

非线性时间序列建模的异方差混合双AR模型

王红军¹, 田 铮^{1,2}, 党怀义³

(1. 西北工业大学 理学院 应用数学系, 陕西 西安 710072;

2. 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; 3. 中国飞行试验研究院, 陕西 西安 710089)

摘要: 研究了可用于非线性时间序列建模的异方差混合双自回归模型(heteroscedastic mixture double-autoregressive model, HMDAR), 给出了HMDAR模型的平稳性条件, 利用ECM (expectation conditional maximization) 算法来估计模型的参数, 运用BIC(Bayes information criterion) 准则来选择模型. HMDAR模型条件分布富于变化的特征使它能够对具有非对称或多峰分布的序列进行建模, 将HMDAR模型应用于几个模拟和实际数据集均得到了较为满意的结果, 特别是对波动较大的序列, HMDAR模型能比其他模型更好地捕捉到数据序列的特征.

关键词: 异方差混合双自回归模型; 平稳性; BIC准则; ECM算法; 非对称分布; 多峰分布; 条件异方差

中图分类号: TP391, N93 **文献标识码:** A

Heteroscedastic mixture double-autoregressive model for modeling nonlinear time series

WANG Hong-jun¹, TIAN Zheng^{1,2}, DANG Huai-yi³

(1. Department of Applied Mathematics, School of Science, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China;

2. National Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China;

3. Chinese Flight Test Establishment, Xi'an Shaanxi 710089, China)

Abstract: The heteroscedastic mixture double-autoregressive (HMDAR) model, which is designed to model nonlinear time series, is investigated in this paper. Firstly, the stationary conditions are derived. Secondly, the parameters are estimated via expectation conditional maximization (ECM) algorithm. Thirdly, the Bayes information criterion (BIC) is used to select the model. The varied feature of conditional distributions makes the HMDAR model capable of modeling time series with asymmetric or multimodal distribution. Finally, the model was applied to several simulated and real data sets with satisfactory results. Especially to a highly time-variant series, the HMDAR model shows better performance than other competing models in data capturing.

Key words: heteroscedastic mixture double-autoregressive model; stationarity; Bayes information criterion; expectation conditional maximization (ECM)algorithm; asymmetric distribution; multimodal distribution; conditional heteroscedasticity

1 引言(Introduction)

时间序列分析的主要目的之一是为了预报. 时间序列模型的最佳预报与最优控制是时序方法在工程应用中的一个重要方面. ARMA模型已被成功的运用于工程系统的预测, 并与控制理论相结合, 用于工程系统的控制^[1]. 但是如果对工程中的非线性系统仍然采用线性时间序列中的ARMA模型来描述, 则不可能获得好的预报结果, 继而也得不到最优控制. 线性时间序列模型的噪声项通常假定服从Gauss分布. 在该假设下, 序列的边缘及条件分布均服从

Gauss分布. 但在工程和金融等大量实际问题中, 具有非Gauss特征的序列却广为存在. 由于混合模型提供了一种可以逼近任何复杂分布形式的有效方法, 该类模型受到研究者的广泛关注^[2]. 一些研究者将混合模型这一有力工具应用于时间序列分析, 得到了一类可用简单结构描述复杂非线性现象的混合时间序列模型.

为了对时间序列中的非Gauss特征(如奇异点、厚尾、变点等)进行描述, Le, Martin和Raftery^[3]提出Gauss混合转移分布模型(Gauss mixture transition

distribution model, GMTD). 该模型对于一些非线性特征给出了很好的解释,但是它不能描述具有循环特征的数据,而且分布形式不够灵活. 为了克服这些缺陷, Wong和Li^[4]将其推广得到混合自回归模型(mixture autoregressive model, MAR). Wang和Tian^[5]进一步将该类模型推广得到了混合ARMA模型(mixture autoregressive moving-average model, MARMA). GMTD模型、MAR模型和MARMA模型有一个共同的假定,即每个分量在不同的时刻具有同方差. 将这一假定用于变化剧烈的时间序列(如某些工程数据或金融市场数据)分析显然是不恰当的. 通过让每个分量的标准差是以往观测序列的函数, Berchtold^[6]进一步将该类模型推广得异方差混合转移分布模型(heteroscedastic mixture transition distribution model, HMTD).

