Ann. Inst. Statist. Math.
Vol. 47, No. 1, 1–20 (1995)

ON THE APPROXIMATION OF CONTINUOUS TIME THRESHOLD ARMA PROCESSES*

P. J. BROCKWELL** AND O. STRAMER***

Statistics Department, Colorado State University, Fort Collins, CO 80521, U.S.A.

(Received May 11, 1993; revised May 6, 1994)

Abstract. Threshold autoregressive (AR) and autoregressive moving average (ARMA) processes with continuous time parameter have been discussed in several recent papers by Brockwell *et al.* (1991, *Statist. Sinica*, **1**, 401–410), Tong and Yeung (1991, *Statist. Sinica*, **1**, 411–430), Brockwell and Hyndman (1992, *International Journal Forecasting*, **8**, 157–173) and Brockwell (1994, *J. Statist. Plann. Inference*, **39**, 291–304). A threshold ARMA process with boundary width $2\delta > 0$ is easy to define in terms of the unique strong solution of a stochastic differential equation whose coefficients are piecewise linear and Lipschitz. The positive boundary-width is a convenient mathematical device to smooth out the coefficient changes at the boundary and hence to ensure the existence and uniqueness of the strong solution of the stochastic differential equation from which the process is derived. In this paper we give a direct definition of a threshold ARMA processes with $\delta = 0$ in the important case when only the autoregressive coefficients change with the level of the process. (This of course includes all threshold AR processes with constant scale parameter.) The idea is to express the distributions of the process in terms of the weak solution of a certain stochastic differential equation. It is shown that the joint distributions of this solution with $\delta = 0$ are the weak limits as $\delta \downarrow 0$ of the distributions of the solution with $\delta > 0$. The sense in which the approximating sequence of processes used by Brockwell and Hyndman (1992, *International Journal Forecasting*, **8**, 157–173) converges to this weak solution is also investigated. Some numerical examples illustrate the value of the latter approximation in comparison with the more direct representation of the process obtained from the Cameron-Martin-Girsanov formula. It is used in particular to fit continuous-time threshold models to the sunspot and Canadian lynx series.

Key words and phrases: Non-linear time series, continuous-time ARMA process, threshold model, stochastic differential equation, approximation of diffusion processes.

---

* Research partially supported by National Science Foundation Research Grants DMS 9105745 and 9243648.

** Now at Mathematics Department, Royal Melbourne Institute of Technology, G.P.O. Box 2476V, Melbourne, Australia 3001.

*** Now at Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA 52242, U.S.A.

1. Introduction

The need for non-linear models in time series analysis has been appreciated for some time now. The classical Gaussian linear models are unable to reproduce some of the features frequently found in observed time series, e.g. time irreversibility, occasional bursts of outlying observations and deviations from Gaussian marginal distributions. In order to obtain more appropriate models for such data, various discrete-time non-linear models have been proposed. Among the more successful of these are the threshold models (Tong (1983, 1990)), bilinear models (Granger and Andersen (1978), Subba-Rao and Gabr (1984)) and random coefficient autoregressive models (Nicholls and Quinn (1982)). Many tests are now available for determining the appropriateness or otherwise of a linear model (see in particular Subba-Rao and Gabr (1984), Tsay (1989), Petruccelli (1990) and Tong (1990)).

Continuous-time (linear) ARMA models for time series have also been found valuable in applications, particularly in fitting models to irregularly spaced data (see Jones (1981), Jones and Ackerson (1990)). The use of continuous-time ARMA processes permits modelling with a larger class of available spectral densities (and corresponding autocovariance functions) than is attainable by continuous-time AR processes (see equations (2.8) and (2.7) below).

Several recent papers (Brockwell *et al.* (1991), Tong and Yeung (1991), Brockwell and Hyndman (1992) and Brockwell (1994)) have investigated the use of continuous-time threshold models, which are especially convenient for modelling irregularly spaced data when linear models are not adequate. Inference for such models depends of course on knowledge of the properties of such processes, particularly the finite dimensional distributions. In this paper our aim is to characterize the joint distributions of an important class of continuous-time threshold ARMA processes and to investigate the relationships between such processes and certain approximating sequences of processes which have been used in numerical studies of their behavior.

The processes considered are defined by stochastic differential equations whose coefficients depend on the value of the process as specified at the beginning of Section 3. They are not the most general continuous-time threshold ARMA processes since we only allow the autoregressive coefficients and the drift term to change with the value of the process. Even with these constraints however we obtain a rich class of non-linear models which appear to provide substantial improvement over the standard linear models for some well-studied data sets. In Section 6 we provide numerical illustrations of the approximations used and fit continuous-time ARMA models to the sunspot series and the logged Canadian lynx series.

2. Linear CARMA processes

We define a CARMA(p,q) process (with $0 \le q < p$) to be a stationary solution of the p-th order linear differential equation,

$$\begin{aligned}(2.1)\qquad & Y^{(p)}(t) + a_1 Y^{(p-1)}(t) + \cdots + a_p Y(t) \\ &= \sigma[W^{(1)}(t) + b_1 W^{(2)}(t) + \cdots + b_q W^{(q+1)}(t) + c],\end{aligned}$$

where the superscript ${}^{(j)}$ denotes j-fold differentiation with respect to t, $\{W(t)\}$ is standard Brownian motion and $a_1, \ldots, a_p$, $b_1, \ldots, b_q$, c and σ (the *scale parameter*) are constants. We assume that $\sigma > 0$ and $b_q \neq 0$ and define $b_j := 0$ for $j > q$. Since the derivatives $W^{(j)}(t)$, $j > 0$ do not exist in the usual sense, we interpret (2.1) as being equivalent to the *observation* and *state* equations,

$$Y(t) = \sigma \boldsymbol{b}' \boldsymbol{X}(t), \qquad t \geq 0, \tag{2.2}$$

and

$$d\boldsymbol{X}(t) = A\boldsymbol{X}(t)dt + \boldsymbol{e}(cdt + dW(t)), \tag{2.3}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{bmatrix}, \qquad \boldsymbol{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{bmatrix}$$

and (2.3) is an Ito 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,

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

of the continuous-time AR(p) process to which $\{Y(t)\}$ reduces when $b_j = 0$, $j \geq 1$.

