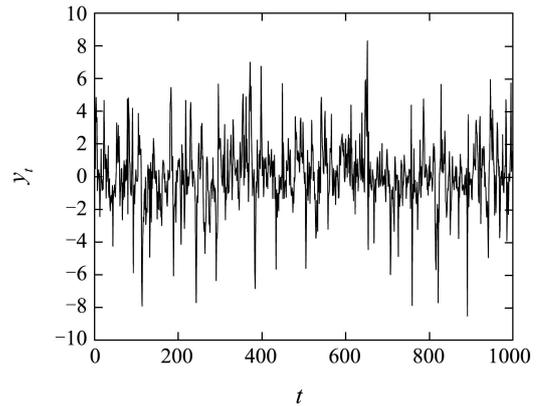


图1 模型(3)模拟产生的随机数序列
Fig. 1 Simulated time series of model (3)

表1 模型(3)模拟产生数据的估计结果

Table 1 Results of the simulation study with model (3)

g	α_g	$\varphi_{g,1}$	$\varphi_{g,2}$	$\theta_{g,0}$	$\theta_{g,1}$					
						250个样本				
真值	0.7000	0.3000	0.2000	0.5000	0.3000	0.7000	0.3000	0.2000	0.5000	0.3000
1 均值	0.6785	0.2937	0.2013	0.4961	0.2895	0.6885	0.2985	0.2007	0.4998	0.2981
标准差	0.0883	0.0986	0.0621	0.1405	0.2044	0.0546	0.0644	0.0414	0.0877	0.1353
						500个样本				
真值	0.3000	1.1000	-0.4000	5.0000		0.3000	1.1000	-0.4000	5.0000	
2 均值	0.3215	1.0358	-0.3686	4.6326		0.3115	1.0694	-0.3913	4.8851	
标准差	0.0883	0.2552	0.2421	1.0476		0.0546	0.1706	0.1523	0.7966	

5 实例分析(Real example)

加拿大山猫捕获量数据(1821~1934, 共计114个观测值)是被众多研究者用不同模型分析过的时间序列. 该序列被广泛认为具有循环特征(周期约为9或10年)和多峰分布的非线性特征. 图2给出原序列经常用对数变换后对应的数据点连线图.

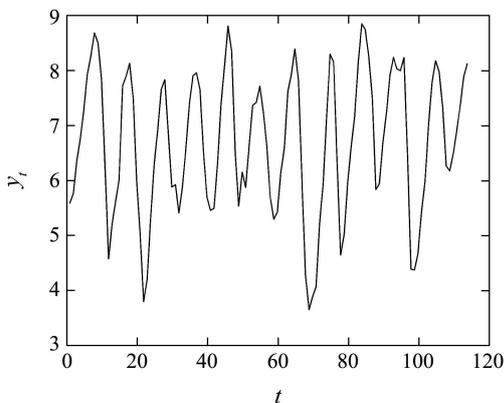


图2 模型(3)加拿大山猫数据经对数变换后的时间点图
Fig. 2 Time plot of the common logarithm of the Canadian lynx data

基于BIC准则, 并综合考虑预报效果, 选用下面的EHMTD(2; 2, 2; 1, 1)模型对上述数据的常用对

数序列进行建模:

$$\begin{aligned} f(y_t|y_{t-2}^{t-1}) &= \\ & \frac{0.316}{\sigma_{1,t}} \phi\left(\frac{y_t - \mu_{1,t}}{\sigma_{1,t}}\right) + \frac{0.684}{\sigma_{2,t}} \phi\left(\frac{y_t - \mu_{2,t}}{\sigma_{2,t}}\right), \quad (4) \\ \mu_{1,t} &= 0.7033 + 1.108y_{t-1} - 0.2862y_{t-2}, \\ \mu_{2,t} &= 0.983 + 1.5279y_{t-1} - 0.8884y_{t-2}, \\ \sigma_{1,t} &= \sqrt{0.007 + 0.0194(y_{t-1} - \mu_{1,t})^2}, \\ \sigma_{2,t} &= \sqrt{0.0445 + 0.0041(y_{t-1} - \mu_{2,t})^2}. \end{aligned}$$

因为混合分布可能呈现多峰, 若采用传统点预报(条件期望)方法往往得不到好的预报结果. 图3给出由模型(4)在 $t = 85 \sim 90$ 上得到的一步预报条件密度函数. 由图可知序列在该区间上的一步预报密度函数的分布形式灵活多变, 并且在 $t = 86 \sim 89$ 时刻均有明显的双峰特征, 说明序列在这些时刻可能有较大的波动. 这与图2观测到的这一时间段居于较窄的波谷从而可能有较大的跳跃是一致的. 尽管在 $t = 85, 90$ 时刻密度函数呈现单峰, 但对应的分布却不再是Gauss分布. 除了直观上观测到得分布具有非对称性以外, 通过 Monte Carlo方法还求出在 $t = 85$ 时刻分布函数

对应的峰度以98%的概率小于0,在 $t = 90$ 时刻对应的峰度以95.5%的概率大于0.这说明模型(4)对序列在这两个时刻分布的薄尾和厚尾特征分别进行了描述.