Berchtold^[6]根据分量的标准差与观测序列之间的不同函数关系,给出了HMTD模型不同的定义形式. 本文主要研究了具有如下结构的HMTD模型,称为异方差混合双自回归模型(heteroscedastic mixture double-autoregressive model, HMDAR). HMDAR模型的定义如下:

$$\begin{cases} F(x_t|x_1^{t-1}) = \sum_{g=1}^K \alpha_g \Phi\left(\frac{x_t - \mu_{g,t}}{\sigma_{g,t}}\right), \\ \mu_{g,t} = \varphi_{g,0} + \sum_{i=1}^{p_g} \varphi_{g,i} x_{t-i}, \quad p_g \geq 1, \\ \sigma_{g,t} = \sqrt{\theta_{g,0} + \sum_{j=1}^{q_g} \theta_{g,j} x_{t-j}^2}, \quad q_g \geq 1. \end{cases} \quad (1)$$

其中: $F(x_t|x_1^{t-1})$ 表示 x_t 的条件分布函数, $x_1^{t-1} = \sigma(x_{t-1}, x_{t-2}, \dots)$ 是由 $(x_{t-1}, x_{t-2}, \dots)$ 生成的 σ -域, $\Phi(\cdot)$ 是标准正态分布函数, 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HMDAR($K; p_1, p_2, \dots, p_K; q_1, q_2, \dots, q_K$).

HMDAR模型实际上是由 K 个具有异方差的AR分量混合得到的. 若假定各分量的标准差是常数,它就是MAR模型. 文中给出了HMDAR模型的平稳性条件. 通过模拟说明几个平稳与非平稳分量经混合可以得到一个平稳序列. 由于HMDAR模型的结构较MAR模型更为复杂,其中某些参数对应的似然方程无显式解,运用EM算法估计参数得不到满意的估计结果. 本文运用GEM算法^[7,8]中的ECM算法来估计HMDAR模型的参数. 模拟及实例分析的结果都表明ECM算法可以得到较好的估计结果. HMDAR模型建模中另一个重要的步骤是阶数的确定问题.

文中利用BIC准则选择和确定HMDAR模型的阶数. 通过对两个波动较大序列的实例分析,演示了HMDAR模型的建模方法,及分布形式灵活多变、具有描述多峰等复杂分布形式的优势.

2 HMDAR模型的平稳性(Stationarity of the HMDAR model)

基于模型(1),可以得到 x_t 的条件期望为

$$E(x_t|x_1^{t-1}) = \sum_{g=1}^K \alpha_g \mu_{g,t}.$$

如果序列 $\{X_t, t \in N\}$ 的波动较大, x_t 的条件分布可能呈现多峰. 此刻 x_t 的条件期望往往得不到好的预报结果.

HMDAR模型另一个重要的特征是可以描述变化的条件方差. x_t 的条件方差为

$$\begin{aligned} \text{var}(x_t|x_1^{t-1}) = \\ \sum_{g=1}^K \alpha_g \mu_{g,t}^2 + \sum_{g=1}^K \alpha_g \sigma_{g,t}^2 - \left(\sum_{g=1}^K \alpha_g \mu_{g,t}\right)^2. \end{aligned}$$

下面给出HMDAR模型的平稳性条件. 定理的证明运用了Le et al(1996)^[3]以及Wong和Li (2000)^[4]中所用的方法,用到Benes(1967)^[9]中的结论. 定理2.1给出HMDAR($K; p_1, \dots, p_K; q_1, \dots, q_K$)模型一阶平稳的充分必要条件. 定理2.2和定理2.3中给出HMDAR($K; 1, \dots, 1; q_1, \dots, q_K$)模型和HMDAR($K; 2, \dots, 2; q_1, \dots, q_K$)二阶平稳的充分必要条件.

定理 2.1 过程 $\{X_t, t \in N\}$ 一阶平稳的充分必要条件是方程

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

的解 $z_1, \dots, z_p$ 都在单位圆内,其中 $\varphi_{g,i} = 0 (i > p_K)$.

定理 2.2 假定 $\{X_t, t \in N\}$ 是由一个HMDAR($K; 1, \dots, 1; q_1, \dots, q_K$)模型所产生的一阶平稳过程,则该过程 X_t 二阶平稳的充分必要条件是方程

$$1 - \left[\sum_{g=1}^K \alpha_g (\varphi_{g,1}^2 + \theta_{g,1}) \right] z^{-1} - \sum_{j=2}^q \left(\sum_{g=1}^K \alpha_g \theta_{g,j} \right) z^{-j} = 0$$

的解 $z_1, \dots, z_p$ 都在单位圆内,其中 $\theta_{g,j} = 0 (j > q_g)$.