The process $\{Y(t), t \geq 0\}$ is thus said to be a CARMA(p, q) process with parameters $(a_1, \ldots, a_p, b_1, \ldots, b_q, \sigma, c)$ if $Y(t) = \sigma[1 \;\; b_1 \;\; \cdots \;\; b_{p-1}]\boldsymbol{X}(t)$ where $\{\boldsymbol{X}(t)\}$ is a stationary solution of (2.3). The solution of (2.3) can be written as

$$\boldsymbol{X}(t) = e^{At}\boldsymbol{X}(0) + \int_0^t e^{A(t-u)}\boldsymbol{e}dW(u) + c\int_0^t e^{A(t-u)}\boldsymbol{e}du,$$

which is stationary if and only if

$$\boldsymbol{X}(0) \sim N\left(a_p^{-1}c[1 \;\; 0 \;\; \cdots \;\; 0]', \int_0^\infty e^{Ay}\boldsymbol{e}\boldsymbol{e}'e^{A'y}dy\right)$$

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. Under these conditions the autocovariance function of the process $\boldsymbol{X}(t)$ is easily found to be

$$E[\boldsymbol{X}^*(t+h)\boldsymbol{X}^*(t)'] = e^{Ah}\Sigma, \qquad h \geq 0,$$

where $\boldsymbol{X}^*(t)$ is the mean-corrected state vector

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

and

$$\Sigma := \int_0^\infty e^{Ay}\boldsymbol{e}\boldsymbol{e}'e^{A'y}dy.$$

The mean and autocovariance function of the CARMA(p,q) process $\{Y(t)\}$ are therefore given by

$$EY(t) = a_p^{-1}\sigma c, \tag{2.5}$$

and

$$\mathrm{Cov}(Y(t+h), Y(t)) = \sigma^2\boldsymbol{b}'e^{Ah}\Sigma\boldsymbol{b}. \tag{2.6}$$

Its spectral density is

$$f_Y(\omega) = \frac{\sigma^2}{2\pi}\frac{|1 + i\omega b_1 + \cdots + (i\omega)^q b_q|^2}{|(i\omega)^p + (i\omega)^{p-1}a_1 + \cdots + a_p|^2}, \qquad -\infty < \omega < \infty. \tag{2.7}$$

For the process to be *minimum phase* the roots of $1 + b_1 z + \cdots + b_q z^q = 0$ must have negative real parts. (This corresponds to *invertibility* for discrete time ARMA processes.)

Notice that the mean-corrected states and observations $\boldsymbol{X}^*(t_i)$ and $Y^*(t_i) = Y(t_i) - a_p^{-1}\sigma c$, at times $t_1, t_2, \ldots$, satisfy the discrete-time state and observation equations,

$$\begin{aligned} \boldsymbol{X}^*(t_{i+1}) &= e^{A(t_{i+1}-t_i)}\boldsymbol{X}^*(t_i) + \boldsymbol{Z}(t_i), \qquad i = 1, 2, \ldots, \\ Y^*(t_i) &= \sigma[1 \;\; b_1 \;\; \cdots \;\; b_p]\boldsymbol{X}^*(t_i), \qquad i = 1, 2, \ldots, \end{aligned}$$

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

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

These equations are in precisely the form needed for application of the Kalman recursions (see e.g. Brockwell and Davis (1991), Chapter 12). From these recursions we can easily compute $m_i = E(Y(t_i) \mid Y(t_j), j < i)$, and $v_i = E((Y(t_i) - m_i)^2 \mid Y(t_j), j < i)$, $i \geq 2$, and hence the likelihood,

$$L = (2\pi)^{-N/2}(v_1 \cdots v_N)^{-1/2} \exp\left[-\sum_{i=1}^{N}(Y(t_i) - m_i)^2/(2v_i)\right], \tag{2.8}$$

where $m_1 = a_p^{-1}\sigma c$ and $v_1 = \sigma^2 \boldsymbol{b}'\Sigma\boldsymbol{b}$.

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 calculations of e^{At} and the integrals are 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 used by Jones (1981) for maximum Gaussian likelihood fitting of CAR processes with possibly irregularly spaced data and by Jones and Ackerson (1990) in their analysis of longitudinal data using CARMA processes.

3. Threshold ARMA models with $\boldsymbol{b}$ and σ constant

To develop a continuous time analogue of the discrete-time self-exciting autoregressive moving average (SETARMA) process of Tong (1983) we would like to allow the parameters $a_1, \ldots, a_p$, $b_1, \ldots, b_q$, c and σ, of the process $\{Y(t)\}$ in the defining equations (2.2) and (2.3) to depend on $Y(t)$, taking fixed values in each of the l regions $r_{i-1} \le Y_t < r_i$, $i = 1, \ldots, l$ where $-\infty = r_0 < r_1 < \cdots < r_l = \infty$. If, instead of allowing the discontinuities in parameter values implicit in this definition, we 'smooth' the transitions at the boundaries by interpolating the parameter values linearly in each of the regions $r_i - \delta \le Y(t) \le r_i + \delta$, $i = 1, \ldots, l-1$, where $\delta > 0$ and δ is small compared with $\min(r_i - r_{i-1})$, then the coefficients of the resulting stochastic differential equation (2.3) satisfy conditions sufficient to guarantee the existence of a unique strong solution of the Ito differential equation (2.3).

In this paper we shall assume that only the autoregressive coefficients $a_1, \ldots,$ a_p and c change with the value of $Y(t)$. This restriction is satisfied of course by all threshold *autoregressive* models with constant σ. We shall also restrict attention to the case of a single threshold, since the extension to more than one is quite straightforward.

Thus we define the CTARMA(p, q) process with threshold at r, boundary-width $2\delta > 0$ and constant $b_1, \ldots, b_q$, σ exactly as in (2.2) and (2.3), except that we allow the parameters $a_1, \ldots, a_p$ and c to depend on $Y(t) = \sigma\boldsymbol{b}'\boldsymbol{X}(t)$ in such a way that

$$a_i(Y(t)) = a_i^{(J)}, \quad i = 1, \ldots, p; \qquad c(Y(t)) = c^{(J)}, \tag{3.1}$$

where $J = 1$ or 2 according as $Y(t) \le r - \delta$ or $Y(t) > r + \delta$. For $r - \delta < Y(t) \le r + \delta$, we define the coefficients $a_i(Y(t))$ and $c(Y(t))$ by linear interpolation. Thus $Y(t)$ is a linear combination of the components of a p-dimensional diffusion process $\{\boldsymbol{X}(t)\}$ whose drift and diffusion parameters are determined by (3.1). If $\boldsymbol{X}(0) = \boldsymbol{x}$, then $\{\boldsymbol{X}(t)\}$ is the unique strong solution of (2.3) subject to this initial condition. (We shall not require $\{\boldsymbol{X}(t)\}$ to be stationary as in Section 2.)

If we allow δ to be zero, a difficulty arises from the breakdown of the Lipschitz condition needed in the standard construction of the strong solution of the Ito equation (2.3). In order to specify uniquely what we mean by a continuous time threshold ARMA (CTARMA) process in this case, one possibility is to consider the well-defined strong solution of the Ito equation (2.3) with $\delta > 0$, i.e. with parameter values as described above, and consider its behaviour as $\delta \downarrow 0$. Under our assumption that the only parameters depending on $Y(t)$ are $a_1, \ldots, a_p$ and c, we shall see (in Section 5) that these processes converge weakly as $\delta \downarrow 0$ to the *weak* solution of the corresponding Ito equation with $\delta = 0$. The remainder of this section is devoted to the characterization (via the Cameron-Martin-Girsanov formula) of the unique (in law) weak solution of the stochastic differential equation (2.3) with arbitrary $\delta \geq 0$ and $\boldsymbol{X}(0) = \boldsymbol{x}$. We shall in fact obtain an explicit representation of its joint distributions in terms of a functional of standard Brownian motion.