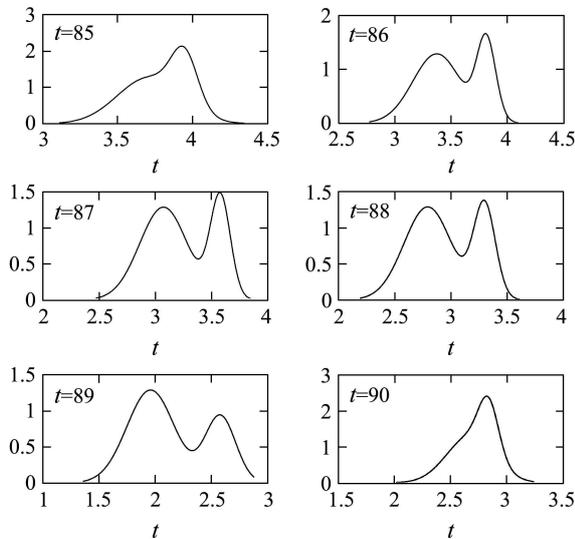


图3 模型(4)在区间 $t = 85 \sim 90$ 上各时刻对应的
一步预报条件密度函数

Fig. 3 One-step predictive distributions of the series($t = 85 \sim 90$)

Moran(1953)^[11]曾用AR模型对该常用对数序列进行分析,得到下面AR(2)模型:

$$y_t = 1.05 + 1.41y_{t-1} - 0.77y_{t-2} + \varepsilon_t,$$

其中 $\varepsilon_t \sim N(0, 0.04591)$.

Tong(1990)^[12]用自激励门限自回归(SETAR)模型研究该组数据,得到的其中一个为SETAR(2; 7, 2)模型:

$$y_t = \begin{cases} 0.546 + 1.032y_{t-1} - 0.173y_{t-2} + \\ 0.171y_{t-3} - 0.431y_{t-4} + 0.332y_{t-5} - \\ 0.284y_{t-6} + 0.210y_{t-7} + \varepsilon_t^{(1)}, & y_{t-2} \leq 3.116; \\ 2.632 + 1.492y_{t-1} - \\ 1.324y_{t-2} + \varepsilon_t^{(2)}, & y_{t-2} > 3.116. \end{cases}$$

其中 $\varepsilon_t^{(1)} \sim N(0, 0.0259)$, $\varepsilon_t^{(2)} \sim N(0, 0.0505)$.

Wong和Li(2000)^[3]运用GMTD模型对该数据建模,得到如下模型:

$$f(y_t|y_{t-1}) = \frac{0.7937}{0.1961} \phi\left(\frac{y_t - 0.6185 - 0.8387y_{t-1}}{0.1961}\right) + \frac{0.2063}{0.2015} \phi\left(\frac{y_t - 0.8154y_{t-1}}{0.2015}\right).$$

Wong和Li(2000)^[3]运用MAR模型对该数据建模,

得到如下最优MAR(2; 2, 2)模型:

$$\mu_{1,t} = 0.7107 + 1.1022y_{t-1} - 0.2835y_{t-2};$$

$$\mu_{2,t} = 0.9784 + 1.5279y_{t-1} - 0.8871y_{t-2};$$

$$f(y_t|y_{t-2}) = \frac{0.3163}{0.0887} \phi\left(\frac{y_t - \mu_{1,t}}{0.0887}\right) + \frac{0.6837}{0.2020} \phi\left(\frac{y_t - \mu_{2,t}}{0.2020}\right).$$

从以上AR(2)、MAR和EHMTD模型中可以观察到一个有趣的结果.除常数项外,MAR模型第一个分量对应的均值方程、EHMTD模型第一个分量对应的均值方程与AR(2)模型彼此非常类似,另外EHMTD模型与MAR模型对应的系数也比很类似,而这几个模型却是用不同的参数估计方法得到的.可见它们之间有着内在的联系,MAR模型是对AR(2)模型的推广,EHMTD模型是对MAR模型的进一步推广.

