International Journal of Forecasting 8 (1992) 157–173
North-Holland

On continuous-time threshold autoregression

P.J. Brockwell

Statistics Department, Colorado State University, Fort Collins, CO 80523, USA

R.J. Hyndman

Statistics Department, University of Melbourne, Parkville, Vic. 3052, Australia

Abstract: The use of non-linear models in time series analysis has expanded rapidly in the last ten years, with the development of several useful classes of discrete-time non-linear models. One family of processes which has been found valuable is the class of self-exciting threshold autoregressive (SETAR) models discussed extensively in the books of Tong (1983, 1990). In this paper we consider problems of modelling and forecasting with continuous-time threshold autoregressive (CTAR) processes. Techniques for analyzing such models have been proposed by Tong and Yeung (1991) and Brockwell, Hyndman and Grunwald (1991). In this paper we define a CTAR(p) process $\{X(t)\}$ with boundary width $2\delta > 0$ as the first component of a p-dimensional Markov process $\{\boldsymbol{X}(t)\}$, defined by a stochastic differential equation. We are primarily concerned with the problems of model-fitting and forecasting when observations are available at times $1, 2, \ldots, N$; however, the techniques considered apply equally well to irregularly spaced observations. For practical computations with CTAR processes we approximate the process $\{X(t)\}$ by a linearly interpolated discrete-time Markov process whose transitions occur at times j/n, $j = 1, 2, \ldots$, with n large. This model is used to fit 'narrow boundary' CTAR models to both simulated and real data.

Keywords: Non-linear forecasting, Threshold autoregression, State–space representation, Maximum likelihood estimation, Continuous-time autoregression

1. Introduction

A discrete-time self-exciting autoregressive (or SETAR) process $\{X_t\}$ with delay parameter d is a solution of the equations,

$$X_t = a_0^{(i)} + \sum_{j=1}^{p} a_j^{(i)} X_{t-j} + \sigma^{(i)} e_t\,, \quad r_{i-1} \le X_{t-d} < r_i\,, \tag{1.1}$$

where $-\infty = r_0 < r_1 < \cdots < r_l = \infty$, $a_j^{(i)}$ and $\sigma^{(i)}$ (>0), $i = 1, \ldots, l$, are constants, and $\{e_t\}$ is a white noise sequence with unit variance. The thresholds are the levels $r_1, \ldots, r_{l-1}$. Thus, the real line is partitioned into l intervals, and X_t satisfies one of l autoregressive equations depending on the interval in which X_{t-d} falls. When $l = 1$, $\{X_t\}$ is an AR process. SETAR processes with appropriately chosen

Correspondence to: P.J. Brockwell, Statistics Department, Colorado State University, Fort Collins, CO 80523, USA.
We wish to acknowledge the valuable comments of J. de Gooijer and two referees which led to a number of improvements in the presentation of the material. The research was partially supported by NSF Grant DMS 9100392 (PJB) and a BHP Research Grant (RJH).

parameters can exhibit important features of real data sets which are not exhibited by Gaussian linear models, e.g. time-irreversibility, non-Gaussian marginal distributions and occasional bursts of outlying observations. For recent work in the theory and applications of SETAR processes, the book of Tong (1990) is an excellent reference. Owing to the difficulty of analyzing non-linear difference equations, theoretical results for the general SETAR model are far from complete. Necessary and sufficient conditions for the existence of a stationary solution of (1.1) when $p = 1$ have however been derived by Chan et al. (1985). There are many tests available for testing the appropriateness or otherwise of a linear model [see Tsay (1989) and Petruccelli (1990)]. The computer package STAR [which can be purchased from Microstar software using the order form in Tong (1990)] can be used for such testing and for the fitting of discrete time threshold models with up to three thresholds.

A continuous-time analogue of the non-linear model (1.1) is of considerable interest for a variety of reasons. Many data sets *are* in fact observations of continuous-time processes at discrete times. For such data one should allow for the possibility of multiple threshold crossings between successive observation times. CTAR models do this. As pointed out in the linear case by Jones (1981), the use of continuous-time models also facilitates the analysis of irregularly spaced data. Jones's technique is to use a state–space representation and the Kalman recursions to compute the Gaussian likelihood for any given set of parameter values. Tong and Yeung (1991), subsequently referred to as TY, have adapted this procedure to continuous-time threshold modelling and applied it to the analysis of several discrete-time series, including IBM closing stock prices, the Hang-sen Index and the Canadian Lynx data. They make the approximation that between successive observations in the same regime, the process evolves as the linear process appropriate to that regime. In this paper we consider a different approach based on an approximating sequence of Markov processes with transitions at times 0, $1/n, 2/n, \ldots$, where n is large. Such approximations are frequently used in the numerical study of diffusion processes [see, for example, Gikhman and Skorokhod (1969)]. We restrict attention here to models with one threshold only since the ideas extend immediately to models with more than one. Given observations at integer times of a continuous-time threshold autoregressive process, we are concerned with the computation of one-step forecasts, the corresponding mean squared errors, and the related problem of parameter estimation by maximization of the Gaussian likelihood. We analyse a number of simulated and real data sets, comparing the results with those obtained from the method of TY.

2. Linear CAR(p) processes

We begin with a brief account of linear continuous-time autoregressive processes of order $p \geq 1$. Formally we would like such a process to satisfy a pth order linear differential equation of the form

$$X^{(p)}(t) + a_1 X^{(p-1)}(t) + \cdots + a_p X(t) = \sigma W^{(1)}(t) + b\,, \tag{2.1}$$

where the superscript (j) denotes j-fold differentiation with respect to t, $\{W(t)\}$ is standard Brownian motion and $a_1, \ldots, a_p$, σ and b are constants. The derivative $W^{(1)}(t)$ does not exist in the usual sense, so we interpret (2.1) as being equivalent to the *observation* and *state* equations,

$$X(t) = [1 \quad 0 \ldots 0]\boldsymbol{X}(t)\,, \quad t \geq 0\,, \tag{2.2}$$

and

$$\mathrm{d}\boldsymbol{X}(t) = A\boldsymbol{X}(t)\,\mathrm{d}t + \boldsymbol{e}(b\,\mathrm{d}t + \sigma\,\mathrm{d}W(t))\,, \tag{2.3}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 \ldots 0 \\ 0 & 0 & 1 \ldots 0 \\ \vdots & \vdots & \vdots \ddots \vdots \\ 0 & 0 & 0 \ldots 1 \\ -a_p & -a_{p-1} & -a_{p-2} \cdots -a_1 \end{bmatrix}, \qquad \boldsymbol{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

b and σ^2 are drift and variance parameters, respectively, and (2.3) is an Itô differential equation for the state vector $\boldsymbol{X}(t)$. [We assume also that $\boldsymbol{X}(0)$ is independent of $\{W(t)\}$.] The state vector $\boldsymbol{X}(t)$ is in fact the vector of derivatives,

$$\boldsymbol{X}(t) = \begin{bmatrix} X(t) \\ X^{(1)}(t) \\ \vdots \\ X^{(p-1)}(t) \end{bmatrix}. \tag{2.4}$$

We say that $\{X(t), t \geq 0\}$ is a (stationary) CAR(p) process with parameters $(a_1, \ldots, a_p, b, \sigma)$ if $X(t) = [1 \quad 0 \ldots 0]\boldsymbol{X}(t)$, where $\{\boldsymbol{X}(t)\}$ is a stationary solution of (2.3). The other components of $\boldsymbol{X}(t)$ are the derivatives of $X(t)$ as in (2.4). The solution of (2.3) can be written as

$$\boldsymbol{X}(t) = \mathrm{e}^{At}\boldsymbol{X}(0) + \sigma \int_0^t \mathrm{e}^{A(t-u)}\boldsymbol{e}\, \mathrm{d}W(u) + b \int_0^t \mathrm{e}^{A(t-u)}\boldsymbol{e}\, \mathrm{d}u\,,$$

which is stationary if and only if

$$\boldsymbol{X}(0) \sim \mathrm{N}\Big(a_p^{-1} b[1 \quad 0 \ldots 0]',\ \sigma^2 \int_0^\infty \mathrm{e}^{Ay}\boldsymbol{e}\boldsymbol{e}'\, \mathrm{e}^{A'y}\, \mathrm{d}y\Big)$$

and all the eigenvalues of A (i.e. the roots of $z^p + a_1 z^{p-1} + \cdots + a_p = 0$) have negative real parts.