(An alternative way to approach the definition of a continuous time analogue of the SETARMA process with $\delta = 0$ is to specify the state vector $\boldsymbol{X}(t)$ as a p-variate diffusion process with appropriately defined boundary behaviour. In general there is no unique way in which to do this (see for example Brockwell and Hyndman (1992) and Brockwell *et al.* (1991) where two different such definitions of a CTAR(1) process are given). We shall reserve the term *modified* CTAR(1) for the process defined in the latter paper.)

Assume now that the coefficients $a_1(Y(t)), \ldots, a_p(Y(t))$ and $c(Y(t))$ are defined as in (3.1) with $\delta \geq 0$, while $\boldsymbol{b}$ and σ are both constant. In this case the CTARMA process $\{Y(t)\}$ can be written (cf. (2.2) and (2.3)) as

$$Y(t) = \sigma \boldsymbol{b}' \boldsymbol{X}(t), \tag{3.2}$$

where the components $X_1, X_2, \ldots, X_p$ of the state vector $\boldsymbol{X}(t)$ satisfy

$$\begin{aligned}
dX_1 &= X_2(t)dt, \\
dX_2 &= X_3(t)dt, \\
&\vdots \\
dX_{p-1} &= X_p(t)dt, \\
dX_p &= [-a_p(Y(t))X_1(t) - \cdots - a_1(Y(t))X_p(t) + c(Y(t))]dt + dW(t).
\end{aligned} \tag{3.3}$$

Our aim in this section is to show, with $\boldsymbol{X}(0) = \boldsymbol{x} = [x_1 \ x_2 \ \cdots \ x_p]'$, that (3.3) has a unique (in law) weak solution $\{\boldsymbol{X}(t)\}$ and to determine the distribution of $\boldsymbol{X}(t)$ for any given $\boldsymbol{X}(0) = \boldsymbol{x}$. These distributions determine in particular the joint distribution of the values of the process $\{Y(t)\}$ at times $t_1, \ldots, t_N$ given $\boldsymbol{X}(0)$.

Assuming that $\boldsymbol{X}(0) = \boldsymbol{x}$, we can write $\boldsymbol{X}(t)$ in terms of $\{X_p(s), 0 \leq s \leq t\}$ using the relations, $X_{p-1}(t) = x_{p-1} + \int_0^t X_p(s)ds, \ldots, X_1(t) = x_1 + \int_0^t X_2(s)ds$. The resulting functional relationship will be denoted by

$$\boldsymbol{X}(t) = \boldsymbol{F}(X_p, t). \tag{3.4}$$

Substituting from (3.4) into the last equation in (3.3), we see that it can be written in the form,

$$dX_p = D(X_p, t)dt + dW(t), \tag{3.5}$$

where $D(X_p, t)$, like $\boldsymbol{F}(X_p, t)$, depends on $\{X_p(s), 0 \leq s \leq t\}$.

Now let B be standard Brownian motion (with $B(0) = x_p$) defined on the probability space $(C[0,\infty), \mathcal{B}[0,\infty), P_{x_p})$ and let $\mathcal{F}_t = \sigma\{B(s), s \leq t\} \vee \mathcal{N}$, where $\mathcal{N}$ is the sigma algebra of P_{x_p}-null sets of $\mathcal{B}[0,\infty)$. The equations

$$\begin{aligned} dZ_1 &= Z_2 dt, \\ dZ_2 &= Z_3 dt, \\ &\vdots \\ dZ_{p-1} &= Z_p dt, \\ dZ_p &= dB(t), \end{aligned} \tag{3.6}$$

with $\boldsymbol{Z}(0) = \boldsymbol{x} = [x_1 \ x_2 \ \cdots \ x_p]'$, clearly have the unique strong solution, $\boldsymbol{Z}(t) = \boldsymbol{F}(B,t)$, where $\boldsymbol{F}$ is defined as in (3.4). Let D be the functional appearing in (3.5) and suppose that $\hat{B}$ is the Ito integral defined by $\hat{B}(0) = x_p$ and

$$d\hat{B}(t) = -D(B,t)dt + dB(t) = -D(Z_p,t)dt + dZ_p(t). \tag{3.7}$$

If we define the new measure $\hat{P}_{x_p}$ on $(C[0,\infty), \mathcal{B}[0,\infty))$ satisfying

$$d\hat{P}_{x_p} = M(B,t)dP_{x_p}, \tag{3.8}$$

where

$$M(B,t) = \exp\left[-\frac{1}{2}\int_0^t D^2(B,s)ds + \int_0^t D(B,s)dB(s)\right], \tag{3.9}$$

then by the Cameron-Martin-Girsanov formula (see e.g. Oksendal (1989), p. 105), $\hat{B}$ is standard Brownian motion under $\hat{P}_{x_p}$. Hence from (3.7) we see that $(B, \hat{B})$ is a weak solution of (3.5) [on $(C[0,\infty), \mathcal{B}[0,\infty), \hat{P}_{x_p}, \{\mathcal{F}_t\})$] with initial condition $X_p(0) = x_p$. Hence $(\boldsymbol{Z}(t), \hat{B}(t))$ is a weak solution (on the same probability space) of the equations (3.3) with initial condition $\boldsymbol{X}(0) = \boldsymbol{x}$. Moreover, by Proposition 5.3.10 of Karatzas and Shreve (1991) and by theorem 10.2.2 of Stroock and Varadhan (1979) the weak solution is unique in law. The characteristic function $\phi(\boldsymbol{\theta} \mid \boldsymbol{x})$ of $\boldsymbol{X}(t)$ given $\boldsymbol{X}(0) = \boldsymbol{x}$ is therefore given by

$$\begin{aligned} \phi_{\boldsymbol{X}(t)}(\boldsymbol{\theta} \mid \boldsymbol{x}) &= \hat{E}_{x_p}[\exp(i\boldsymbol{\theta}'\boldsymbol{Z}(t))] \\ &= E_{x_p}[\exp(i\boldsymbol{\theta}'\boldsymbol{Z}(t))M(B,t)] \\ &= E_{x_p}[\exp(i\boldsymbol{\theta}'\boldsymbol{F}(B,t))M(B,t)], \end{aligned} \tag{3.10}$$

where $\hat{E}_{x_p}$ and E_{x_p} denote expectation relative to $\hat{P}_{x_p}$ and P_{x_p} respectively, $\boldsymbol{F}$ is defined as in (3.4) and M as in (3.9). The importance of equation (3.10) is that it gives the conditional characteristic function of the state-vector $\boldsymbol{X}(t)$ (and in particular of the CTARMA process, $Y(t) = \sigma\boldsymbol{b}'\boldsymbol{X}(t)$) given $\boldsymbol{X}(0) = \boldsymbol{x}$, as an expectation of a functional of the standard Brownian motion $\{B(t)\}$ starting at x_p.