下面在表2中给出上述模型由模拟实验得到的一步预报的结果比较:

表2 不同模型基于序列一步预报的结果

Table 2 Empirical coverages of the $(1 - \alpha)100\%$ prediction intervals for the series

模型	95%	90%	80%	70%	60%	50%
EHMTD	95.18	89.29	81.34	70.63	62.50	51.96
AR	92.86	88.61	75.89	68.75	58.93	50.00
SETAR	95.33	90.65	81.83	69.16	56.07	44.86
GMTD	95.54	91.96	81.25	72.32	59.82	45.54
MAR	96.43	88.39	82.14	70.54	63.39	52.68

由表2可见,AR模型、SETAR模型整体上给出的预报结果相对较差.因为该数据已通过多种方法证明为非线性序列^[11],要用线性AR模型对非线性序列建模,结果自然不会很好.尽管SETAR是发展比较成熟的非线性时间序列模型,但由于其分布形式不够灵活,也未能给出满意的预报结果. GMTD, MAR, EHMTD模型给出了相对较好的结果.但由于GMTD模型施加了许多的限制,使其描述分布的能力受到影响.有上表可见,在50%水平上, GMTD模型的预报结果较差. MAR和EHMTD模型在各个水平上的预报结果都比较稳定,同时从整体上也可以看到EHMTD模型的预报结果要优于MAR模型.这是因为EHMTD模型考虑到了异方差,是对MAR模型进一步的推广,能够描述更为复杂的非线性现象.

6 结论(Conclusions)

本文研究了EHMTD模型的平稳性条件、厚尾特征及建模方法. 运用ECM算法来估计模型的参数. 通过仿真分析, 说明了该法在样本量增大的情形下, 估计结果是渐近趋于真值的. 由实例分析结果可见, EHMTD模型结构简单, 易于建模. 由于条件分布形式灵活, 能够描述非对称、多峰、厚尾(或薄尾)等非Gauss特征. EHMTD模型是一类有着较强应用背景和广泛应用前景的非线性时间序列模型.

参考文献(References):

- [1] MCLACHLAN G J, PEEI D. *Finite Mixture Model*[M]. A Wiley-Interscience Publication: John Wiley & Sons, INC, 2006.
- [2] LE N D, MARTIN R, RAFTERY A E. Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models[J]. *Journal of Robotic System*, 1984, 12(2): 123 – 140.
- [3] WONG C S, LI W K. On a mixture autoregressive model[J]. *Journal of Royal Statistics Society B*, 2000, 62(1): 95 – 115.
- [4] WONG C S, LI W K. On a mixture autoregressive conditional heteroscedastic model[J]. *Journal of the American Statistical Association*, 2001, 96(455): 982 – 995.
- [5] WONG C S, LI W K. On a logistic mixture autoregressive model[J]. *Biometrika*, 2001, 88(3): 833 – 846.
- [6] SHAO Q. Mixture periodic autoregressive time series models[J]. *Statistics & Probability Letters*, 2006, 76(1): 609 – 618.
- [7] 王红军, 田铮. 一非线性时间序列建模的混合ARMA模型[J]. *控制理论与应用*, 2005, 22(6): 875 – 881.
- (WANG Hongjun, TIAN Zheng. Mixture autoregressive moving average model for modeling nonlinear time series[J]. *Control Theory & Applications*, 2005, 22(6): 875 – 881.)
- [8] BERCHTOLD A. Mixture transition distribution(MTD) modeling of heteroscedastic time series[J]. *Computational Statistics & Data Analysis*, 2003, 41(3): 399 – 411.
- [9] 王红军, 田铮, 杨政. 异方差混合双AR模型[J]. *系统仿真学报*, 2007, 19(3): 497 – 501.
(WANG Hongjun, TIAN Zheng, YANG Zheng. Heteroscedastic mixture double-autoregressive model[J]. *Journal of System Simulation*, 2007, 19(3): 497 – 501.)
- [10] BENES V E. Existence of finite invariant measures for markov processes[J]. *Proceedings of the American Mathematical Society*, 1967, 18(6): 1058 – 1061.
- [11] FAN J Q, YAO Q W. *Nonlinear Time series: Nonparametric and Parametric Methods*[M]. Beijing: Science Press, 2006.
- [12] TONG H. *Non-linear Time Series* [M]. New York: Oxford University Press, 1990.

作者简介:

王红军 (1975—), 男, 理学博士, 西安电子科技大学理学院数学系教师, 主要研究方向为非线性时间序列分析与信息处理, E-mail: hjwang80@163.com;

田铮 (1948—), 女, 西北工业大学理学院应用数学系博士生导师、计算机科学与工程系博士生导师, 主要从事非线性时间序列分析与信息处理、非线性多尺度随机模型与图像处理等方面的研究, E-mail: Zhtian@nwpu.edu.cn;

韩四儿 (1975—), 男, 西北工业大学理学院应用数学系博士研究生, 主要研究方向为非线性时间序列分析与信息处理, E-mail: hansier2003@126.com.