Notice that for the stationary CAR(p) process, the mean-corrected states and observations at times $t_1 < t_2 < \cdots$, i.e.

$$\boldsymbol{X}^*(t_i) = \begin{bmatrix} X(t_i) - b/a_p \\ X^{(1)}(t_i) \\ \vdots \\ X^{(p-1)}(t_i) \end{bmatrix}$$

and $X^*(t_i) = X(t_i) - b/a_p$, satisfy the discrete-time state and observation equations

$$\boldsymbol{X}^*(t_{i+1}) = \mathrm{e}^{A(t_{i+1}-t_i)}\boldsymbol{X}^*(t_i) + \boldsymbol{Z}(t_i)\,, \quad i = 1, 2, \ldots,$$

$$X^*(t_i) = [1 \quad 0 \ldots 0]\boldsymbol{X}^*(t_i)\,, \qquad i = 1, 2, \ldots,$$

where $\{\boldsymbol{Z}(t_i),\ i = 1, 2, \ldots\}$ is an independent sequence of Gaussian random vectors with mean, $\mathrm{E}[\boldsymbol{Z}(t_i)] = \boldsymbol{0}$ and covariance matrices

$$\mathrm{E}[\boldsymbol{Z}(t_i)\boldsymbol{Z}(t_i)'] = \sigma^2 \int_0^{t_{i+1}-t_i} \mathrm{e}^{Ay}\boldsymbol{e}\boldsymbol{e}'\, \mathrm{e}^{A'y}\, \mathrm{d}y\,.$$

These equations are in precisely the form needed for application of the Kalman recursions [see, for example, Brockwell and Davis (1991, Ch. 12)]. From these recursions we can easily compute $m_i = \mathrm{E}(X(t_i) \mid X(t_j), j < i)$, and $v_i = \mathrm{E}((X(t_i) - m_i)^2 \mid X(t_j), j < i)$, $i \leq 2$, and hence the likelihood of the observations $X(t_1), \ldots, X(t_N)$,

$$L = (2\pi)^{-N/2}(v_1 \ldots v_N)^{-1/2} \exp\Big[-\sum_{i=1}^{N} (X(t_i) - m_i)^2/(2v_i)\Big], \tag{2.5}$$

where $m_1 = a_p^{-1} b$ and v_1 is the top left component of the stationary state covariance matrix, $\sigma^2 \int_0^\infty e^{Ay} \boldsymbol{e}\boldsymbol{e}' e^{A'y}\, dy$. A non-linear optimization algorithm can then be used in conjunction with the expression for L to find maximum likelihood estimates of the parameters. The calculation of e^{At} and the integrals is most readily performed using the spectral representation of the matrix A. (The eigenvectors of A can easily be written down in terms of the eigenvalues and the coefficients, $a_1, \ldots, a_p$.) This is the method of Jones (1981) for maximum Gaussian likelihood fitting of CAR processes with possibly irregularly spaced data.

Example 1. If we use the above procedure to fit a CAR(2) model to the annual sunspot numbers, 1770–1869 [see, for example, Brockwell and Davis (1991, p. 6)], we obtain the model

$$X^{(2)}(t) + 0.495X^{(1)}(t) + 0.433X(t) = 24.7W^{(1)}(t) + 21.0\,,$$

with AIC $= -2 \ln L + 2p = 827.6$, which is less than that of the minimum AIC discrete time AR(p) model for the same data with $p \le 20$ (but not as small as that of the minimum AIC *subset* AR model in discrete time). The estimated standard deviations of the estimators of a_1, a_2, σ and b are 0.13, 0.07, 2.0 and 3.9, respectively.

If a continuous-time linear autoregressive process is observed at integer times only, the observed sequence (or skeleton) is then a discrete time ARMA($p, p-1$) process [see Phillips (1959)] with constraints on the coefficients. The joint distribution of the state vectors $\{\boldsymbol{X}(t),\ t = 1, \ldots, N\}$ is the limit as $n \to \infty$ of the joint distribution of $\{\boldsymbol{X}_n(t),\ t = 1, \ldots, N\}$, where $\boldsymbol{X}_n(0)$ has the same distribution as $\boldsymbol{X}(0)$ and

$$\boldsymbol{X}_n(t + n^{-1}) = (I + n^{-1}A)\boldsymbol{X}_n(t) + \boldsymbol{e}(n^{-1}b + n^{-1/2}\sigma Z(t))\,, \quad t = 0, 1/n, 2/n, \ldots, \tag{2.6}$$

where I is the $p \times p$ unit matrix and $\{Z(t)\}$ is an iid sequence with $P[Z(t) = 1] = P[Z(t) = -1] = 1/2$. The process $\{\boldsymbol{X}(t),\ t \ge 0\}$ itself is a multivariate diffusion process whose generator satisfies

$$Gf(\boldsymbol{x}) = \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x_p^2} - a_1 x_p \frac{\partial f}{\partial x_p} - \cdots - a_p x_1 \frac{\partial f}{\partial x_p} + b\frac{\partial f}{\partial x_p} + x_p \frac{\partial f}{\partial x_{p-1}} + \cdots + x_2 \frac{\partial f}{\partial x_1}\,,$$

for all twice continuously differentiable functions f on $\mathbb{R}^p$ such that each term on the right is bounded.

3. Continuous-time threshold autoregressive models

We define the CTAR(p) process with a single threshold at r and boundary width $2\delta > 0$ exactly as in (2.2) and (2.3), except that we allow the parameters $a_1, \ldots, a_p$, σ and b to depend on $X(t) = [1 \quad 0 \ldots 0]\boldsymbol{X}(t)$ in such a way that

$$\begin{aligned}
&a_1(X(t)) = a_1^{(1)}, \ldots, a_p(X(t)) = a_p^{(1)}\,, \qquad \sigma(X(t)) = \sigma^{(1)}\,, \qquad b(X(t)) = b^{(1)}\,,\\
&\quad \text{for } X(t) \le r - \delta\,,\\
&a_1(X(t)) = a_1^{(2)}, \ldots, a_p(X(t)) = a_p^{(2)}\,, \qquad \sigma(X(t)) = \sigma^{(2)}\,, \qquad b(X(t)) = b^{(2)}\,,\\
&\quad \text{for } X(t) > r + \delta\,,\\
&\text{and } a_1(X(t)), \ldots, b(X(t)) \text{ are defined by linear interpolation for } r - \delta < X(t) \le r + \delta\,.
\end{aligned} \tag{3.1}$$

Equivalently we can regard $X(t)$ as the first component of the p-dimensional diffusion process $\{\boldsymbol{X}(t)\}$ whose drift and diffusion parameters are determined by (3.1).

For numerical calculations we consider the approximating sequence of processes

$$X_n(t) = [1 \quad 0 \ldots 0]\boldsymbol{X}_n(t), \quad t = 0, 1/n, 2/n, \ldots, \tag{3.2}$$

where

$$\boldsymbol{X}_n(t + n^{-1}) = [I + n^{-1}A(X_n(t))]\boldsymbol{X}_n(t) + n^{-1}b(X_n(t))\boldsymbol{e} + n^{-1/2}\sigma(X_n(t))Z(t)\boldsymbol{e}, \tag{3.3}$$

and $A(X(t))$ is defined as in (2.3) with $a_1, \ldots, a_p$, b and σ dependent on $X(t)$ as in (3.1). We also assume that the initial state $\boldsymbol{X}_n(0)$ has the same distribution as $\boldsymbol{X}(0)$ and is independent of the binary iid sequence $\{Z(t), t = 1/n, 2/n, \ldots\}$ with $P[Z(t) = 1] = P[Z(t) = -1] = 1/2$. An adaptation of the argument in Theorem 1 of Gikhman and Skorokhod (1969, p. 460) shows that for $\delta > 0$, the finite-dimensional distributions of the process $\{\boldsymbol{X}_n(t), t \geq 0\}$ (with sample paths linear between $t = j/n$ and $t = (j+1)/n$, $j = 0, 1, 2, \ldots$) converge to those of $\{\boldsymbol{X}(t)\}$ as $n \to \infty$.

Note. Since numerical results based on the approximating discrete-time process are found to be essentially the same for all sufficiently small $\delta > 0$ as for $\delta = 0$, we shall base all our numerical results on the approximating process defined by (3.2) and (3.3) with $\delta = 0$ without making explicit reference each time to the value of δ. The results can be regarded as applying to a threshold CTAR process with *narrow boundary*.