The preceding construction of the weak solution $\{\boldsymbol{X}(t)\}$ of (2.3) with coefficients as in (3.1) is valid both when $\delta > 0$ and when $\delta = 0$. If $\delta > 0$ then the solution coincides in law with the unique strong solution $\{\boldsymbol{X}^{(\delta)}(t)\}$. In either case equation (3.10) determines the transition function of the Markov process $\{\boldsymbol{X}(t)\}$. (The Markov property follows from Theorem 4.20 of Karatzas and Shreve (1991).) In the same way the moments of the transition function can be expressed as expected values of functionals of standard Brownian motion.

In Section 5 we shall show that $\boldsymbol{X}^{(\delta)}$ converges weakly as $\delta \downarrow 0$ to $\boldsymbol{X}^{(0)}$, the weak solution when $\delta = 0$. First however we consider a useful sequence of approximating processes, $\boldsymbol{X}_n^{(\delta)}$, which converge weakly to $\boldsymbol{X}^{(\delta)}$ as $n \to \infty$, for each fixed $\delta > 0$.

4. A simple approximating sequence of processes

For the numerical study of continuous time threshold AR models, Brockwell and Hyndman (1992) used an approximating sequence of processes $\{Y_n^{(\delta)}(t)\}$. The analogous approximating sequence for the threshold ARMA process defined in Section 3 is as follows:

$$Y_n^{(\delta)}(t) = \sigma[1 \ b_1 \ b_2 \ \cdots \ b_q]\boldsymbol{X}_n^{(\delta)}(t), \tag{4.1}$$

where

$$\begin{aligned}\boldsymbol{X}_n^{(\delta)}(t+n^{-1}) = [I + n^{-1}A(Y_n^{(\delta)}(t))]\boldsymbol{X}_n^{(\delta)}(t) + n^{-1}b(Y_n^{(\delta)}(t))\\ + n^{-1/2}\sigma(Y_n^{(\delta)}(t))Z(t),\end{aligned} \tag{4.2}$$

and $A(Y(t))$ is defined as in (2.3) with $a_1, \ldots, a_p$, and c dependent on $Y(t)$ as in (3.1). It is assumed also that the initial state $\boldsymbol{X}_n^{(\delta)}(0)$ has the same distribution as $\boldsymbol{X}^{(\delta)}(0)$ 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 $\{\boldsymbol{X}_n^{(\delta)}(t), t \geq 0\}$ (with sample paths linear between $t = j/n$ and $t = (j+1)/n$, $j = 0, 1, 2, \ldots$) converge to those of $\{\boldsymbol{X}^{(\delta)}(t)\}$ as $n \to \infty$.

The process $\{\boldsymbol{X}_n^{(\delta)}(t), t = 0, 1/n, 2/n, \ldots\}$ defined by (4.2) is clearly Markovian. The conditional expectations,

$$m_n(\boldsymbol{x}, t) = E(Y_n^{(\delta)}(t) \mid \boldsymbol{X}_n^{(\delta)}(0) = \boldsymbol{x}),$$

satisfy the backward Kolmogorov equations,

$$\begin{aligned}m_n(\boldsymbol{x}, t+n^{-1}) = \frac{1}{2}m_n(\boldsymbol{x} + n^{-1}(A(y)\boldsymbol{x} + c(y)\boldsymbol{e}) + n^{-1/2}\sigma(y)\boldsymbol{e}, t)\\ + \frac{1}{2}m_n(\boldsymbol{x} + n^{-1}(A(y)\boldsymbol{x} + c(y)\boldsymbol{e}) - n^{-1/2}\sigma(y)\boldsymbol{e}, t),\end{aligned} \tag{4.3}$$

with the initial condition,

$$m_n(\boldsymbol{x}, 0) = y, \tag{4.4}$$

where

$$y := \sigma \boldsymbol{b}' \boldsymbol{x}.$$

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

$$m_n^{(j)}(\boldsymbol{x}, t) = E([Y_n^{(\delta)}(t)]^j \mid \boldsymbol{X}_n^{(\delta)}(0) = \boldsymbol{x}),$$

satisfy the same equation (4.3), and the slightly modified initial condition (4.4) with the right-hand side replaced by y^j.

Let us denote by $\{\boldsymbol{X}_n^{(\delta)}(t)\}$ the solution of (4.2) with $\{Z(t)\}$ given and $\boldsymbol{X}(0) = \boldsymbol{x}$. Note that $\{\boldsymbol{X}_n^{(\delta)}(t)\}$ is well-defined both for $\delta > 0$ and $\delta = 0$.

PROPOSITION 4.1. *Let $A^{(1)}$ and $A^{(2)}$ denote the matrices obtained when the values $a_i^{(1)}$ and $a_i^{(2)}$ respectively are substituted into the matrix A of (2.3). Then provided n is large enough for $(I + n^{-1}A^{(1)})$ and $(I + n^{-1}A^{(2)})$ to be non-singular, the finite-dimensional distributions of $\{\boldsymbol{X}_n^{(\delta)}(t)\}$ converge weakly to those of $\{\boldsymbol{X}_n^{(0)}(t)\}$ as $\delta \downarrow 0$ for almost all initial states $\boldsymbol{x}$ (i.e. for all $\boldsymbol{x}$ outside a set of Lebesgue measure zero).*

PROOF. For all values of n sufficiently large, the recursions (4.2) with $\delta = 0$ define a mapping, $\boldsymbol{x} \to \boldsymbol{X}(t)$, of $\boldsymbol{R}^p$ onto $\boldsymbol{R}^p$ which maps sets of positive Lebesgue measure into sets of positive Lebesgue measure. The inverse image, N of $\{\boldsymbol{x} : \sigma \boldsymbol{b}' \boldsymbol{x} = r\}$ therefore has zero Lebesgue measure, and for each $\boldsymbol{x} \notin N$ and any finite set of times $t_1, \ldots, t_m$ there exists $\delta(\boldsymbol{x}, t_1, \ldots, t_m)$ such that if $\delta < \delta(\boldsymbol{x}, t_1, \ldots, t_m)$, then with probability one,

$$\boldsymbol{X}_n^{(\delta)}(t_i) = \boldsymbol{X}_n^{(0)}(t_i), \qquad i = 1, \ldots, m.$$

This completes the proof. □

In the following section, we shall show that the joint distributions of $\boldsymbol{X}^{(\delta)}$, the unique strong solution of the equations (2.3) with $\boldsymbol{X}(0) = \boldsymbol{x}$ and with $a_1, \ldots, a_p$ defined as in (3.1), converge weakly as $\delta \downarrow 0$, to those of $\boldsymbol{X}^{(0)}$, the unique (in law) weak solution of (2.3) constructed in Section 3. Moreover $\boldsymbol{X}^{(\delta)}$ converges weakly as $\delta \downarrow 0$ to $\boldsymbol{X}^{(0)}$.

