Hysteretic autoregressive time series models

By GUODONG LI, BO GUAN, WAI KEUNG LI and PHILIP L. H. YU

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong

gdli@hku.hk   boyd.guan@gmail.com   hrntlwk@hku.hk   plhyu@hku.hk

Summary

This paper extends the classical two-regime threshold autoregressive model by introducing hysteresis to its regime-switching structure, which leads to a new model: the hysteretic autoregressive model. The proposed model enjoys the piecewise linear structure of a threshold model but has a more flexible regime switching mechanism. A sufficient condition is given for geometric ergodicity. Conditional least squares estimation is discussed, and the asymptotic distributions of its estimators and information criteria for model selection are derived. Simulation results and an example support the model.

Some key words: Autoregression; Conditional least squares; Geometric ergodicity; Hysteresis; Threshold model.

1. Introduction

Threshold time series models have been very successful since their introduction by Tong (1978) and Tong & Lim (1980); see also Chan (1993), Tsay (1998), Hansen (2000), and Ling & Tong (2005). Due to its piecewise linear nature, the threshold model can mimic nonlinear features such as resonance, limit cycles and time-irreversibility, and it is also easy to fit numerically. Tong (1990) gives a comprehensive exposition of such models.

However, there is a sudden change in the probability structure when a threshold process switches regimes, which may not be the case in the real world. It has been observed that threshold models usually work well except around the boundaries between different regimes (Wu & Chen, 2007). The smooth-transition threshold model (Chan & Tong, 1986; van Dijk et al., 2002) can reduce this problem to some extent, but it may not function well in complicated cases, and it is not piecewise linear. Hamilton (1989) and McCulloch & Tsay (1994) used the discrete-state Markov switching model to analyse financial and economic time series. Regime switching in that model is completely controlled by a latent random variable. The model enjoys some flexibility in the switching mechanism, but the fitted model may be difficult to interpret. Wu & Chen (2007) considered a threshold variable-driven switching model in which the switching mechanism is jointly controlled by a latent variable and some observable variables, but this lacks a physical interpretation.

Hysteresis has been widely observed in economics, engineering, mechanics, material science, etc. Consider the simple microeconomic example in Gocke (2002). A previously inactive firm may have to bear a market entry cost to produce a certain product, i.e., production will become active only when the price of the product, p_t, is high enough to cover both the entry cost, e_t, and the cost of ingredients, c_t. However, once production has started, the firm will become inactive if and only if p_t becomes too low to cover the cost c_t. As a result, when $c_t < p_t < c_t + e_t$, the active or inactive status remains unchanged. The original magnetic hysteresis of a single iron-crystal at the micro level exhibits exactly this pattern (Kneller, 1962), but hysteresis at the macro level has different and complicated patterns. All extant discussions of hysteresis in economics focus on the macro level, and there is no specific hysteretic model in econometrics and statistics. Camarero et al. (2006) and Proietti (2006) used nonstationarity to interpret hysteresis in macroeconomic time series, whilst Kannebley (2008) and Perez-Alonso & Sanzo (2011) considered threshold models.

From the viewpoint of threshold models, the hysteretic pattern at the micro level defines a regime switching mechanism, in which the active and inactive statuses correspond to two different regimes. This motivates us to propose a somewhat different type of piecewise linear model, which we call the hysteretic time series model. Consider a simple hysteretic process with hysteresis variable z_t and hysteresis zone $(r_L, r_U]$. The time series is at the lower regime when $z_t \leqslant r_L$, and at the upper regime when $z_t > r_U$, and the regime remains unchanged as long as z_t falls within the hysteresis zone. The Supplementary Material gives more details on the regime switching mechanisms of this model. This paper concentrates on the self-exciting hysteretic autoregressive model.