定理 2.3 假定 $X_t, t \in N$ 是由一个HMDAR($K; 2, \dots, 2; q_1, \dots, q_K$)模型所产生的一阶平稳过程,则该过程 X_t 二阶平稳的充分必要条件是方程

$$\begin{aligned} 1 - \left[\sum_{g=1}^K \alpha_g (\varphi_{g,1}^2 + \theta_{g,1}) \right] z^{-1} - \\ \left(\sum_{g=1}^K \alpha_g \varphi_{g,2}^2 + \frac{\varphi_{g,1}}{1 - \sum_{g=1}^K \alpha_g \varphi_{g,2}} \right) z^{-2} - \end{aligned}$$

$$\sum_{j=3}^q \left(\sum_{g=1}^K \alpha_g \theta_{g,j} \right) z^{-j} = 0$$

的解 $z_1, \dots, z_q$ 都在单位圆内, 其中 $\theta_{g,j}=0(j>q_g)$.

3 HMDAR模型的参数估计及模型的选择(Parameter estimation of the HMDAR model and the model selection)

由于HMDAR模型的一些参数对应的似然方程不存在显式解, 运用普通EM算法往往得不到好的估计结果. 文中采用GEM算法中的ECM算法来估计模型的参数. 假定 $X = (x_1, x_2, \dots, x_n)$ 是产生于HMDAR模型的一组观测值. $Z = (z_1, \dots, z_n)$ 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 来自模型(1)中第 g 个分量, 则 $z_{g,t} = 1$; 否则 $z_{g,t} = 0$; $\alpha = (\alpha_1, \dots, \alpha_{K-1})^T$, $\varphi_g = (\varphi_{g,0}, \dots, \varphi_{g,p_g})^T$, $\theta_g = (\theta_{g,0}, \dots, \theta_{g,q_g})^T$, $\Psi = (\phi_1^T, \dots, \phi_K^T; \theta_1^T, \dots, \theta_K^T)^T$, $\Theta = (\alpha^T, \varphi_1^T, \dots, \varphi_K^T, \theta_1^T, \dots, \theta_K^T)^T$, $g=1, \dots, K$. 令 $r = \max\{p_1, \dots, p_K; q_1, \dots, q_K\}$, 则引入添加数据 Z 后, 完整数据的对数似然函数为

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

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 z_t 的期望为

$$\begin{aligned} \tau_{g,t} = E(z_{g,t}|X, \Theta) = & \frac{(\alpha_g/\sigma_{g,t})f_g(\varepsilon_{g,t}/\sigma_{g,t})}{\sum_{g=1}^K (\alpha_g/\sigma_{g,t})f_g(\varepsilon_{g,t}/\sigma_{g,t})}, \\ & g=1, \dots, K. \end{aligned}$$

M步 把 $\tau_{g,t}$ 代入式(2), 通过极大化式(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 $\phi = (\phi_1^T, \dots, \phi_K^T)^T$ 满足方程

$$\begin{aligned} \sum_{t=r+1}^n \frac{\tau_{g,t} x_t w_j}{\sigma_{g,t}^2} = & \sum_{s=0}^{p_g} \hat{\varphi}_{g,s} \sum_{t=r+1}^n \frac{\tau_{g,t} w_j w_s}{\sigma_{g,t}^2}, \quad j=0, \dots, p_g. \end{aligned}$$

其中

$$w_j = \begin{cases} 1, & j=0, \\ x_{t-j}, & j \neq 0. \end{cases}$$

参数 $\theta = (\theta_1^T, \dots, \theta_K^T)^T$ 满足方程

$$\begin{cases} \sum_{t=r+1}^n \frac{\tau_{g,t} u_j}{\sum_{s=0}^{q_g} \hat{\theta}_{g,s} u_s} = \\ \sum_{t=r+1}^n \frac{\tau_{g,t} (x_t - \mu_{g,t})^2 u_j}{(\sum_{s=0}^{q_g} \hat{\theta}_{g,s} u_s)^2}, \\ j=0, \dots, q_g. \end{cases} \quad (3)$$

其中

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

当分量的方差为一常数时(即为MAR模型),

$$\hat{\theta}_{g,0} = \frac{\sum_{t=r+1}^n \tau_{g,t} (x_t - \mu_{g,t})^2}{\sum_{t=r+1}^n \tau_{g,t}},$$