The process $\{\boldsymbol{X}_n(t)\}$ defined by (3.3) is clearly Markovian. The conditional expectations

$$m_n(\boldsymbol{x}, t) = \mathrm{E}(X_n(t) \mid \boldsymbol{X}_n(0) = \boldsymbol{x})$$

satisfy the backward Kolmogorov equations

$$\begin{aligned} m_n(\boldsymbol{x}, t + n^{-1}) = {} & \tfrac{1}{2}m_n(\boldsymbol{x} + n^{-1}(A(x)\boldsymbol{x} + b(x)\boldsymbol{e}) + n^{-1/2}\sigma(x)\boldsymbol{e}, t) \\ & + \tfrac{1}{2}m_n(\boldsymbol{x} + n^{-1}(A(x)\boldsymbol{x} + b(x)\boldsymbol{e}) - n^{-1/2}\sigma(x)\boldsymbol{e}, t), \end{aligned} \tag{3.4}$$

with the initial condition

$$m_n(\boldsymbol{x}, 0) = x. \tag{3.5}$$

These equations clearly determine the moments $m_n(\boldsymbol{x}, t)$ uniquely. The higher order moments,

$$m_n^{(j)}(\boldsymbol{x}, t) = \mathrm{E}(X_n^j(t) \mid \boldsymbol{X}_n(0) = \boldsymbol{x}),$$

satisfy the same equation (3.4), and the slightly modified initial condition (3.5) with the right-hand side replaced by x^j.

Example 2. Exhibit 1 shows the one-step predictors $m_n((x, 0), 1)$ for the CTAR(2) process with parameters $r = 0$ and

$$a_1^{(1)}(x) = 5.0, \quad a_2^{(1)}(x) = 5.0, \quad \sigma^{(1)}(x) = 5, \quad b^{(1)}(x) = 20.0,$$

$$a_1^{(2)}(x) = 2.0, \quad a_2^{(2)}(x) = 2.0, \quad \sigma^{(2)}(x) = 10, \quad b^{(2)}(x) = 25.0.$$

Notice the deviation from linearity in x of the predictors. Values are given for $n = 10$, $n = 100$ and $n = 500$. The values for $n = 100$ and $n = 500$ are virtually indistinguishable and close also to the corresponding values for $n = 10$.

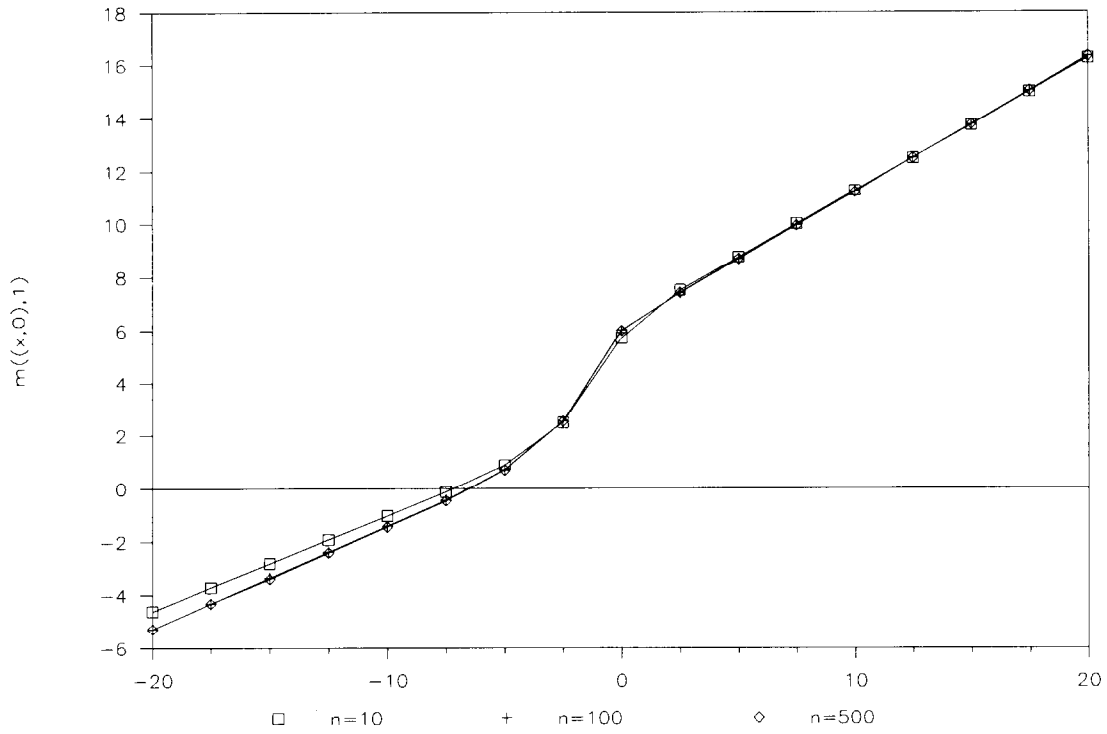

Exhibit 1. One-step predictors, Example 2.

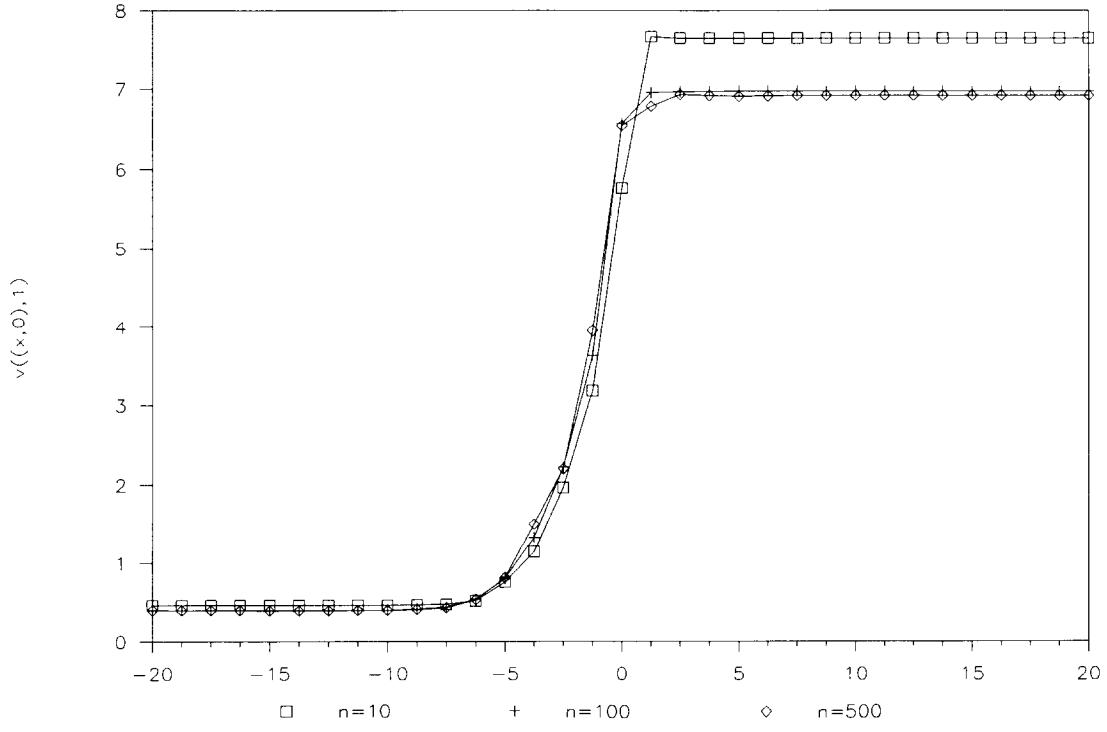

Exhibit 2. One-step MSE, Example 2.

Exhibit 2 shows the one-step mean squared errors,

$$v_n((x,0),1) = m_n^{(2)}((x,0),1) - m_n^2((x,0),1)\,,$$

for the same parameter values. Notice that as we move away from the boundary at $x=0$, the mean squared errors approach the corresponding values for the processes without any threshold. For $n=100$ and 500 these values are extremely close to the mean squared errors, 0.399 ($x<0$) and 6.874 ($x>0$) for the CAR(2) processes corresponding to the parameters below and above the threshold.

4. CTAR(1) processes

We will discuss the CTAR(1) process $\{X(t)\}$ in a little more detail since it is possible to state results more explicitly than in the case when $p \geq 2$. We shall assume that there is a single threshold at r with boundary width $2\delta > 0$.