2. HYSTERETIC AUTOREGRESSIVE PROCESSES

Consider a self-exciting hysteretic autoregressive model with regime indicator R_t:

$$y_t = \begin{cases} x_t^{\mathrm{T}}\phi + \sigma_1\varepsilon_t, & R_t = 1, \\ x_t^{\mathrm{T}}\psi + \sigma_2\varepsilon_t, & R_t = 0, \end{cases} \qquad R_t = \begin{cases} 1, & y_{t-d} \leqslant r_L, \\ 0, & y_{t-d} > r_U, \\ R_{t-1}, & \text{otherwise,} \end{cases} \tag{1}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})^{\mathrm{T}}$, $\phi = (\gamma_1, \phi_1, \ldots, \phi_p)^{\mathrm{T}}$, $\psi = (\gamma_2, \psi_1, \ldots, \psi_p)^{\mathrm{T}}$, the ε_ts are independent and identically distributed random variables with mean zero and variance one, $\sigma_1 > 0$ and $\sigma_2 > 0$ are scalars, integer $d > 0$ is the delay parameter, and $r_L \leqslant r_U$ are the boundary parameters of the hysteresis zone. Model (1) includes the traditional two-regime threshold autoregressive model as a special case when $r_L = r_U$. The hysteretic model is also referred to as the buffered threshold model in Zhu et al. (2014).

From (1), the regime indicator takes the form

$$R_t = I(y_{t-d} \leqslant r_L) + I(r_L < y_{t-d} \leqslant r_U)R_{t-1}$$

$$= I(y_{t-d} \leqslant r_L) + \sum_{j=0}^{\infty} \prod_{i=0}^{j} I(r_L < y_{t-d-i} \leqslant r_U)I(y_{t-d-j-1} \leqslant r_L)$$

almost surely. When $r_L < r_U$, the regime indicator R_t depends on past observations that are infinitely far away, which renders the hysteretic model different from traditional threshold models (Tong, 1990; Hansen, 2000).

Suppose that $\sigma_1 = \sigma_2 = \sigma$. Denote $Y_t = (y_t, \ldots, y_{t-p+1}, R_t)^{\mathrm{T}}$, $e_t = (\sigma\varepsilon_t, 0, \ldots, 0)^{\mathrm{T}}$, $M_{0t} = \{m_{0t}, 0, \ldots, 0, I(y_{t-d} \leqslant r_L)\}^{\mathrm{T}}$,

$$M_{1t} = \begin{pmatrix} m_{1t} & m_{2t} & \cdots & m_{pt} & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & I(r_L < y_{t-d} \leqslant r_U) \end{pmatrix},$$

$m_{0t} = \gamma_1 I(A_t) + \gamma_2 I(A_t^c)$, and $m_{it} = \phi_i I(A_t) + \psi_i I(A_t^c)$ with $i = 1, \ldots, p$, where A_t is the event $\{y_{t-d} \leqslant r_L\} \cup \{r_L < y_{t-d} \leqslant r_U, R_{t-1} = 1\}$ and A_t^c is its complement. It can be verified that $Y_t = g(Y_{t-1}) + e_t$ with $g(Y_{t-1}) = M_{0t} + M_{1t}Y_{t-1}$. As a result, $\{Y_t\}$ is a Markov chain, and, by a method similar to that in Chan & Tong (1985), we can obtain its geometric ergodicity.

THEOREM 1. *Suppose that ε_t has a density function that is positive everywhere on $\mathbb{R}$ and $E(|\varepsilon_t|) < \infty$. If $\sum_{i=1}^{p} |\phi_i| < 1$ and $\sum_{j=1}^{p} |\psi_j| < 1$, then the multivariate process $\{Y_t\}$ is geometrically ergodic, and hence we can obtain the geometric ergodicity of the hysteretic process $\{y_t\}$.*

The case with $\sigma_1 \neq \sigma_2$ can be proved similarly. Chan & Tong (1985) derived the geometric ergodicity of two-regime threshold autoregressive models under the same conditions as in Theorem 1.

3. CONDITIONAL LEAST SQUARES ESTIMATION

3·1. *Estimation procedure*

Let $\lambda = (\theta^{\mathrm{T}}, r_L, r_U, d)^{\mathrm{T}}$ denote the parameter vector of model (1), where $\theta = (\phi^{\mathrm{T}}, \psi^{\mathrm{T}})^{\mathrm{T}}$. Let Θ be a compact subset of $\mathbb{R}^{2p+2}$, $[a, b]$ be a predetermined interval and $d_{\max}$ be a predetermined positive integer. For the true values of parameter vector λ, we assume that $\theta_0 = (\phi_0^{\mathrm{T}}, \psi_0^{\mathrm{T}})^{\mathrm{T}}$ is an interior point of Θ, $a < r_{0L} < r_{0U} < b$ and $d_0 \in D = \{1, \ldots, d_{\max}\}$.