否则, 式(3)的解是不唯一的, 因此采用更为广义的ECM算法. 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

$$m = 2K + \sum_{g=1}^K p_g + \sum_{g=1}^K q_g$$

次条件极大化: 在第 $i+1$ 迭代中, 记

$$\Psi^{(i)} = (\psi_1^{(i)}, \dots, \psi_m^{(i)}),$$

在得到 $Q(\Psi | \Psi^{(i)}, X)$ 后, 首先在 $\psi_2^{(i)}, \dots, \psi_m^{(i)}$ 保持不变条件下求使 $Q(\Psi | \Psi^{(i)}, X)$ 达到最大的 $\psi_1^{(i+1)}$, 然后在 $\psi_1 = \psi_1^{(i+1)}, \psi_j = \psi_j^{(i)}, j=3, \dots, m$ 的条件下求使 $Q(\Psi | \Psi^{(i)}, X)$ 达到最大的 $\psi_2^{(i+1)}$, 如此继续, 经过 m 次条件极大化, 就得到 $\Psi^{(i+1)} = (\psi_1^{(i+1)}, \dots, \psi_m^{(i+1)})$, 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.

HMDAR模型的选择包含两个方面的问题: 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自身的阶数 $p_1, \dots, p_K; q_1, \dots, q_K$. 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

$$\begin{aligned} \text{BIC} = & -2L + (2K-1 + \sum_{g=1}^K p_g + \sum_{g=1}^K q_g) \log(n-r). \end{aligned} \quad (4)$$

其中

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

4 仿真与实例分析(Simulation and example analysis)

4.1 仿真分析(Simulation analysis)

由前面的定理可知, 平稳性条件既要依赖于混合元的权重系数 α_k , 又要依赖于各HMDAR分量自回归部分的系数. 几个平稳AR过程的混合仍然为平稳过程, 但几个平稳与非平稳线性AR过程的混合也可能为平稳过程. 对于这个性质, 将通过下面HMDAR模型(5)和(6)模拟产生的数据进行说明. 由模型式(5)易见, 第1个分量对应的是一个非平稳且具有异方差的AR(1)过程, 第2个分量对应的是一个平稳的带有异方差的AR(1)过程, 但如图1(a)这两个带有异方差的AR模型混合后模拟产生的序列却是平稳的; 由模型(6)可知, 第1个分量对应的是一个非平稳且具有异方差的AR(1)过程, 第2, 3个分量对应的是平稳AR(1)过程, 如图1(b)这3个AR模型混合后模拟产生的序列却是平稳的.

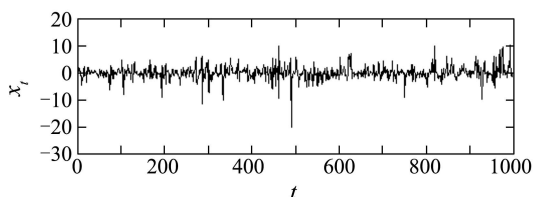


图 1(a) 模型(5)模拟产生的随机数序列

Fig. 1(a) Simulated time series of model (5)

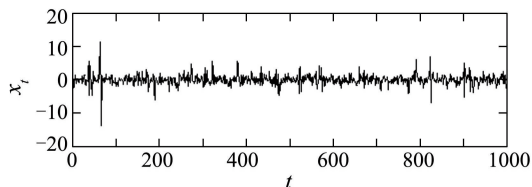


图 1(b) 模型(6)模拟产生的随机数序列

Fig. 1(b) Simulated time series of model (6)

$$F(x_t|x_1^{t-1}) = 0.3\Phi\left(\frac{x_t - 1.3x_{t-1}}{\sqrt{5 + 0.2x_{t-1}^2}}\right) + 0.7\Phi\left(\frac{x_t - 0.2x_{t-1}}{\sqrt{0.6 + 0.5x_{t-1}^2}}\right), \quad (5)$$

$$F(x_t|x_1^{t-1}) = 0.5\Phi\left(\frac{x_t - 0.2x_{t-1}}{\sqrt{0.6 + 0.5x_{t-1}^2}}\right) + 0.4\Phi\left(\frac{x_t - 0.3x_{t-1}}{\sqrt{0.5}}\right) + 0.1\Phi\left(\frac{x_t - 1.1x_{t-1}}{\sqrt{0.5}}\right). \quad (6)$$

下面运用ECM算法通过HMDAR模型模拟产生数据对其参数进行估计. 实验共进行500次, 每次模拟产生500个样本点. 表1和表2分别给出模型(5)和(6)模拟产生数据的估计结果. 由表中结果可知ECM算法对HMDAR模型参数估计的结果精度较高.