The backward equation for the conditional expectation, $m(x,t) = \mathrm{E}(X(t) \mid X(0)=x)$, can be written as

$$\frac{\partial m}{\partial t} = \frac{\sigma^2(x)}{2}\frac{\partial^2 m}{\partial x^2} - a(x)x\frac{\partial m}{\partial x} + b(x)\frac{\partial m}{\partial x}\,, \tag{4.1}$$

with initial condition

$$m(x,0) = x\,. \tag{4.2}$$

[The jth moments, $m^{(j)}(x,1) = \mathrm{E}(X(1)^j \mid X(0)=x)$ satisfy the same equation (4.1) with initial condition $m^{(j)}(x,0) = x^j$.]

The discrete-time approximation (3.3) when $p=1$ takes the form

$$X_n(t+n^{-1}) = [1 - n^{-1}a(X_n(t))]X_n(t) + n^{-1}b(X_n(t)) + n^{-1/2}\sigma(X_n(t))Z(t)\,, \tag{4.3}$$

where $a(x)(=a_1(x))$, $b(x)$ and $\sigma(x)$ are defined as in (3.1). The corresponding backward equation for the conditional expectation $m_n(x,t) = \mathrm{E}(X_n(t) \mid X_n(0)=x)$ is therefore

$$\begin{aligned} m_n(x, t+n^{-1}) = {} & \tfrac{1}{2}m_n(x - n^{-1}(a(x)x - b(x)) + n^{-1/2}\sigma(x), t) \\ & + \tfrac{1}{2}m_n(x - n^{-1}(a(x)x - b(x)) - n^{-1/2}\sigma(x), t)\,, \end{aligned} \tag{4.4}$$

with initial condition

$$m_n(x,0) = x\,. \tag{4.5}$$

Example 3. Exhibit 3 shows the one-step predictors $m_n(x,1)$ for the CTAR(1) process with parameters $r=0$ and

$$a^{(1)}(x) = 0.5\,, \qquad \sigma^{(1)}(x) = 0.5\,, \qquad b^{(1)}(x) = 0\,,$$

$$a^{(2)}(x) = 1.0\,, \qquad \sigma^{(2)}(x) = 1.0\,, \qquad b^{(2)}(x) = 0\,.$$

Values are plotted for $n=10$ and $n=100$ together with a piecewise linear graph showing the one-step predictors for the linear processes defined by the parameter values below and above the threshold. The three graphs are virtually indistinguishable.

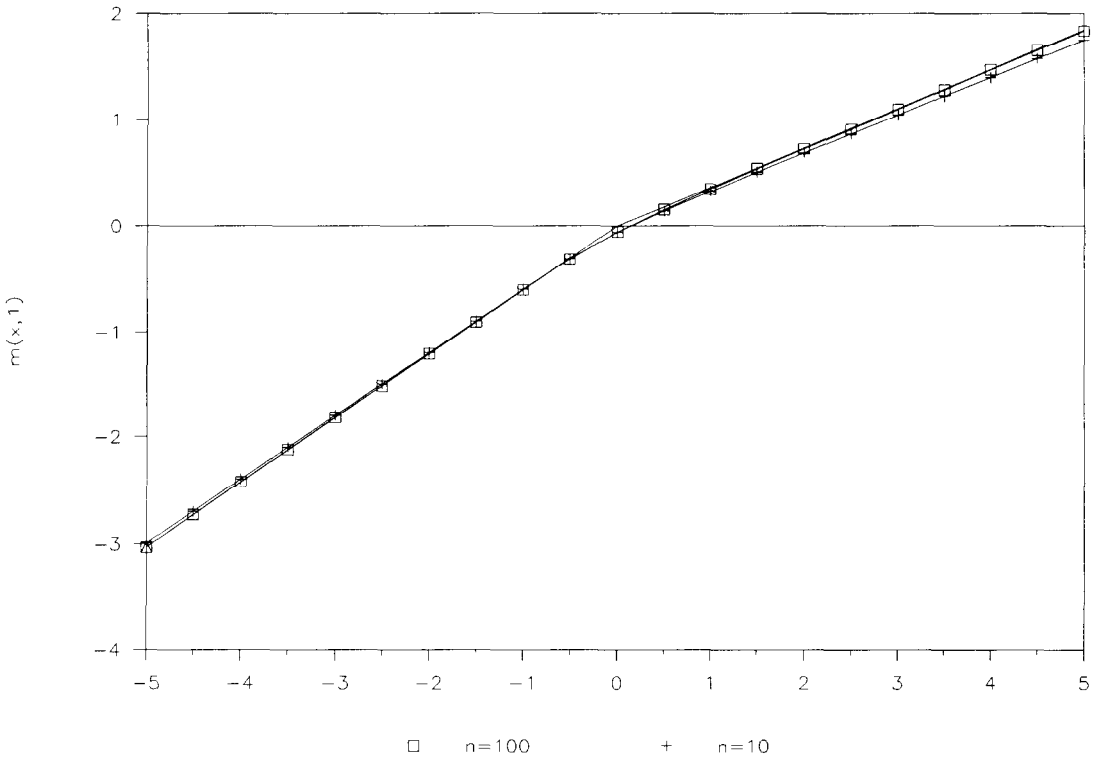

Exhibit 3. One-step predictors, Example 3.

Exhibit 4 shows the one-step mean squared errors,

$$v_n(x,1) = m_n^{(2)}(x,1) - m_n^2(x,1)\,,$$

for the same parameter values. The two constant values plotted above and below the threshold again correspond to the linear processes defined by the parameter values above and below the threshold.

Using a general result from Karlin and Taylor (1981, p. 221) we find that $\{X(t)\}$ has a stationary distribution if and only if $\lim_{x\to\pm\infty}[a(x)x^2 - 2b(x)x] > 0$ and that the stationary distribution, if this condition is satisfied, has probability density

$$\pi_\delta(x) = k\sigma^{-2}(x)\exp[-I(x)]\,,$$

where

$$I(x) = \int^x 2\sigma^{-2}(y)[a(y)y - b(y)]\,\mathrm{d}y$$

and k is the uniquely determined constant such that $\int_{-\infty}^{\infty}\pi_\delta(x)\,\mathrm{d}x = 1$. Substituting from (3.1) for $a(y)$, $b(y)$ and $\sigma(y)$, we can easily write down $\pi_\delta(x)$ quite explicitly. Here, however, we simply observe that the limit of this density as $\delta\to 0$ is

$$\pi_0(x) := \lim_{\delta\to 0}\pi_\delta(x) = c^{(i)}\exp\{-\sigma^{(i)-2}[a^{(i)}x^2 - 2b^{(i)}x]\}\,, \tag{4.6}$$

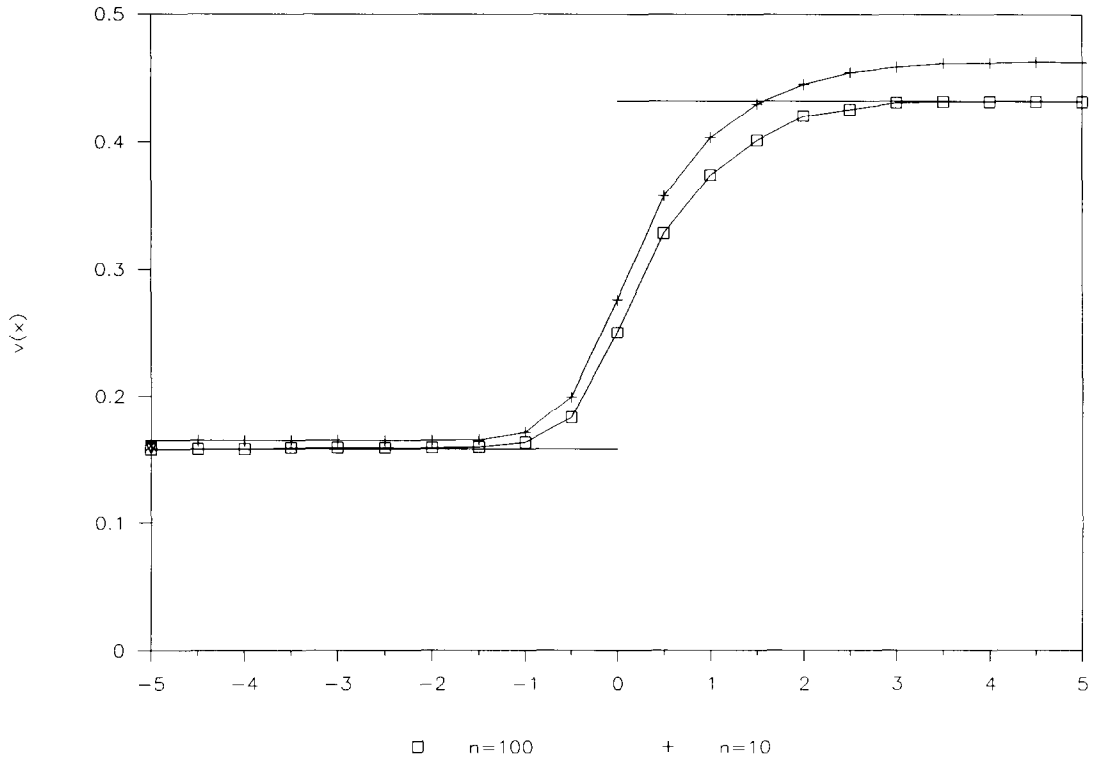

Exhibit 4. One-step MSE, Example 3.

where $i=1$ if $x<r$ and $i=2$ if $x>r$. The constants $c^{(1)}$ and $c^{(2)}$ are uniquely determined by the conditions $\int \pi_0(x)\,dx=1$ and $\sigma^{(1)^2}\pi(r-)=\sigma^{(2)^2}\pi(r+)$. Thus the density π_0 has a discontinuity at the boundary r if $\sigma^{(1)}\neq\sigma^{(2)}$.