Denote the function for the sum of squared errors by $L_n(\lambda) = \sum_{t=1}^{n} \{\varepsilon_t(\lambda)\}^2$, where

$$\varepsilon_t(\lambda) = (y_t - x_t^{\mathrm{T}}\phi) R_t(r_L, r_U, d) + (y_t - x_t^{\mathrm{T}}\psi)\{1 - R_t(r_L, r_U, d)\}.$$

Let $n_0 = \max(p, d_{\max})$. For time series $\{y_t, -n_0 + 1 \leqslant t \leqslant n\}$ generated by model (1), the regime indicator function $R_t(r_L, r_U, d)$ in $\varepsilon_t(\lambda)$ depends on past observations that are infinitely far away, and hence initial values are needed for fitting.

For fixed r_L, r_U and d, the first few observations of the hysteresis variable, say $y_{1-d}, \ldots, y_{t_0-d}$, may fall into the hysteresis zone $(r_L, r_U]$, such that we fail to identify the regimes of $y_1, \ldots, y_{t_0}$. We can simply assign them to the lower regime, and then denote the resulting regime indicator function by $\tilde{R}_t(r_L, r_U, d)$. The exact value of $R_{t_0+1}(r_L, r_U, d)$ is known because y_{t_0+1-d} lies outside the hysteresis zone, and it can be verified that $\tilde{R}_t(r_L, r_U, d) = R_t(r_L, r_U, d)$ as $t_0 < t \leqslant n$.

By replacing $R_t(r_L, r_U, d)$ in the definition of $\varepsilon_t(\lambda)$ with $\tilde{R}_t(r_L, r_U, d)$, we can denote the function $\tilde{\varepsilon}_t(\lambda)$, and hence $\tilde{L}_n(\lambda)$. Then the conditional least squares estimator can be defined as

$$\hat{\lambda}_n = (\hat{\theta}_n^{\mathrm{T}}, \hat{r}_L, \hat{r}_U, \hat{d})^{\mathrm{T}} = \underset{\lambda}{\operatorname{argmin}} \, \tilde{L}_n(\lambda).$$

Write $\tilde{R}_t = \tilde{R}_t(\hat{r}_L, \hat{r}_U, \hat{d})$ for simplicity. We further estimate σ_1^2 and σ_2^2 by $\hat{\sigma}_{1n}^2 = n_1^{-1} \sum_{t=1}^{n} (y_t - x_t^{\mathrm{T}}\hat{\phi}_n)^2 \tilde{R}_t$ and $\hat{\sigma}_{2n}^2 = n_2^{-1} \sum_{t=1}^{n} (y_t - x_t^{\mathrm{T}}\hat{\psi}_n)^2 (1 - \tilde{R}_t)$, respectively, where $\hat{\theta}_n = (\hat{\phi}_n^{\mathrm{T}}, \hat{\psi}_n^{\mathrm{T}})^{\mathrm{T}}$, $n_1 = \sum_{t=1}^{n} \tilde{R}_t$ and $n_2 = n - n_1$.

Let $X_t = [x_t^{\mathrm{T}}\tilde{R}_t(r_L, r_U, d), x_t^{\mathrm{T}}\{1 - \tilde{R}_t(r_L, r_U, d)\}]^{\mathrm{T}}$. For each fixed (r_L, r_U, d), the minimizer of $\tilde{L}_n(\lambda)$ has a closed form:

$$\tilde{\theta}_n(r_L, r_U, d) = \underset{\theta}{\operatorname{argmin}} \, \tilde{L}_n(\lambda) = \left(\sum_{t=1}^{n} X_t X_t^{\mathrm{T}} \right)^{-1} \sum_{t=1}^{n} X_t y_t. \tag{2}$$

The step function $\tilde{L}_n\{\tilde{\theta}_n(r_L, r_U, d), r_L, r_U, d\}$ can be minimized by searching amongst all possible jumps, i.e., $d \in D$ and $(r_L, r_U) \in \{(y_{t-d}, y_{s-d}) : 1 \leqslant t, s \leqslant n, a \leqslant y_{t-d} \leqslant y_{s-d} \leqslant b\}$; see Li & Li (2008, 2011).