表 1 模型(5)模拟产生数据的估计结果

Table 1 Results of the simulation study with model (5)

k	α_k	ϕ_{k1}	$\theta_{k,0}$	$\theta_{k,1}$
真值	0.3000	1.3000	5.0000	0.2000
估计均值	0.3034	1.2740	5.0234	0.1903
估计值标准差	0.0753	0.2222	1.0497	0.0729
真值	0.7000	0.2000	0.6000	0.5000
估计均值	0.6966	0.1948	0.5910	0.4869
估计值标准差	0.0753	0.0974	0.1444	0.1009

表 2 模型(6)模拟产生数据的估计结果

Table 2 Results of the simulation study with model (6)

k	α_k	ϕ_{k1}	$\theta_{k,0}$	$\theta_{k,1}$
真值	0.5000	0.2000	0.6000	0.5000
估计均值	0.4799	0.2124	0.5813	0.5109
估计值标准差	0.1563	0.1576	0.2909	0.3823
真值	0.4000	0.3000	0.5000	
估计均值	0.4153	0.2976	0.5227	
估计值标准差	0.1496	0.1149	0.2710	
真值	0.1000	1.1000	5.0000	
估计均值	0.1048	1.1014	4.8576	
估计值标准差	0.0707	0.8592	2.8394	

4.2 实例分析(Example analysis)

化工过程温度读数序列(226个观测值^[10])是一组曾被许多研究者用不同模型分析过的数据. 原始序列见图2(a), 显然是非平稳的; 一阶差分序列见图2(b), 该序列难以确定其是否平稳, 许多研究者曾对该序列进行了建模分析, 将在下面介绍; 也有研究者认为有必要对该序列进行二阶差分, 以得到图2(c)中的平稳序列.