We observe that π_0 is not the same as the stationary distribution of the CTAR(1) process (with boundary width zero) as defined by Brockwell, Hyndman and Grunwald (1991). The latter is a one-dimensional diffusion process with a boundary condition at the threshold r which gives (under the same conditions as above) a stationary distribution of the same algebraic form as $\pi_0(x)$, but satisfying a different boundary condition, namely

$$\sigma^{(1)}\pi(r-)=\sigma^{(2)}\pi(r+)\,.$$

This results in different values for the constants $c^{(1)}$ and $c^{(2)}$. We shall refer to the process defined by Brockwell, Hyndman and Grunwald as a *modified* CTAR(1) process. It can be approximated by modifying the approximating sequence (4.3) with $\delta=0$ so that whenever $X_n(t+n^{-1})$ as defined by (4.3) is on the opposite side of the boundary r from $X_n(t)$ we multiply $(X_n(t+n^{-1})-r)$ by $\sigma(X_n(t+n^{-1}))/\sigma(X_n(t))$ to allow for the change in σ on crossing the boundary.

5. Inference for CTAR(1) processes

In this section we consider the fitting of CTAR(1) models to observations $x(1),\ldots,x(N)$ at the discrete times $1,\ldots,N$. Our aim, like that of TY, will be to maximize the Gaussian likelihood of the data, namely

$$L(\boldsymbol{\theta}; x(1), \ldots, x(N)) = (2\pi)^{-N/2}(v(1) \ldots v(N))^{-1/2} \exp -\left[\sum_{j=1}^{N} (x(j) - m(j))^2/(2v(j))\right], \tag{5.1}$$

where $\boldsymbol{\theta}$ denotes the vector of parameters to be estimated, $m(j) = \mathrm{E}[X(j) \mid X(j-1) = x(j-1)]$ and $v(j) = \mathrm{E}[(X(j) - m(j))^2 \mid X(j-1) = x(j-1)]$ for $j \geq 2$, $m(1) = \mathrm{E}(X(1))$ and $v(1) = \mathrm{Var}(X(1))$. Our method of calculating the Gaussian likelihood, however, is different from theirs. For given parameter values $a^{(j)}$, $\sigma^{(j)}$, $b^{(j)}$, $j = 1, 2$, and threshold value r, we first compute the density (4.6) and its mean $m(1)$ and variance $v(1)$. The one-step predictors $m(j)$ and mean squared errors $v(j)$, $j = 2, \ldots, N$, are then approximated, for large n, by the solution of (4.4) and the analogous equation for the second moment, $\mathrm{E}[X_n(t)^2 \mid X_n(0) = x]$ (the one-step predictors and mean squared errors as a function of the initial state x are shown for Example 3 in Exhibits 3 and 4, respectively). For estimation of the parameters, $a^{(j)}$, $\sigma^{(j)}$, $b^{(j)}$ and r, a non-linear maximization algorithm is then used to maximize the Gaussian likelihood as computed from (5.1).

The method used by TY is analogous to that of the preceding paragraph except that it uses the piecewise linear predictors and piecewise constant mean square prediction errors for $X(t+1)$ computed from one of two linear processes depending on whether $X(t) < r$ or $X(t) > r$. These are illustrated for Example 3 in Exhibits 3 and 4. TY also make an adjustment when $X(t)$ and $X(t+1)$ are on opposite sides of the threshold. We have omitted this adjustment in our use of their method, since by allowing the predictor of $X(t+1)$ and its mean squared error to depend on $X(t+1)$, we obtained artificially low one-step mean squared errors and estimators which were not as good as those obtained without the adjustment.

Example 4. Twenty realizations of length 100 of the CTAR(1) process in Example 3 were generated using eq. (4.3) with $n = 1000$ and $\delta = 0$. For each realization we estimated the parameters by maximizing the Gaussian likelihood (a) using the Gaussian likelihood under the model (4.3) with $n = 10$ and (b) using the piecewise linear approximation to the one-step predictors and mean squared errors. The results were as shown in Exhibit 5.

Both methods gave reasonably good estimates of the true parameters, although Method (b) was notably less successful in estimating the AR coefficients. Method (b) did however locate the threshold well in each case, possibly because of the sharp discontinuity in the one-step prediction mean squared error at the boundary in the piecewise linear approximation, which makes the corresponding Gaussian likelihood sensitive to changes in r. Method (b) is also faster computationally. However, since the piecewise linear approximation is the same for both the CTAR(1) and modified CTAR(1) processes it is clear that Method (b) cannot distinguish between these processes while Method (a) can be used (by appropriate adjustment of the difference equations) to fit models of either type. Because of its speed, Method (b) is a useful source of preliminary estimates with which to initialize subsequent maximization with Method (a). The calculation of the Gaussian likelihood by Method (a) can of course be improved by increasing the chosen value of n, but only at the expense of increased computational time.

Example 5. We next consider the modelling of the IBM closing stock prices, 18 May 1961–30 March 1962 [listed in Tong (1990, p. 512)]. Following TY we shall consider the series of relative daily price

Exhibit 5

Estimated parameter	r	$a^{(1)}$	$\sigma^{(1)}$	$b^{(1)}$	$a^{(2)}$	$\sigma^{(2)}$	$b^{(2)}$
Actual value	0.000	0.500	0.500	0.000	1.000	1.000	0.000
Sample mean (a)	−0.044	0.503	0.467	−0.004	1.147	0.921	0.074
St. deviation (a)	0.151	0.174	0.158	0.143	0.307	0.146	0.223
Sample mean (b)	0.016	0.706	0.580	−0.099	2.796	1.113	0.245
St. deviation (b)	0.046	0.408	0.099	0.122	4.998	0.668	0.917

Exhibit 6

Estimated parameter	r	$a^{(1)}$	$\sigma^{(1)}$	$b^{(1)}$	$a^{(2)}$	$\sigma^{(2)}$	$b^{(2)}$	$-2\ln(GL)$	MSE
Linear model	0.00	1.55	1.69	0.07	1.55	1.69	0.07	590.6	0.879
CTAR(1) model (a)	−0.56	2.77	2.69	−0.84	0.40	1.13	−0.58	578.7	0.844
Piecewise model (b)	−0.36	4.54	2.94	−0.15	0.73	1.21	−0.20	577.9	0.833

changes, $X(t) = 100(P(t) - P(t-1))/P(t-1)$. We shall treat the data as uniformly spaced, although it would not be difficult to treat the data as irregularly spaced to take into account weekends and holidays (TY treat the data both ways). The estimated model parameters are shown in Exhibit 6 for the fitted CTAR(1) models, together with the values of -2 times the corresponding log likelihood and the observed one-step mean squared errors. Model (a) was found by maximizing the Gaussian likelihood under model (4.3) with $n = 10$ and Model (b) was found using the piecewise linear approximation. The linear model is the maximum likelihood CAR(1) model for the data, found as described in Section 2.

Model (b) differs from the corresponding model of TY since the latter does not include drift parameters, $b^{(j)}$, $j = 1, 2$, and we have not used their boundary adjustment.