For the initial values of the regime indicator function, we can instead assign these t_0 observations to the upper regime, and then denote the resulting function by $\tilde{R}_t^*(r_L, r_U, d)$. Let $\tilde{L}_n^*(\lambda)$ be the corresponding sum of squared errors, and $\hat{\lambda}_n^* = \operatorname{argmin}_{\lambda} \tilde{L}_n^*(\lambda)$. To improve accuracy, we can choose $\hat{\lambda}_n^*$ as the conditional least squares estimator when $\tilde{L}_n^*(\hat{\lambda}_n^*) < \tilde{L}_n(\hat{\lambda}_n)$.

3·2. *Asymptotic results*

Assumption 1. It holds that $\phi_0 \neq \psi_0$, $\operatorname{pr}(y_t \in [a, b]) < 1$ and that ε_t has a bounded, continuous and positive density on $\mathbb{R}$.

THEOREM 2. *If the time series $\{y_t\}$ is strictly stationary and ergodic with $E(|y_t|^{2+\delta}) < \infty$ for a small $\delta > 0$, and Assumption 1 holds, then $\hat{\lambda}_n \to \lambda_0$, $\hat{\sigma}_{1n}^2 \to \sigma_{01}^2$ and $\hat{\sigma}_{2n}^2 \to \sigma_{02}^2$ almost surely, where $\lambda_0 = (\theta_0^{\mathrm{T}}, r_{0L}, r_{0U}, d_0)^{\mathrm{T}}$, σ_{01}^2 and σ_{02}^2 are the corresponding true parameters.*

We can prove Theorem 2 by standard arguments for strong consistency. The delay parameter d takes only integer values, so $\hat{d}$ will equal d_0 when sample size n is sufficiently large. Without loss of generality, we assume that the true delay parameter, d_0, is known for the reminder of this subsection, and it is then deleted from parameter vector λ and corresponding functions.

Assumption 2. The time series $\{y_t\}$ is strictly stationary with $E(y_t^{4+\delta}) < \infty$ for a small $\delta > 0$, and $E(\varepsilon_t^4) < \infty$.

Assumption 3. The autoregressive function is discontinuous on the hysteresis zone $[r_{0L}, r_{0U}]$; i.e., there exist $p - 1$ constants $z_{p-1}, \ldots, z_{p-d+1}, z_{p-d-1}, \ldots, z_0$, such that $z^\mathrm{T}(\psi_0 - \phi_0) \neq 0$ for all $z_{p-d} \in [r_{0L}, r_{0U}]$, where $z = (1, z_{p-1}, \ldots, z_0)^\mathrm{T}$, and it is assumed that $d \leqslant p$ without loss of generality.

Let $Y_t = (y_t, \ldots, y_{t-p+1}, R_t)^\mathrm{T}$; then, from §2, $\{Y_t\}$ is a Markov chain. Denote its m-step transition probability function by $P^m(x, A)$, where $x \in \mathbb{R}^p \times \{0, 1\}$, $A \in \mathcal{B}_p \times \mathcal{U}$, $\mathcal{B}_p$ is the class of Borel sets of $\mathbb{R}^p$, and $\mathcal{U} = \{\varnothing, \{0\}, \{1\}, \{0, 1\}\}$.

Assumption 4. The time series $\{Y_t\}$ admits a unique invariant measure $\pi(\cdot)$, such that there exist $K > 0$ and $0 \leqslant \rho < 1$, for any $x \in \mathbb{R}^p \times \{0, 1\}$ and any m, $\|P^m(x, \cdot) - \pi(\cdot)\|_v \leqslant K(1 + \|x\|)\rho^m$, where $\|\cdot\|_v$ and $\|\cdot\|$ are, respectively, the total variation norm and Euclidean norm.