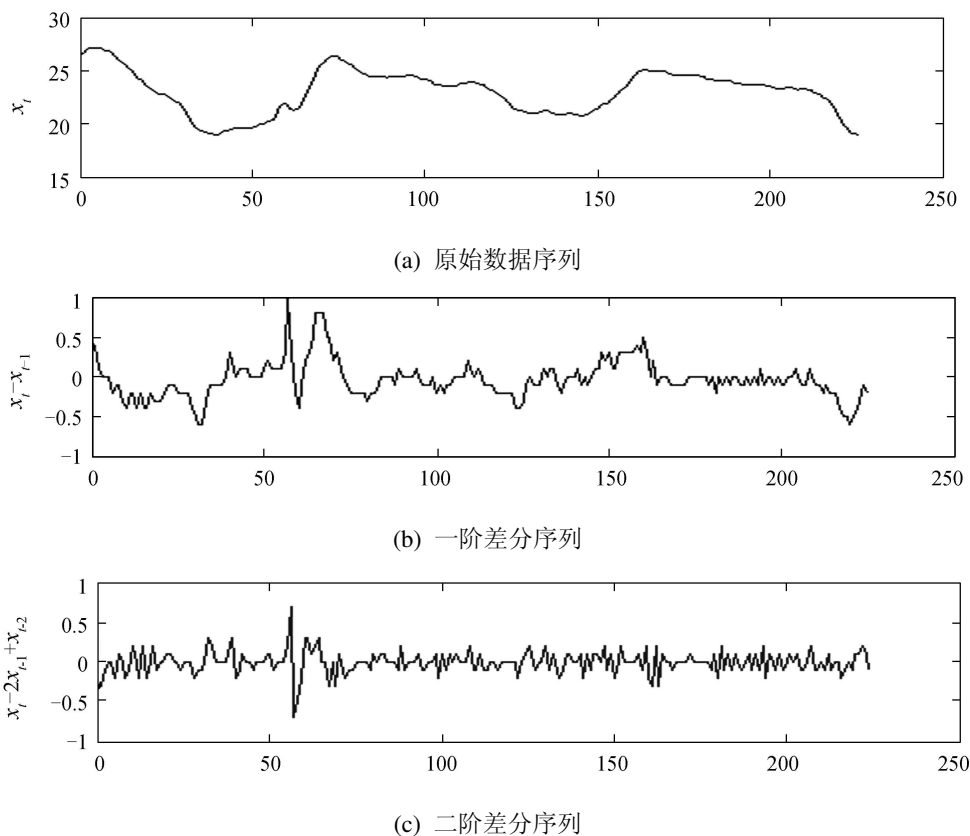


图2 化工过程温度读数序列

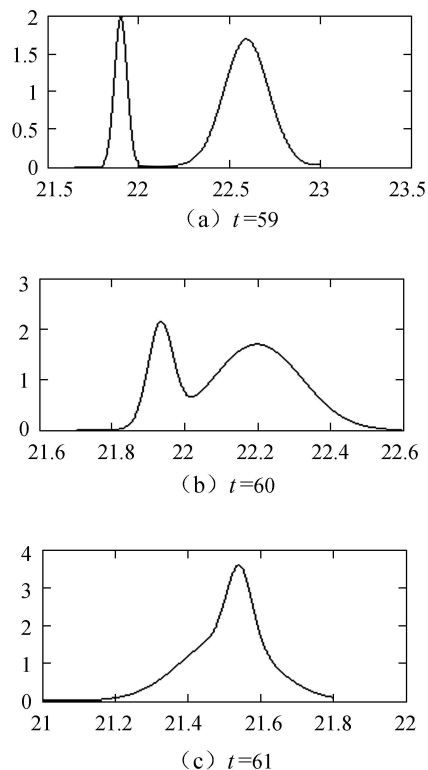
Fig. 2 Chemical process temperature series

用HMDAR模型对化工过程温度读数一阶差分序列进行建模分析, 根据BIC准则, 选用如下的HMDAR(3; 1, 1, 0; 0, 1, 1):

$$\begin{aligned}
 F(x_t|x_1^{t-1}) = & 0.5185\Phi\left(\frac{x_t - 0.0027 - 1.9911x_{t-1} + 0.9911x_{t-2}}{\sqrt{0.0149}}\right) + \\
 & 0.1733\Phi\left(\frac{x_t + 0.0814 - 1.3826x_{t-1} + 0.3826x_{t-2}}{\sqrt{0.0012(x_{t-1} - x_{t-2})^2}}\right) + \\
 & 0.3094\Phi\left(\frac{x_t - x_{t-1}}{\sqrt{0.7070(x_{t-1} - x_{t-2})^2}}\right).
 \end{aligned}
 \quad (7)$$

图3给出模型(7)在 $t = 59 \sim 64$ 上的一步预报的条件密度函数. 由图可知序列在区间 $t = 59 \sim 64$ 上的一步预报条件密度函数的分布形状灵活多变, 并且除了在 $t = 61, 63$ 时刻呈现出非对称的分布形式, 在 $t = 59, 60, 62, 64$ 时刻均有明显的双峰特征. 说明化工温度在该区间上变化剧烈. 这与图2所反映的原数据序列自 $t = 50$ 至 $t = 75$ 时刻内序列波动性较大是吻合的. 拟合模型得到的 x_t 的条件方差 $\text{var}(x_t|x_1^{t-1})$ 见图4. 由图可见, 该序列波动的大小是随时间而变化的, 并且在

自 $t = 50$ 至 $t = 75$ 时刻内波动较大, 这与前面观测到的信息一致. 表明该模型能够描述化工温度读数序列具有的条件异方差.