5. Convergence as $\delta \downarrow 0$

In Section 3 we showed that the characteristic function of any weak solution $\boldsymbol{X}(t)$ of (2.3) with coefficients as defined in (3.1) and with $\boldsymbol{X}(0) = \boldsymbol{x}$ is given by (3.10). The dependence of (3.10) on δ was not made explicit, however it is clear from the definitions of the functionals M and $\boldsymbol{F}$ that M depends on δ while $\boldsymbol{F}$ does not. For clarity we shall write $M^{(\delta)}$ for M in this section. As was also pointed out in Section 3, the result (3.10) is valid for all $\delta \geq 0$.

The aim of this section is to examine the relation between the processes $\{\boldsymbol{X}^{(\delta)}(t)\}$ with $\delta > 0$ (which are defined as strong solutions) and the process $\{\boldsymbol{X}^{(0)}(t)\}$ (which is defined in law only). Before proving the main result of this section we need a preliminary lemma concerning the process $\{\boldsymbol{F}(B,t)\}$ which appears in the key representation (3.10).

LEMMA 5.1. *If $\{B(t)\}$ is standard Brownian motion, then the process $\{\boldsymbol{F}(B,t)\}$ defined as in* (3.4) (*with deterministic initial state $\boldsymbol{x}$*) *is a Gaussian diffusion process. The covariance matrix $V(t)$ of $\boldsymbol{F}(B,t)$* (*strictly positive definite for $t > 0$*) *can be written as*

$$V(t) = \int_0^t e^{Hy}\boldsymbol{e}\boldsymbol{e}'e^{H'y}dy,$$

where

$$H = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

If $0 < \epsilon < T$ and $\boldsymbol{b} \in \boldsymbol{R}^p$, then

$$\min_{\epsilon \le s \le T} \boldsymbol{b}'V(s)\boldsymbol{b} = \boldsymbol{b}'V(\epsilon)\boldsymbol{b}.$$

PROOF. Observe that $\boldsymbol{F}(B,t)$ satisfies the stochastic differential equation (2.3) with $c = a_1 = \cdots = a_p = 0$. Since H is the value of the coefficient matrix A when $a_1 = \cdots = a_p = 0$, it suffices, by Proposition 6.2 of Ichihara and Kunita (1974), to check that

$$\operatorname{rank}(\boldsymbol{e}\boldsymbol{e}', H\boldsymbol{e}\boldsymbol{e}', \dots, H^{p-1}\boldsymbol{e}\boldsymbol{e}') = p.$$

This elementary calculation completes the proof. □

We now turn to the main result of this section.

THEOREM 5.1. *In the notation of* (3.10), *as $\delta \downarrow 0$,*

$$E_{x_p}[\exp(i\boldsymbol{\theta}'\boldsymbol{F}(B,t))M^{(\delta)}(B,t)] \to E_{x_p}[\exp(i\boldsymbol{\theta}'\boldsymbol{F}(B,t))M^{(0)}(B,t)]. \tag{5.1}$$

PROOF. By Scheffé's theorem and Example 21.21 of Billingsley ((1986), p. 289), it suffices to show that $M^{(\delta)}(B,t)$ converges in probability to $M^{(0)}(B,t)$ as $\delta \downarrow 0$. To establish this we shall show (see (3.9)) that for each fixed t, $\int_0^t D^{(\delta)}(B,s)^2 ds$ and $\int_0^t D^{(\delta)}(B,s)dB(s)$ converge in probability as $\delta \to 0$ to the corresponding integrals with $\delta = 0$. (We use the notation $D^{(\delta)}$ to display the dependence on δ of the functional D introduced in (3.5)).

For the first of these we apply the Cauchy-Schwarz inequality twice to obtain the inequalities,

$$
\begin{aligned}
(5.2)\qquad & E_{x_p}\left[\int_0^t D^{(\delta)}(B,s)^2 ds - \int_0^t D^{(0)}(B,s)^2 ds\right]^2 \\
&\le tE_{x_p}\left[\int_0^t (D^{(\delta)}(B,s)^2 ds - \int_0^t D^{(0)}(B,s)^2)^2 ds\right] \\
&\le tE_{x_p}\Bigg[\left(\max_{0\le s\le t} |D^{(\delta)}(B,s)^2 - D^{(0)}(B,s)^2|^2\right) \\
&\qquad \cdot (L\{s \in [0,t] : |\sigma \boldsymbol{b}'\boldsymbol{F}(B,s) - r| < \delta\})\Bigg] \\
&\le t\left[E_{x_p}\left[\max_{0\le s\le t} |D^{(\delta)}(B,s)^2 - D^{(0)}(B,s)^2|^2\right]^2\right]^{1/2} \\
&\qquad \times [E_{x_p}(L\{s \in [0,t] : |\sigma \boldsymbol{b}'\boldsymbol{F}(B,s) - r| < \delta\})^2]^{1/2},
\end{aligned}
$$

where L denotes Lebesgue measure. From the definitions of $D^{(\delta)}$ and $\boldsymbol{F}$, the first factor on the right of the last inequality is clearly bounded above by an expression of the form

$$
t\left[E_{x_p}\left[\max_{0\le s\le t} K(1 + \|\boldsymbol{F}(B,s)\|^2)\right]^2\right]^{1/2},
$$

where K depends only on the AR coefficients $a_i^{(J)}$ in (3.1). Note also, from Karatzas and Shreve ((1991), p. 306), that

$$
E_{x_p}\left(\max_{0\le s\le t} \|\boldsymbol{F}(B,s)\|^{2m}\right) \le C(1 + \|\boldsymbol{x}\|^{2m})\exp(ct) \qquad \text{for} \quad m \ge 1,
$$

where C is a positive constant. Hence

$$
(5.3)\qquad t\left[E_{x_p}\left(\max_{0\le s\le t} |D^{(\delta)}(B,s)^2 - D^{(0)}(B,s)^2|^2\right)^2\right]^{1/2} \le Dt,
$$

where D is a positive constant. The final factor in the last inequality of (5.2) is bounded above by

$$
[tE_{x_p}(L\{s \in [0,t] : |\sigma \boldsymbol{b}'\boldsymbol{F}(B,s) - r| < \delta\})]^{1/2},
$$

and application of Fubini's theorem followed by Lemma 5.1 shows that for any $\epsilon > 0$,

$$
\begin{aligned}
E_{x_p}(L\{s \in [0,t] : |\sigma \boldsymbol{b}'\boldsymbol{F}(B,s) - r| < \delta\}) &= \int_0^t P[|\sigma \boldsymbol{b}'\boldsymbol{F}(B,s) - r| < \delta] ds \\
&\le \epsilon + \int_\epsilon^T 2\delta[2\pi\sigma^2 \boldsymbol{b}'V_\epsilon \boldsymbol{b}]^{-1} ds.
\end{aligned}
$$

Since the last line converges to ϵ as $\delta \downarrow 0$ and since ϵ is arbitrary, we conclude that

$$\int_0^t P[|\sigma \boldsymbol{b}'\boldsymbol{F}(B,s) - r| < \delta]ds \to 0 \quad \text{as} \quad \delta \to 0.$$