Under Assumption 4, $\{Y_t\}$ is said to be V-uniformly ergodic with $V(x) = K(1 + \|x\|)$, a condition stronger than geometric ergodicity; see Chapter 16 in Meyn & Tweedie (1993). Following Chan & Tong (1985) and Chan (1989), Assumption 4 is also implied by the conditions in Theorem 1.

THEOREM 3. *If Assumptions* 1–4 *hold, then*

(i) $n(\hat{r}_L - r_{0L}) = O_p(1)$, $n(\hat{r}_U - r_{0U}) = O_p(1)$ *and*

(ii) $n^{1/2} \sup_{n(|r_L - r_{0L}| + |r_L - r_{0L}|) \leqslant B} \|\tilde{\theta}_n(r_L, r_U) - \tilde{\theta}_n(r_{0L}, r_{0U})\| = o_p(1)$ *for any fixed* $0 < B < \infty$, *where* $\tilde{\theta}_n(r_L, r_U)$ *is defined as in* (2).

Furthermore,

$$n^{1/2}(\hat{\theta}_n - \theta_0) \to N\{0, \mathrm{diag}(\sigma_{01}^2 \Sigma_1^{-1}, \sigma_{02}^2 \Sigma_2^{-1})\}$$

in distribution as $n \to \infty$, *where* $\Sigma_1 = E(x_t x_t^\mathrm{T} R_t)$ *and* $\Sigma_2 = E\{x_t x_t^\mathrm{T}(1 - R_t)\}$.

Denote $\xi_{1t} = \sum_{j=0}^\infty [\{x_{t+j}^\mathrm{T}(\psi_0 - \phi_0)\}^2 + 2\sigma_2 x_{t+j}^\mathrm{T}(\psi_0 - \phi_0)\varepsilon_{t+j}]H_{t+j,j}$ and $\xi_{2t} = \sum_{j=0}^\infty [\{x_{t+j}^\mathrm{T}(\psi_0 - \phi_0)\}^2 - 2\sigma_1 x_{t+j}^\mathrm{T}(\psi_0 - \phi_0)\varepsilon_{t+j}]H_{t+j,j}$, where $H_{t,j} = \prod_{l=1}^j I(r_{0L} < y_{t-d+1-l} \leqslant r_{0U})$ with the convention $\prod_{l=1}^0 = 1$. For $i = 1$ and 2, let $F_{i,L}(\cdot \mid r)$ and $F_{i,U}(\cdot \mid r)$ be the conditional distribution functions of $\xi_{it}\{1 - R_{t-1}(r_0)\}$ and $\xi_{it} R_{t-1}(r_0)$ given $y_{t-d} = r$, respectively. Denote by $\pi(\cdot)$ the density function of y_t. We then define two independent one-dimensional two-sided compound Poisson processes:

$$\wp_L(s) = I(s < 0) \sum_{k=1}^{N_1^{(L)}(-s)} \zeta_k^{(1,L)} + I(s \geqslant 0) \sum_{k=1}^{N_2^{(L)}(s)} \zeta_k^{(2,L)},$$

and

$$\wp_U(s) = I(s < 0) \sum_{k=1}^{N_1^{(U)}(-s)} \zeta_k^{(1,U)} + I(s \geqslant 0) \sum_{k=1}^{N_2^{(U)}(s)} \zeta_k^{(2,U)},$$

where $\{N_1^{(L)}(s), s \geqslant 0\}$ and $\{N_2^{(L)}(s), s \geqslant 0\}$ are two Poisson processes with $N_1^{(L)}(0) = N_2^{(L)}(0) = 0$ and the same jump rate $\pi(r_{0L})$, $\{N_1^{(U)}(s), s \geqslant 0\}$ and $\{N_2^{(U)}(s), s \geqslant 0\}$ are another two Poisson processes with $N_1^{(U)}(0) = N_2^{(U)}(0) = 0$ and the same jump rate $\pi(r_{0U})$, $N_1^{(L)}(\cdot)$ and $N_1^{(U)}(\cdot)$ are left-continuous, $N_2^{(L)}(\cdot)$ and $N_2^{(U)}(\cdot)$ are right-continuous, and these four Poisson processes are independent. Each of the four sequences, $\{\zeta_k^{(i,j)}, k \geqslant 1\}$ with $i = 1$ or 2 and $j = L$ or U, consists of independent and identically distributed random variables, where $\zeta_k^{(i,L)}$ and $\zeta_k^{(i,U)}$ have distribution functions $F_{i,L}(\cdot \mid r_{0L})$ and $F_{i,U}(\cdot \mid r_{0U})$, respectively.