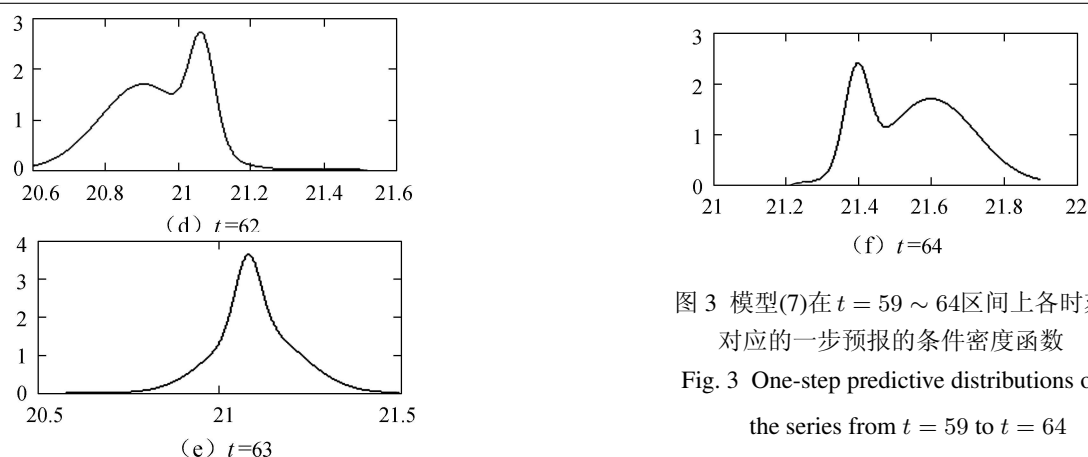


图3 模型(7)在 $t = 59 \sim 64$ 区间上各时刻对应的一步预报的条件密度函数

Fig. 3 One-step predictive distributions of the series from $t = 59$ to $t = 64$

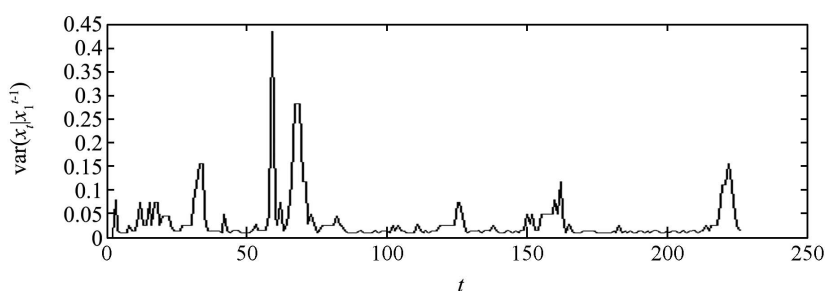


图4 用HMDAR模型拟合的化工温度序列

Fig. 4 Volatility computed from the fitted HMDAR model for the chemical process temperature series

Box^[10]用ARIMA模型分别对化工温度读数一阶差分序列和二阶差分序列进行建模,得到最优ARIMA(1, 1, 0)模型为

$$y_t - 1.82y_{t-1} + 0.82y_{t-2} = \varepsilon_t.$$

其中 $\varepsilon_t \sim N(0, 0.018)$. 得到最优ARIMA(0, 2, 2)模型为

$$y_t - 2y_{t-1} + y_{t-2} = \varepsilon_t - 0.13\varepsilon_{t-1} - 0.12\varepsilon_{t-2}.$$

其中 $\varepsilon_t \sim N(0, 0.019)$.

Wong和Li^[4]运用AR-ARCH模型对一阶差分序列建模,得到如下模型:

$$y_t = 1.8427y_{t-1} - 0.8427y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, h_t),$$

$$h_t = E(\varepsilon_t^2 | x_1^{t-1}) = 0.0098 + 0.4101\varepsilon_{t-1}^2.$$

在表3中给出对各模型对应的BIC值.从表3可见,仅从BIC值来比较,HMDAR(3; 1, 1, 0; 0, 1, 1)模型显然要优于其他模型.

表3 一些模型对应的BIC值

Table 3 BIC value of some models

模型	ARIMA (1,1,0)	ARIMA (0,2,2)	AR-ARCH	HMDAR (3;1,1,0;0,1,1)
BIC	-254	-230	-293	-541

下面在表4中给出上述模型由模拟实验得到的

基于整个序列一步预报的结果比较.由表4可知上述4种模型基于整个序列一步预报的结果难分优劣,在不同的置信水平上各有胜负.

表4 不同模型基于整个序列的一步预报结果的比较

Table 4 Empirical coverages of the $(1 - \alpha)100\%$ prediction intervals for the series %

模型	95	90	80	70	60	50
HMDAR	94.62	91.03	84.75	73.09	65.92	55.64
AR-ARCH	92.83	88.79	82.96	76.68	59.19	50.22
ARIMA(1, 1, 0)	96.86	93.72	82.51	77.58	69.51	57.40
ARIMA(0, 2, 2)	94.62	93.72	81.17	78.03	72.65	53.36

下面在表5中给出对这几个模型在波动较大的区间 $t = 30 \sim 75$ 上的预报结果.由表5可见HMDAR模型和AR-ARCH模型在区间 $t = 30 \sim 75$ 上的预报结果明显优于ARIMA(1, 1, 0)模型和ARIMA(0, 2, 2)模型.这是因为HMDAR模型和AR-ARCH模型都能描述条件异方差.同时由于AR-ARCH模型的分布形式相对HMDAR模型比较简单,HMDAR模型的预报结果又明显优于AR-ARCH模型.由于HMDAR模型既可以描述异方差,又可以描述复杂的分布形式,因此在波动较大的区间 $t = 30 \sim 75$ 上给出了相对其他3个模型较好的预报结果.