From (5.2) and (5.3), we conclude that for each fixed t,

$$E_{x_p}\left[\left(\int_0^t D^{(\delta)}(B,s)^2 ds - \int_0^t D^{(0)}(B,s)^2 ds\right)\right]^2 \to 0 \quad \text{as} \quad \delta \to 0.$$

Finally we observe, from the fundamental property of the Ito integral, that

$$\begin{aligned}(5.4) \qquad & E_{x_p}\left[\int_0^t (D^{(\delta)}(B,s) - D^{(0)}(B,s))dB(s)\right]^2 \\ &= E_{x_p}\left[\int_0^t |D^{(\delta)}(B,s) - D^{(0)}(B,s)|^2 ds\right].\end{aligned}$$

The proof that the right-hand side of (5.4) converges to zero as $\delta \to 0$ uses inequalities exactly analogous to those in (5.2) with D replacing D^2 throughout. This completes the proof that for each fixed t, $\int_0^t D^{(\delta)}(B,s)^2 ds$ and $\int_0^t D^{(\delta)}(B,s)dB(s)$ converge in probability as $\delta \to 0$ to the corresponding integrals with $\delta = 0$, and hence the result (5.1). □

Remark 5.1. The argument which gave us the representation (3.10) immediately provides a similar expression for the finite-dimensional distributions of $\boldsymbol{X}^{(\delta)}$. The weak convergence of these as $\delta \downarrow 0$ is proved in the same way as Theorem 5.1. The following still stronger result is also true. (It can also be proved with the aid of Theorem 2.4.2 of Kushner (1984).)

THEOREM 5.2. *$\boldsymbol{X}^{(\delta)}$ converges weakly as $\delta \downarrow 0$ to $\boldsymbol{X}^{(0)}$.*

PROOF. Since the finite-dimensional distributions of the processes $\boldsymbol{X}^{(\delta)}$ converge as $\delta \downarrow 0$ to the finite-dimensional distributions of the process $\boldsymbol{X}^{(0)}$, it is enough to prove that $\boldsymbol{X}^{(\delta)}$ is tight. This follows directly from problems 2.4.11 and 5.3.15 of Karatzas and Shreve (1991). □

6. Examples

The representation (3.10) determines the transition function of the state process $\boldsymbol{X}(t)$ both for $\delta > 0$ and $\delta = 0$. In the same way the moments of the transition function can be expressed as expected values of functionals of standard Brownian motion. For example, the conditional mean of $\boldsymbol{X}(t)$ given $\boldsymbol{X}(0) = \boldsymbol{x}$ is

$$(6.1) \qquad E[(\boldsymbol{X}(t)) \mid \boldsymbol{X}(0) = \boldsymbol{x}] = E_{x_p}[(\boldsymbol{F}(B,t))M(B,t)], \qquad r = 1, 2, \ldots.$$

Analytical evaluation of the right hand-side appears to be extremely difficult. Even for the CTAR(1) process with σ constant and (see (3.1)) $a_1^{(1)} = a_1^{(2)} = \delta = 0$, i.e.

for Brownian motion with two-valued state-dependent drift, the results of Karatzas and Shreve (1991), Proposition 6.5.1, demonstrate the complexity of the transition probabilities.

Equation (6.1) does however suggest the possibility of simulating B and estimating the expectation on the right by averaging the observed values of the functional over a large number of independent realizations. In the following example we compare this approach with the approximation used by Brockwell and Hyndman (1992). For the latter we solved the equations (4.3) and (4.4) with $n = 10$ and for the former we carried out the simulation with $r = 9000$ replicates.

Example 6.1. Consider the CTAR(1) process $Y(t)$ defined (cf. (3.2) and (3.3)) by,

$$Y(t) = \sigma X(t),$$

where

$$dX(t) = -a(X(t))X(t)dt + dW(t)$$

and

$$a(y) = \begin{cases} a_1 & \text{if } y < 0 \\ a_2 & \text{if } y \geq 0. \end{cases}$$

In this case we have $F(X(t)) = X(t)$ and $D(X,t) = -a(X(t))$. The conditional expectation $m(x,t)$ of $X(t)$ given $X(0) = x$ is

$$m(x,t) = E_x \left[B(t) \exp \left(-\frac{1}{2} \int_0^t a^2(B(s))B^2(s)ds + \int_0^t a(B(s))B(s)dB(s) \right) \right]. \tag{6.2}$$

For the case $a_1 = 0.5$ and $a_2 = 1$, the function $m(x,1) = E[X(1) \mid X(0) = x] = E_x[F(B,1)M(B,1)]$ was estimated using the sample mean of 9000 simulated realizations of $F(B,1)M(B,1)$. In Fig. 1 the resulting estimates, with x ranging from -5 to $+5$, are shown together with the approximation of Brockwell and Hyndman computed from (4.3) and (4.4) with $n = 10$. The agreement between the two methods is extremely good for initial levels x close to the threshold. However as x moves further away from the threshold, the increasing conditional variance of $F(B,1)M(B,1)$ (reflected by the sample variances in Table 1) causes the accuracy of the simulation method to deteriorate.

Example 6.2. In this example we shall examine in more detail the form taken by the conditional expectation of a CTAR(2) process conditional on the initial value of the state vector. Let $Y(t)$ be the process defined (cf. (3.2) and (3.3)) by,

$$Y(t) = \sigma X_1(t)$$

where the components X_1, X_2, of the state vector $\boldsymbol{X}(t)$ satisfy

$$dX_1 = X_2(t)dt,$$
$$dX_2 = [-a_2(Y(t))X_1(t) - a_1(Y(t))X_2(t) + c(Y(t))]dt + dW(t).$$

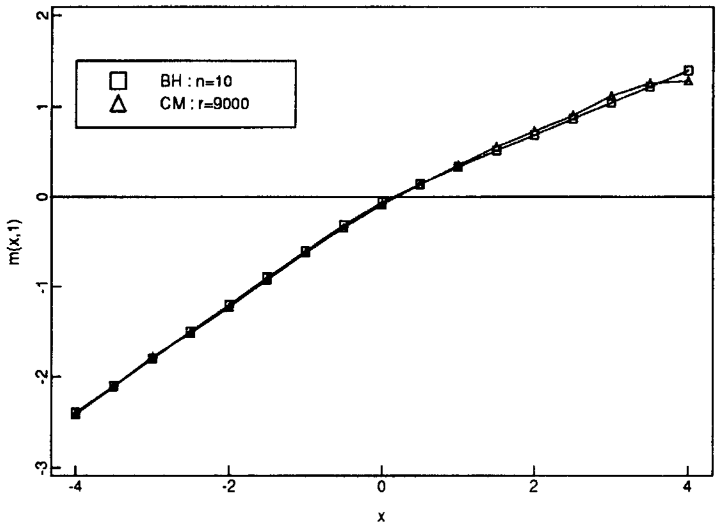

Fig. 1. The one-step predictors for Example 6.1.