For $j = L$ and U, Assumption 3 implies that $E(\zeta_k^{(1,j)}) = E(\zeta_k^{(2,j)}) > 0$, and then $\wp_j(s)$ tends to $+\infty$ as $|s| \to \infty$. As in Li & Ling (2012), there exists a unique random square $[M_-^{(L)}, M_+^{(L)}) \times [M_-^{(U)}, M_+^{(U)})$ on which the process $\wp(z) = \wp_L(z_L) + \wp_U(z_U)$ attains its global minimum, where $z = (z_L, z_U)^\mathrm{T} \in \mathbb{R}^2$ and $[M_-^{(j)}, M_+^{(j)}) = \mathrm{argmin}_{s \in \mathbb{R}} \wp_j(s)$ with $j = L$ or U.

Table 1. *Percentages of correctly selected, overfitted, underfitted and
wrongly selected models by the* AIC *and* BIC, *respectively*

	AIC				BIC			
n	Correct	Over	Under	Wrong	Correct	Over	Under	Wrong
			Hysteretic autoregressive processes					
200	43·0	9·2	0·3	47·5	69·2	1·3	0·7	28·8
400	70·0	12·7	0·0	17·3	95·0	0·0	1·0	4·0
800	87·0	12·3	0·0	0·7	100·0	0·0	0·0	0·0
			Threshold autoregressive processes					
200	83·3	15·2	1·3	0·2	86·2	12·7	0·5	0·6
400	87·5	12·5	0·0	0·0	99·0	0·3	0·7	0·0
800	91·2	8·8	0·0	0·0	100·0	0·0	0·0	0·0

THEOREM 4. *If Assumptions 1–4 hold, then* $n(\hat{r}_L - r_{0L}) \to M_-^{(L)}$ *and* $n(\hat{r}_U - r_{0U}) \to M_-^{(U)}$ *in distribution as* $n \to \infty$. *Moreover,* $n(\hat{r}_L - r_{0L})$, $n(\hat{r}_U - r_{0U})$ *and* $n^{1/2}(\hat{\theta}_n - \theta_0)$ *are asymptotically independent.*

Theorem 4 can be obtained by combining the methods of proof for Theorem 3.3 in Li & Ling (2012) and Theorem 2.3 in Li et al. (2013).

3·3. *Model selection*

To select the order p of model (1), we consider the Bayesian information criterion,

$$\text{BIC}(p) = n_1 \log \hat{\sigma}_{1n}^2 + (p+1) \log n_1 + n_2 \log \hat{\sigma}_{2n}^2 + (p+1) \log n_2. \tag{3}$$

Similarly, we can define the Akaike information criterion, AIC, and the corrected AIC. Let $\hat{p}_n = \text{argmin}_{0 \leqslant p \leqslant p_{\max}} \text{BIC}(p)$, where $p_{\max}$ is a predetermined large order.

THEOREM 5. *If* $p_{\max} \geqslant p_0$ *and Assumptions 1–4 hold, then* $\text{pr}(\hat{p}_n = p_0) \to 1$ *as* $n \to \infty$, *where* p_0 *is the true order, i.e.,* $|\phi_{0p_0}| + |\psi_{0p_0}| > 0$.

We can show that minimization of the AIC tends to select an order that is greater than or equal to p_0. Wong & Li (1998) suggested information criteria for selecting the order of threshold models, and they are similar in form to (3). Moreover, we can consider different orders, say p_1 and p_2, for the two regimes of model (1) in the information criteria proposed above.