Although the non-linear models show a substantial increase in the Gaussian likelihood (as compared with the linear model), the reduction in mean squared error (the average of the squares of the observed one-step prediction errors) is quite small. There are still gains to be made, however, from the non-linear models. Unlike linear models, they have the property that forecasting mean squared errors *depend on the level of the process* [cf. Pemberton (1989) and Tong (1990, p. 349)]. For Model (a) in particular, the one-step mean squared error has a minimum value of 0.69. Exhibit 7 shows the one-step predictors and their mean squared errors as functions of x. It clearly indicates those values of x for which the model one-step mean squared error is small. Simulations with Model (a) are also found to give marginal distributions quite similar to the empirical marginal distribution of the data.

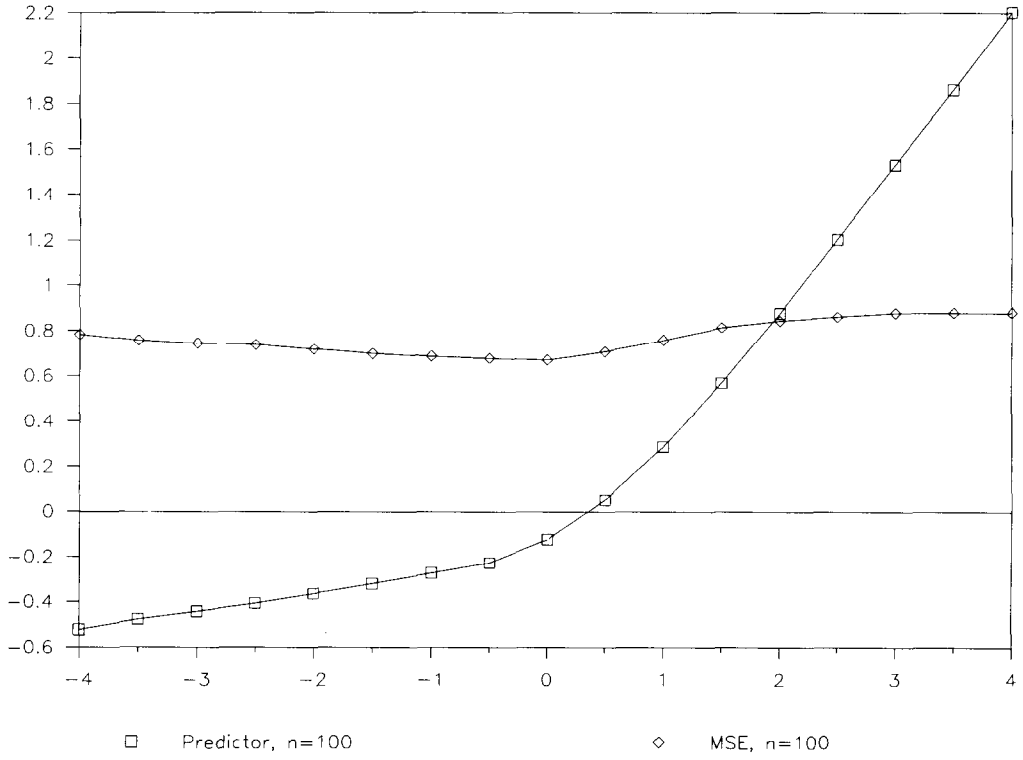

Exhibit 7. $m(x)$ and $v(x)$, IBM Model (a).

Notice from Exhibit 7 that the piecewise constant approximation to the mean squared error is not good over the range of observed values (it has the values 1.30 for $x < -0.56$ and 0.88 for $x > -0.56$). The piecewise constant approximation is also poor for the corresponding CTAR(1) process with the parameters of Model (b). The Gaussian likelihood and mean squared error for Model (b) are computed with the TY approximation as described above. These quantities are exact for the discrete-time Markov process with the appropriate Gaussian initial distribution and Gaussian one-step transition probabilities determined (for the time interval $[t, t+1]$) using one of two CAR(1) processes, chosen in accordance with the value of $X(t)$. Although this process is a very good model for the data (as indicated by the corresponding Gaussian likelihood and mean squared error), it differs substantially from a CTAR(1) process with narrow boundary and the same parameters, because of the deviation from piecewise constancy of the mean-squared error for the latter.

For fitting a CTAR(1) model we therefore prefer method (a) to method (b) in this example. As indicated previously we could, at the expense of computer time, improve on the calculation of the Gaussian likelihood in the fitting of Model (a) by choosing a value of n greater than 10.

Both models identify a threshold with negative value, confirming the conclusion of TY regarding this series. As they point out, this suggests that the dynamics of the closing stock prices change according as the stocks are falling at greater or less than a critical relative rate.

6. Inference for CTAR(p) processes with $p > 1$

When $p = 1$ it is a straightforward matter to compute the Gaussian likelihood of observations $x_1, \ldots, x_N$ of $X(t)$ at times $t_1, \ldots, t_N$ to a high degree of accuracy by choosing a suitably large value of the parameter n of the approximating process defined by (4.3). When $p > 1$ the problem is more difficult since the process $\{X(t)\}$ is no longer Markov.

One approach to this difficulty is to use a piecewise linear approximation, applying the linear Kalman recursions with model parameters changing according to the current value of $X(t)$ in order to compute the one-step predictor of $X(t+1)$ and its mean squared error. This is the method proposed by TY, of which we considered the special case with $p = 1$ in Section 5. Although this is a convenient and fast algorithm, it is difficult to specify the sense in which this approximates the Gaussian likelihood of a CTAR(p) process when $p > 1$.

Our approach will be to use a set of recursions which determine the *exact* likelihood of the observations $\{X(t_1), \ldots, X(t_N)\}$ in terms of the distribution of $\boldsymbol{X}(t_1)$ and the transition density of the Markov process $\{\boldsymbol{X}(t)\}$. The Gaussian likelihood is computed by replacing the transition density by a Gaussian density with the same moments of order up to 2. These moments are computed numerically with the aid of the approximating process defined by (3.3).

In order to obtain the required recursions, it is convenient to introduce a slightly different notation, writing the state vector $\boldsymbol{X}(t)$ as

$$\boldsymbol{X}(t) = \begin{bmatrix} X(t) \\ \boldsymbol{V}(t) \end{bmatrix},$$

where $\boldsymbol{V}(t)$ is the $(p-1) \times 1$ vector consisting of the first $p-1$ derivatives of $X(t)$.

Consider now the joint probability density, f_r, of $X(t_r)$, $\boldsymbol{V}(t_r)$, $X(t_{r-1})$, $X(t_{r-2})$, $\ldots$, $X(t_1)$. From the Markov property of $\{\boldsymbol{X}(t)\}$ it is easy to check that

$$f_{r+1}(x_{r+1}, \boldsymbol{v}_{r+1}, x_r, x_{r-1}, \ldots, x_1) = \int p(x_{r+1}, \boldsymbol{v}_{r+1}, t_{r+1} - t_r \mid x_r, \boldsymbol{v}_r) \times f_r(x_r, \boldsymbol{v}_r, x_{r-1}, \ldots, x_1)\, \mathrm{d}\boldsymbol{v}_r, \tag{6.1}$$

where $p(x_{r+1}, \boldsymbol{v}_{r+1}, t_{r+1} - t_r \mid x_r, \boldsymbol{v}_r)$ is the probability density of $(X(t_{r+1}), \boldsymbol{V}(t_{r+1})')'$, given $\boldsymbol{X}(t_r) =$

$(x_r, \boldsymbol{v}_r')'$. For a given set of observed values $x_1, \ldots, x_N$, at times $t_1, \ldots, t_N$, the functions $f_2, \ldots, f_N$ are functions of $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$, respectively. These functions can easily be computed recursively from (6.1) in terms of f_1 and the functions $p(x_{r+1}, \cdot, t_{r+1} - t_r \mid x_r, \cdot)$. The likelihood of the observations $x_1, \ldots, x_N$ is then clearly

$$L(\boldsymbol{\theta}; x_1, \ldots, x_N) = \int_{\boldsymbol{v}_N} f_N(\boldsymbol{v}_n)\, \mathrm{d}\boldsymbol{v}_n . \tag{6.2}$$

Notice that the filtered value of the unobserved vector $\boldsymbol{V}(t_r)$, $r = 1, \ldots, N$ [i.e. the conditional expectation of $\boldsymbol{V}(t_r)$ given $X(t_i) = x_i$, $i = 1, \ldots, r$] is readily obtained from the function f_r as

$$\tilde{\boldsymbol{v}}_r = \frac{\int \boldsymbol{v} f_r(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}}{\int f_r(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}} . \tag{6.3}$$

On the other hand, the calculation of the expected value of $X(t_{r+1})$ given $X(t_i) = x_i$, $i = 1, \ldots, r$, involves a much more complicated higher dimensional multiple integration. An alternative natural predictor of $X(t_{r+1})$ which is easy to compute can be found from

$$\tilde{x}_{r+1} = m((x_r, \tilde{\boldsymbol{v}}_r')', t_{r+1} - t_r) , \tag{6.4}$$

where $m((x_r, \tilde{\boldsymbol{v}}_r')', t_{r+1} - t_r)$ is approximated by $m_n((x_r, \tilde{\boldsymbol{v}}_r')', t_{r+1} - t_r)$ as defined in Section 3.