Table 1. The sample variance of 9000 simulated values of $F(B,1)M(B,1)$ in Example 6.

Initial level	Variance of $F(B,1)M(B,1)$
−4.0	16.5
−3.5	7.24
−3.0	2.99
−2.5	1.26
−2.0	0.54
−1.5	0.36
−1.0	0.36
−0.5	0.40
0.0	0.41
0.5	0.41
1.0	0.43
1.5	0.59
2.0	1.15
2.5	2.86
3.0	8.32
3.5	24.5
4.0	59.9

Assume now that $\boldsymbol{X}(0) = \boldsymbol{x} = [x_1, x_2]'$. In this case we have

$$\boldsymbol{F}(B,t) = \left[x_1 + \int_0^t B(s)ds, B(t)\right]',$$

$$D(B,t) = -a_2(Y(t))\left(x_1 + \int_0^t B(s)ds\right) - a_1(Y(t))B(t) + c(Y(t)),$$
$$M(B,t) = \exp\left[-\frac{1}{2}\int_0^t D^2(B,s)ds + \int_0^t D(B,s)dB(s)\right],$$

where

$$a_1(y) = \begin{cases} a_1^1 & \text{if } y < 0, \\ a_1^2 & \text{if } y \geq 0, \end{cases}$$
$$a_2(y) = \begin{cases} a_2^1 & \text{if } y < 0, \\ a_2^2 & \text{if } y \geq 0, \end{cases}$$
$$c(y) = \begin{cases} c_1 & \text{if } y < 0, \\ c_2 & \text{if } y \geq 0. \end{cases}$$

The conditional mean of $\boldsymbol{X}(t)$ given $\boldsymbol{X}(0) = \boldsymbol{x}$ is

$$E[\boldsymbol{X}(t) \mid \boldsymbol{X}(0) = \boldsymbol{x}] = E_{x_2}[\boldsymbol{F}(B,t)M(B,t)], \tag{6.3}$$

and the conditional mean $m(\boldsymbol{x},t)$ of $Y(t)$ given $\boldsymbol{X}(0) = \boldsymbol{x}$ is therefore

$$m(\boldsymbol{x},t) = \sigma\ E_{x_2}[F_1(B,t)M(B,t)] = \sigma E_{x_2}\left[\left(x_1 + \int_0^t B(s)ds\right)M(B,t)\right]. \tag{6.4}$$

For the CTAR(2) process of this Example with $a_1^1 = 1$, $a_1^2 = 0.5$, $a_2^1 = 0.5$, $a_2^2 = 1$, $c_1 = 0.5$, $c_2 = 1$, $\sigma = 1$ and $x_2 = 0$, numerical results similar to those in Example 6.1 were obtained. They are shown below in Fig. 2 and Table 2.

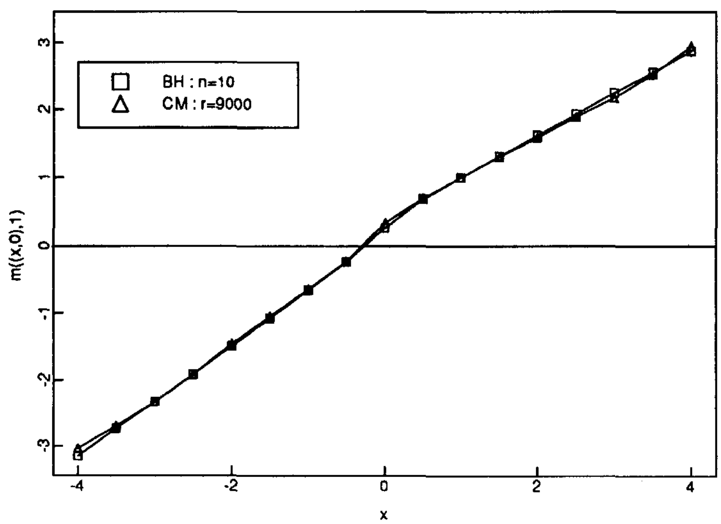

Fig. 2. The one-step predictors for Example 6.2.

Table 2. The sample variance of 9000 simulated values of $F_1(B,1)M(B,1)$ in Example 6.2.

Initial level	Variance of $F_1(B,1)M(B,1)$
−4.0	38.43
−3.5	20.18
−3.0	9.98
−2.5	4.67
−2.0	1.89
−1.5	0.71
−1.0	0.25
−0.5	0.16
0.0	0.56
0.5	0.51
1.0	0.37
1.5	0.46
2.0	1.15
2.5	3.77
3.0	11.70
3.5	38.18
4.0	130.27

Example 6.3. Brockwell and Hyndman (1992) fitted a continuous-time threshold AR(2) model to the annual sunspot numbers, 1770–1869, by maximizing the Gaussian likelihood conditional on the initial state $\boldsymbol{X}(0) = (101, 0)'$. The conditional Gaussian likelihood (GL) was computed by assuming the transition probabilities of the state vector for a time-interval of one year to be Gaussian with transition mean and covariance matrices equal to those of the approximating process defined in Section 5 with $n = 10$. The fitted model has a boundary at $r = 10.0$ and a value of $-2\ln(\mathrm{GL})$ equal to 796.6, as compared with the maximum-likelihood linear CAR(2) model,

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

for which $-2\ln(\mathrm{GL}) = 813.1$. If the maximization is constrained to the class of CTAR(2) models with constant σ, we obtain the fitted model,

$$\begin{aligned} X^{(2)}(t) + 7.39X^{(1)}(t) + 0.43X(t) &= 33.8W^{(1)}(t) + 30.3, \qquad X(t) < 10,\\ X^{(2)}(t) + 0.75X^{(1)}(t) + 0.46X(t) &= 33.8W^{(1)}(t) + 21.3, \qquad X(t) > 10, \end{aligned}$$

for which $-2\ln(\mathrm{GL}) = 800.9$. In order to check the use of the approximating process with $n = 10$ in computing GL for this model, a comparison was made of the mean and variance of $X(1)$ given $\boldsymbol{X}(0) = (x, 0)'$ for values of x between 0 and 100. Figures 3 and 4 show the comparisons between the approximations with $n = 10$ and corresponding values obtained by simulation of the approximating

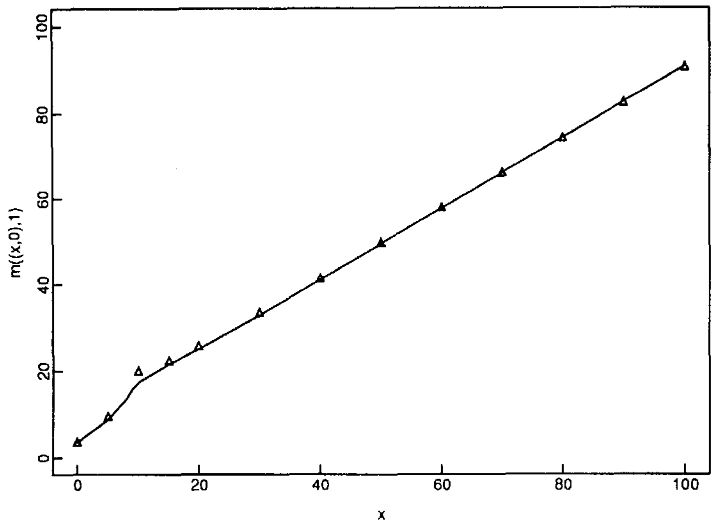

Fig. 3. The mean of $X(1)$ conditional on $\boldsymbol{X}(0) = (x, 0)'$, $0 \leq x \leq 100$, for Example 6.3. The continuous curve is based on the approximating process of Section 4 with $n = 10$ and the points marked by triangles are obtained by simulation with $n = 100$ and Gaussian noise.