4. NUMERICAL STUDIES

4·1. *Simulation experiment*

In an experiment to evaluate the information criteria in §3·3, two data-generating processes are employed. The first is a three-regime threshold autoregressive model,

$$y_t = \begin{cases} -0\cdot02y_{t-1} + 0\cdot76y_{t-2} + \varepsilon_t, & y_{t-2} \leqslant -0\cdot9, \\ -0\cdot70y_{t-1} + 0\cdot20y_{t-2} + \varepsilon_t, & -0\cdot9 < y_{t-2} \leqslant 0\cdot5, \\ 0\cdot65y_{t-1} + 0\cdot32y_{t-2} + \varepsilon_t, & 0\cdot5 < y_{t-2}, \end{cases}$$

and the second is a hysteretic autoregressive model with the same structure in the lower and upper regimes and the same delay and boundary parameters, where $\{\varepsilon_t\}$ are independent random variables with the standard normal distribution. These models both have two boundary or threshold parameters.

We consider sample sizes, $n = 200$, 400 and 800, with 1000 replications for each sample size and data-generating process. The hysteretic model and three-regime threshold model are both fitted to each generated series with $d_{\max} = 4$, and the values of a and b are taken to be the 10th and 90th percentiles of each sample, respectively. The AIC and BIC are used in turn to select the model and order p with $p_{\max} = 4$. For the information criteria for the threshold model, see Wong & Li (1998). Table 1 lists the percentages

Table 2. *Information criteria of six fitted models*

	HAR	AR	TAR3-A	TAR3-B	TAR2-A	TAR2-B
BIC	1568·06	1692·45	1610·86	1588·96	1577·90	1575·48
AIC	1512·80	1651·67	1522·54	1529·62	1522·83	1526·01
AICc	1820·95	1955·76	1842·17	1842·99	1831·02	1833·74

AICc, corrected AIC; AR, autoregressive model; HAR, hysteretic autoregressive model; TAR3-A and TAR3-B, three-regime threshold autoregressive model selected by the AIC and BIC, respectively; TAR2-A and TAR2-B, two-regime threshold autoregressive model selected by the AIC and BIC, respectively.

of correctly selected, overfitted, underfitted and wrongly selected models. Wrong selection refers to the case of a threshold model being selected for a hysteretic process or a hysteretic model being chosen for a threshold process. The percentages of correct selection increase for both the AIC and BIC as sample size n increases, whilst the AIC has a slight tendency to select a bigger model even when $n = 800$. A larger proportion of hysteretic processes is wrongly selected as threshold processes, particularly when the sample size is as small as $n = 200$. This may be because the hysteretic model includes the two-regime threshold model as a special case.

4·2. *Annual sunspot numbers*

The sequence of Wolf sunspot numbers is very popular in time series analysis, and various linear and nonlinear models have been applied to explore it. See, for example, the two-regime threshold models in Tong (1990) and the three-regime threshold models in Tsay (1989). Hysteresis has been observed in some solar activity cycles (Dmitriev et al., 2002; Suyal et al., 2012), and the sunspot number is a measure of solar activity. As a result, the hysteretic model may be more suitable for sunspot numbers than the threshold model.

We consider the annual mean Wolf sunspot numbers from 1700 to 2013, with 314 observations in total. The hysteretic autoregressive model is first applied to this sequence with $d_{max} = 6$, and the values of a and b are the 10th and 90th percentiles of the data, respectively. We allow the orders in both regimes to differ for the sake of parsimony. The AIC and BIC are employed to perform model selection with $p_{max} = 13$, and they choose the same model. Moreover, as a comparison, we consider two other classes of competing models, namely, autoregressive models and three-regime threshold autoregressive models. Both information criterion select the same autoregressive model but different threshold autoregressive models. As a result, a total of four models are selected.

The fitted delay parameter is $\hat{d} = 2$ for both the hysteretic and threshold models, which is consistent with the results in Tsay (1989). The Supplementary Material gives the sample autocorrelation functions of the residuals from these four fitted models, and we can conclude that they are all adequate. Their information criterion values, including those of the AIC, BIC and corrected AIC, are listed in Table 2. It can be seen that the hysteretic model outperforms all three other models, especially the autoregressive model.

We also consider the class of two-regime threshold autoregressive models. Two models with different autoregressive orders are selected by the AIC and BIC. Their information criteria values are also given in Table 2. As expected, they are larger than the corresponding values of the fitted hysteretic model as the two-regime threshold model is a special case of the hysteretic model.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of Theorems 1–5, additional illustrations of the hysteretic models, more simulation results, more results on annual sunspot numbers and another example.