In all the numerical calculations given below we have taken $f_1(x, \boldsymbol{v})$ to be the Dirac delta function assigning mass one to $(x_1, \mathbf{0}')'$. This means that the likelihood in (6.2) is actually that of $x_2, \ldots, x_N$, *conditional* on $X(t_1) = x_1$ and $\boldsymbol{V}(t_1) = \mathbf{0}$. The first- and second-order moments of the transition density $p(x_{r+1}, \boldsymbol{v}_{r+1}, t_{r+1} - t_r \mid x_r, \boldsymbol{v}_r)$ are found using the approximating process defined by (3.2) and (3.3) (with $n = 10$). The 'Gaussian likelihood' is then found by replacing the transition densities by Gaussian densities with the same first- and second-order moments. The integrals in the recursions (6.1) are replaced by approximating Riemann sums.

There are two natural criteria for model-fitting based on the calculations just described:

(a) maximization of the Gaussian likelihood and

(b) minimization of the mean squared error of the predictors defined by (6.4).

Both criteria are used below to find models for the sunspot data (Example 1), the resulting models being referred to as Models (a) and (b), respectively. Model (c) is obtained by the method of TY.

Example 6. The parameter estimates for the CTAR(2) models fitted to the sunspot numbers (Example 1) by the three methods described above are shown in Exhibit 8. In each case we condition on $X(1) = x_1 (= 101)$ and $V(1) = 0$.

The optimization was constrained for each model so that the roots of $z^2 + a_1^{(i)} z + a_2^{(i)} = 0$, $i = 1,2$, all have negative real parts. The values of $-2\ln(L)$ and (observed) mean squared error for the linear model of Example 1 [computed conditionally on $X(1) = 101$ and $V(1) = 0$] are 812.3 and 216.1. Comparing with the corresponding values in Exhibit 8, we see that the non-linear models all result in a reduction in mean squared error, although for Model (a) the reduction is quite small.

As a check on the fitted models, we simulated 2000 observations of each of the CTAR(2) processes with the parameters of Models (a), (b) and (c) and compared the resulting marginal distributions with

Exhibit 8

Estimated parameter	r	$a_1^{(1)}$	$a_2^{(1)}$	$\sigma^{(1)}$	$b^{(1)}$	$a_1^{(2)}$	$a_2^{(2)}$	$\sigma^{(2)}$	$b^{(2)}$	$-2\ln(L)$	MSE
Model (a)	10.0	8.74	0.33	43.3	31.6	0.55	0.46	28.4	23.0	796.6	215.2
Model (b)	36.3	0.21	0.34	7.10	17.3	0.70	0.44	42.3	21.8	939.6	177.5
Model (c)	48.8	0.54	0.68	23.5	23.0	0.00	0.19	12.7	11.1	792.2	176.5

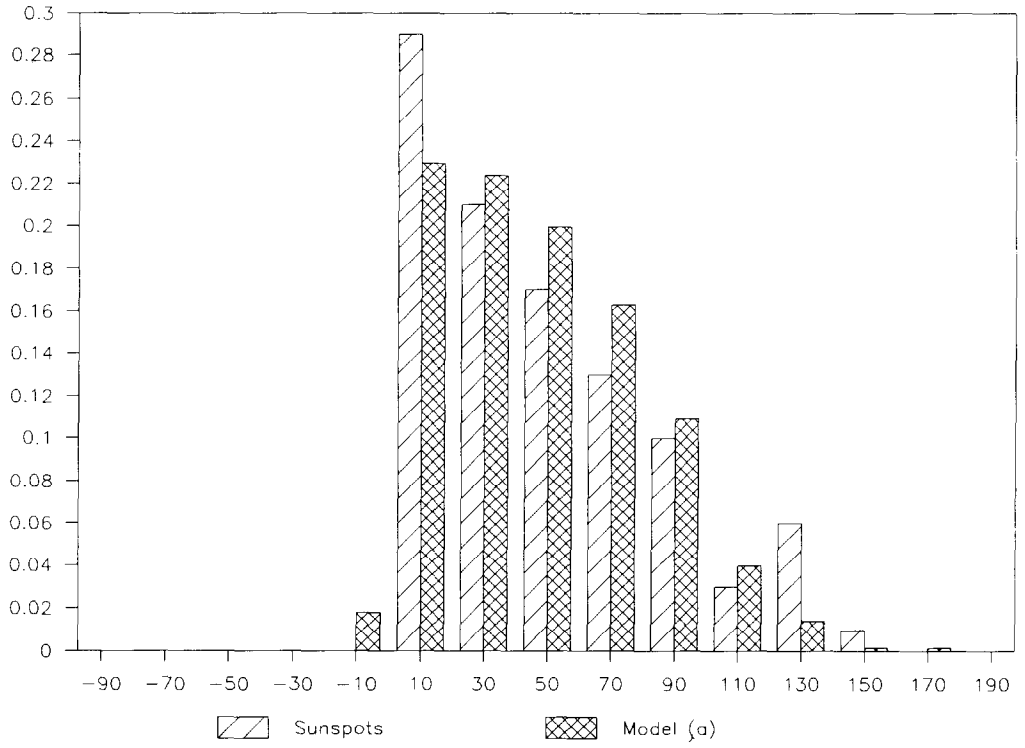

Exhibit 9. Histograms, Example 6.

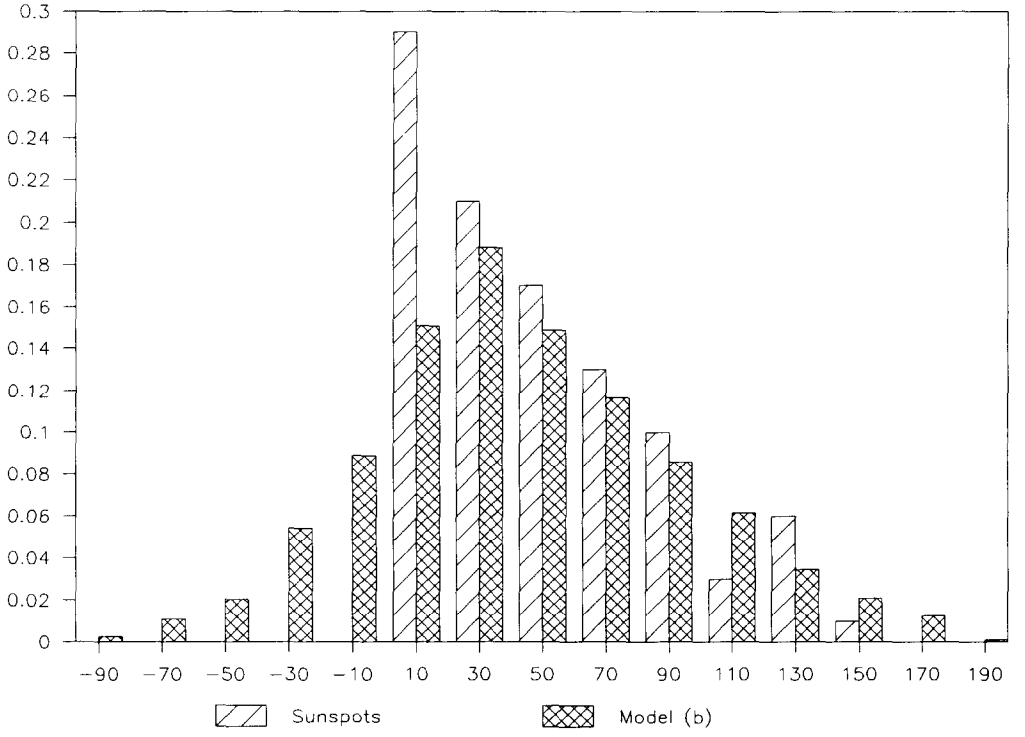

Exhibit 10. Histograms, Example 6.