表5 不同模型在区间 $t = 30 \sim 75$ 上的一步预报结果的比较

Table 5 Empirical coverages of the $(1 - \alpha)100\%$ prediction intervals for the series

from $t = 30$ to $t = 75$							%
模型	95	90	80	70	60	50	
HMDAR	89.13	82.61	78.26	65.22	63.04	56.52	
AR-ARCH	89.13	78.26	79.31	67.39	50.00	36.96	
ARIMA(1, 1, 0)	84.78	76.09	69.57	54.35	47.83	38.27	
ARIMA(0, 2, 2)	78.26	73.91	65.22	60.87	50.00	41.30	

5 结论(Conclusion)

本文研究得到了HMDAR模型的平稳性条件及建模方法. 通过仿真分析, 说明在满足一定条件下, 几个具有异方差的平稳与非平稳AR过程的混合可得到一个平稳过程. 运用ECM算法来估计HMDAR模型的参数, 得到了较好的估计结果. 由实例分析结果可见, HMDAR模型结构简单, 易于建模. 条件分布形式灵活, 能够根据数据的特征及时实现从单峰到多峰或从多峰到单峰的相互转化, 并能描述条件异方差. 特别是对于波动较大的序列, HMDAR模型能够更好的抓住数据的特征, 给出了较好的预报结果.

参考文献(References):

- [1] 杨叔子, 吴雅. 时间序列分析的工程应用(第二册)[M]. 武汉: 华中理工大学出版社, 1994: 1-47.
(YANG Suzi, WU Ya. *Time Series Analysis in Engineering Application (The second volume)*[M]. Wuhan: Huazhong University of Science and Technology Press, 1994: 1-47.)
- [2] MCLACHLAN G J, PEEI D. *Finite Mixture Model*[M] // SHEWHART W A, WILKS S S. New York: John Wiley & Sons, Inc, 2001: 1-37.
- [3] LEND, MARTIN R, RAFTERY A E. Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution mod-

els[J]. *J of the American Statistical Associatio*, 1996, 91(436): 1504-1515.

- [4] WONG C S, LI W K. On a mixture autoregressive model[J]. *J of the Royal Statistical Society-B*, 2000, 62(part 1): 95-115.
- [5] 王红军, 田铮. 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 serie[J]. *Control Theory & Applications*, 2005, 22(6): 875-881.)
- [6] BERCHTOLD A. Mixture transition distribution(MTD) modeling of heteroscedastic time serie[J]. *Computational Statistics & Data Analysis*, 2003, 41(3): 399-411.
- [7] MCLACHLAN G J, KRISHNAN T. *The EM Algorithm and Extensions*[M] // SHEWHART W A, WILKS S S. New York: John Wiley & Sons, Inc, 1997: 166-175.
- [8] 茆诗松, 王静龙, 濮小龙. 高等数理统计[M]. 北京: 高等教育出版社; 柏林: Springer, 1998: 428-443.
(MAO Shisong, WANG Jinglong, PU Xiaolong. *Advanced Mathematical Statistics*[M]. Beijing: China Higher Education Press; Heidelberg, Berlin: Springer-Verlag, 1998: 428-443.)
- [9] BENES V E. Existence of Finite invariant measures for Markov processes[J]. *Proceedings of the American Mathematical Society*, 1967, 18(6): 1058-1061.
- [10] BOX G E P, JENKINS G M, REINSEL G C. *Time Series Analysis: Forecasting and Control*[M]. 3rd ed. Engle-wood Cliffs: Prentice-Hall, 1994: 101-102.

作者简介:

王红军 (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: hjwang_99@163.com;

田铮 (1948—), 女, 教授, 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;

党怀义 (1967—), 男, 研究员, 主要从事航空遥测数据处理技术、大型工程数据库系统技术、分布式网络数据处理技术研究等工作, 曾主持研究开发了多项大型工程网络应用系统, 曾获省、部级科技成果进步二等奖2项, 发表学术论文数篇, E-mail: dhy2000@sina.com.