REFERENCES

CAMARERO, M., CARRION-I-SILVESTRE, J. L. & TAMARIT, C. (2006). Testing for hysteresis in unemployment in OECD countries: new evidence using stationarity panel tests with breaks. *Oxford Bull. Econ. Statist.* **68**, 167–82.

CHAN, K. S. (1989). A note on the geometric ergodicity of a Markov chain. *Adv. Appl. Prob.* **21**, 702–4.

CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520–33.

CHAN, K. S. & TONG, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Prob.* **17**, 666–78.

CHAN, K. S. & TONG, H. (1986). On estimating thresholds in autoregressive models. *J. Time Ser. Anal.* **7**, 178–90.

DMITRIEV, A. V., SUVOROVA, A. V. & VESELOVSKY, I. S. (2002). Expected hysteresis of the 23rd solar cycle in the heliosphere. *Adv. Space Res.* **29**, 475–9.

GOCKE, M. (2002). Various concepts of hysteresis applied in economics. *J. Econ. Surv.* **16**, 167–88.

HAMILTON, J. (1989). A new approach to the economics analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–84.

HANSEN, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68**, 575–603.

KANNEBLEY, S. J. (2008). Tests for the hysteresis hypothesis in Brazilian industrialized exports: a threshold cointegration analysis. *Econ. Mod.* **25**, 171–90.

KNELLER, E. (1962). *Ferromagnetismus*. Berlin: Springer.

LI, D. & LING, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *J. Economet.* **167**, 240–53.

LI, D., LING, S. & LI, W. K. (2013). Asymptotic theory on the least squares estimation of threshold moving-average models. *Economet. Theory* **29**, 482–516.

LI, G. & LI, W. K. (2008). Testing for threshold moving average with conditional heteroscedasticity. *Statist. Sinica* **18**, 647–65.

LI, G. & LI, W. K. (2011). Testing a linear time series model against its threshold extension. *Biometrika* **98**, 243–50.

LING, S. & TONG, H. (2005). Testing for a linear MA model against threshold MA models. *Ann. Statist.* **33**, 2529–52.

MCCULLOCH, R. & TSAY, R. S. (1994). Statistical analysis of economic time series via Markov switching models. *J. Time Ser. Anal.* **15**, 235–50.

MEYN, S. P. & TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. New York: Springer.

PEREZ-ALONSO, A. & SANZO, S. D. (2011). Unemployment and hysteresis: a nonlinear unobserved components approach. *Studies Nonlin. Dynam. Economet.* **15**, 1–27.

PROIETTI, T. (2006). Trend-cycle decompositions with correlated components. *Economet. Rev.* **25**, 61–84.

SUYAL, V., PRASAD, A. & SINGH, H. P. (2012). Hysteresis in a solar activity cycle. *Solar Phys.* **276**, 407–14.

TONG, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, Ed. C. H. Chen. Amsterdam: Sijthoff and Noordhoff, pp. 575–86.

TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

TONG, H. & LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. R. Statist. Soc.* B **42**, 245–92.

TSAY, R. S. (1989). Testing and modeling threshold autoregressive processes. *J. Am. Statist. Assoc.* **84**, 231–40.

TSAY, R. S. (1998). Testing and modeling multivariate threshold models. *J. Am. Statist. Assoc.* **93**, 1188–1202.

VAN DIJK, D., TERASVIRTA, T. & FRANSES, P. H. (2002). Smooth transition autoregressive models – a survey of recent developments. *Economet. Rev.* **21**, 1–47.

WONG, C. S. & LI, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models. *J. Time Ser. Anal.* **19**, 113–24.

WU, S. & CHEN, R. (2007). Threshold variable determination and threshold variable driven switching autoregressive models. *Statist. Sinica* **17**, 241–64.

ZHU, K., YU, P. L. H. & LI, W. K. (2014). Testing for the buffered autoregressive processes. *Statist. Sinica* **24**, 971–84.

[*Received February* 2014. *Revised March* 2015]