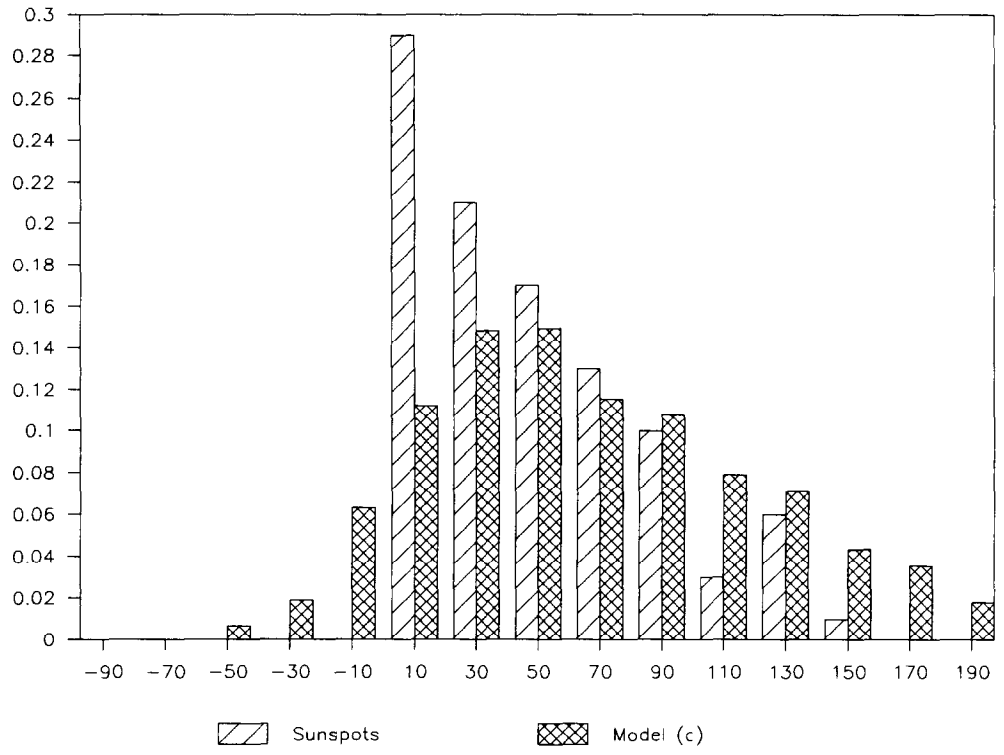

Exhibit 11. Histograms, Example 6.

the (decidedly non-Gaussian) marginal distribution of the sunspot numbers themselves. The results are shown in Exhibits 9, 10 and 11 and indicate a strong resemblance between the marginal distribution for Model (a) and the empirical distribution of the sunspot numbers.

The values of $-2\ln(GL)$ and MSE given in Exhibit 8 for Model (c) were computed using the innovations form of the Gaussian likelihood (5.1) [conditional on $X(1) = 101$ and $V(1) = 0$] with the one-step predictors $m(j)$ and mean squared errors $v(j)$ computed by applying the linear Kalman recursions with parameter values chosen according as $X(j-1)$ is greater than r or otherwise. Although this technique provides us with a fast, well-defined algorithm for computing one-step predictors and mean squared errors, the resulting likelihood does not have the same clearly-defined interpretation which it has when $p = 1$. In fact the CTAR *model* with these parameters is a poor fit to the data (in terms of Gaussian likelihood, one-step mean squared error and marginal distribution). There is a generalized non-linear regression model implicit in the TY algorithm however in which $EX(n)$ is a complicated n-dependent function of all the preceding observations. Although this model has no clear relation to a non-linear continuous time autoregression, it is a very good model for the sunspot numbers as can be seen from Exhibit 8.

We examined the forecasting performance of the three models by using them to compute one-step forecasts of the annual sunspot numbers for the years 1870–1889 [these can be found in Tong (1990, p. 470)]. The observed mean squared errors were

Model:	(*a*)	(*b*)	(*c*)	(*d*)
MSE :	450.9	419.3	391.5	469.8.

Model (d) is the linear model fitted in Example 1. Of the non-linear models, Model (c) gives the greatest reduction in mean squared error. Although the relation between the TY algorithm and continuous-time autoregression is only tenuous, its success here suggests that it is worthy of further study as a practical tool for non-linear prediction.

7. Conclusions

We have defined a CTAR(p) process with threshold r and boundary width $2\delta > 0$ as the first component of a p-dimensional diffusion process whose finite dimensional distributions are the limits of those of the approximating sequence of processes defined by (3.2) and (3.3). It is found numerically that the behaviour of the approximating sequence is essentially the same for small $\delta > 0$ as for $\delta = 0$. All numerical calculations with the approximating sequence were therefore carried out with $\delta = 0$. Two methods of inference for CTAR(1) processes were discussed and compared. Maximum Gaussian likelihood was found to perform well in the examples considered, both when one computes the exact Gaussian likelihood for the approximating process (4.3) with $n = 10$ and when one uses the piecewise linear approximation of TY. For the CTAR(p) process with $p \geq 2$, three methods were examined. One was the TY method which applies the linear Kalman recursions, with parameters chosen at time t in accordance with the location of $X(t)$ relative to thresholds. The other methods were based on a set of recursions which determine the exact likelihood of the observations from the initial distribution and transition densities of the p-dimensional state process. The 'Gaussian likelihood' is computed by replacing the latter by Gaussian densities whose first- and second-order moments are computed from the approximating process (3.3). The conditional expectation of the $(p-1)$-dimensional vector of derivatives $\boldsymbol{V}(t_r)$ given the observations $X(t_j) = x_j$, $j \leq r$, is also easily computed from the recursions and these can be used to compute the one-step predictors defined by (6.4). Parameter estimates can be computed either by maximizing the Gaussian likelihood or by minimizing the one-step prediction mean squared error. When these techniques were applied to the sunspot series it was found that the model obtained by maximization of the Gaussian likelihood as computed from the approximating process (3.3) had a marginal distribution quite similar to the observed marginal distribution of the data itself. Although the CTAR model with the TY parameters is not a good fit to the data, the TY algorithm was found to give very good one-step predictors for the sunspot numbers, suggesting that the non-linear model implicit in the algorithm is a good one for this data set. An important property of all the models fitted (and of non-linear models generally) is the non-constancy of the prediction mean squared error, which enables us to identify conditions under which more accurate forecasts can be expected.

References

Brockwell, P.J. and R.A. Davis, 1991, *Time Series: Theory and Methods, 2nd edn.* (Springer-Verlag, New York).

Brockwell, P.J., R.J. Hyndman and G.K. Grunwald, 1991, "Continuous time threshold autoregressive models", *Statistica Sinica*, 1, 401–410.

Chan, K.S., J.D. Petruccelli, H. Tong and S.W. Woolford, 1985, "A multiple threshold AR(1) model", *Journal of Applied Probability*, 22, 267–279.

Gikhman, I.I. and A.V. Skorokhod, 1969, *Introduction to the Theory of Random Processes* (W.B. Saunders, Philadelphia).

Jones, R.H., 1981, "Fitting a continuous time autoregression to discrete data", in: D.F. Findley, ed., *Applied Time Series Analysis II* (Academic Press, New York), 651–682.

Karlin, S. and H.M. Taylor, 1981, *A Second Course in Stochastic Processes* (Academic Press, New York).

Pemberton, R.H., 1989, "Forecasting accuracy of non-linear time series models", Tech. Report, Dept. of Mathematics, University of Salford.

Petruccelli, J.D., 1990, "A comparison of tests for SETAR-type non-linearity in time series", *Journal of Forecasting*, 9, 25–36.

Phillips, A.W., 1959, "The estimation of parameters in systems of stochastic differential equations", *Biometrika*, 46, 67–76.

Tong, H., 1983, *Threshold Models in Non-linear Time Series Analysis*, Springer Lecture Notes in Statistics, 21 (Springer-Verlag, New York).

Tong, H., 1990, *Non-linear Time Series: A Dynamical System Approach* (Clarendon Press, Oxford).

Tong, H. and I. Yeung, 1991, "Threshold autoregressive modelling in continuous time", *Statistica Sinica*, 1, 411–430.

Tsay, R.S., 1989, "Testing and modeling threshold autoregressive processes", *Journal of the American Statistical Association*, 84, 231–240.

Biographies: P.J. BROCKWELL is a Professor in the departments of Statistics and Electrical Engineering at Colorado State University. He has worked previously at Argonne National Laboratory, Michigan State University, La Trobe University, Kuwait University and The University of Melbourne. His research interests are in stochastic modelling and time series analysis.

R.J. HYNDMAN is employed by the Statistical Consulting Centre at The University of Melbourne and is studying for a Ph.D. in Statistics, working on time series analysis.