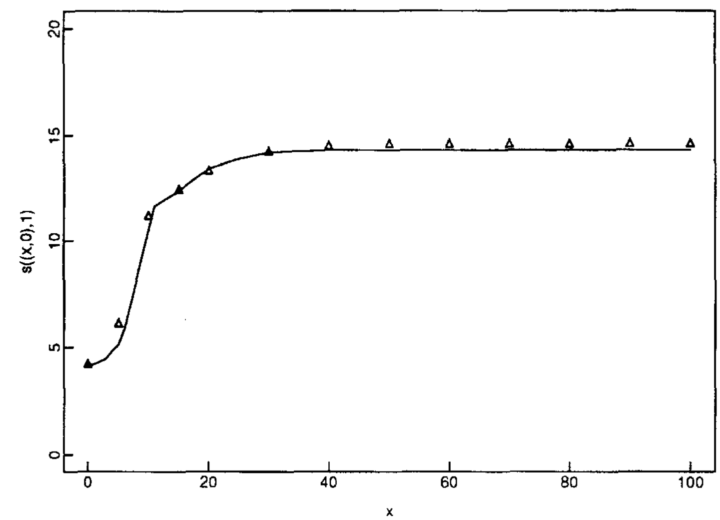

Fig. 4. The standard deviation of $X(1)$ conditional on $\boldsymbol{X}(0) = (x, 0)'$, $0 \leq x \leq 100$, for Example 6.3. The continuous curve is based on the approximating process of Section 4 with $n = 10$ and the points marked by triangles are obtained by simulation with $n = 100$ and Gaussian noise.

process with $n = 100$ and with Gaussian instead of binary noise process $Z(t)$. Each of the plotted points for the simulation is based on independent samples of size 100,000. Such a simulation is much slower than direct computation using the approximating process with $n = 10$, but the graphs indicate that both methods give very similar results.

Example 6.4. To illustrate the use of the non-linear threshold ARMA process with one threshold we consider the logged (to base 10) Canadian lynx series, 1821–1920. By maximizing GL (the Gaussian likelihood conditional on the initial state vector) as described in Brockwell (1994), the following CTARMA(2, 1) model was fitted to the data.

$$\begin{aligned}
&X^{(2)}(t) + 0.458X^{(1)}(t) + 0.349X(t) \\
&\qquad = 0.207[W^{(1)}(t) - 0.688W^{(2)}(t) + 5.348], \qquad X(t) < 3.242, \\
&X^{(2)}(t) + 0.003X^{(1)}(t) + 0.407X(t) \\
&\qquad = 0.223[W^{(1)}(t) - 0.388W^{(2)}(t) + 5.206], \qquad X(t) > 3.242,
\end{aligned}$$

with $-2\ln(\mathrm{GL}) = -26.37$. The observed root mean square error of the one-step predictors of the 14 values for the years 1921–1934 is 0.104, comparing favourably with the values 0.120, 0.116 and 0.115 obtained using the threshold model of Tong (1983), the random coefficient autoregressive model of Nicholls and Quinn (1982) and the bilinear model of Subba-Rao and Gabr (1984) respectively.

If we constrain the maximization to the class of CTARMA(2, 1) models with constant values of the moving average coefficient and the parameter σ, we obtain the model,

$$\begin{aligned}
&X^{(2)}(t) + 0.553X^{(1)}(t) + 0.344X(t) \\
&\qquad = 0.195[W^{(1)}(t) - 0.575W^{(2)}(t) + 5.574], \qquad X(t) < 3.092, \\
&X^{(2)}(t) + 0.008X^{(1)}(t) + 0.402X(t) \\
&\qquad = 0.195[W^{(1)}(t) - 0.575W^{(2)}(t) + 5.877], \qquad X(t) > 3.092,
\end{aligned}$$

for which the value of $-2\ln(\mathrm{GL})$ is -25.40 and the observed root mean square one-step error for the years 1921–1934 is 0.105. Thus the constrained model achieves a value of $-2\ln(\mathrm{GL})$ close to that of the unconstrained model and gives a root mean square error for the years 1921–1934 which, although not quite as good as the unconstrained model, is still better than the competing non-linear models referred to in the preceding paragraph.

7. Conclusions

We have used the Cameron-Martin-Girsanov formula to investigate the relation between the joint distributions of the well-defined strong solutions of (2.3) with coefficients as in (3.1) and $\delta > 0$ and those of the weak solution, which is defined also when $\delta = 0$. The representation (3.10) also suggests an obvious simulation technique for computing moments of the transition distribution of the solution both for $\delta > 0$ and $\delta = 0$. Numerical examples suggest good agreement between the results from this technique and those from the approximation of Brockwell and Hyndman (1992), provided the initial state $\boldsymbol{X}(0)$ is close to the threshold. However the performance of the Cameron-Martin-Girsanov simulation method deteriorates as $X(0)$ moves away from the threshold since the functionals of Brownian motion required in the simulation have large variance. The approximation technique of Brockwell and Hyndman (1992) was used to fit non-linear continuous time threshold models to the sunspot and logged Canadian lynx trappings. For a more detailed discussion of the model-fitting procedure see Brockwell and Hyndman (1992) and Brockwell (1994).

Acknowledgements

We wish to thank the National Science Foundation for its support of this work through Grants DMS 9105745 and 9243648 and two referees for valuable suggestions for improvement.

References

Billingsley, P. (1986). *Probability*, 2nd ed., Springer, New York.

Brockwell, P. J. (1994). On continuous time threshold ARMA processes, *J. Statist. Plann. Inference*, **39**, 291–304.

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

Brockwell, P. J. and Hyndman, R. J. (1992). On continuous time threshold autoregression, *International Journal Forecasting*, **8**, 157–173.

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

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

Granger, C. W. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*, Vanderhoek and Ruprecht, Göttingen.

Ichihara, K. and Kunita, H. (1974). A classification of the second order degenerate elliptic operators and its probabilistic characterization, *Zeit. Wahrscheinlichkeitsth.*, **30**, 235–254.

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

Jones, R. H. and Ackerson, L. M. (1990). Serial correlation in unequally spaced longitudinal data, *Biometrika*, **77**, 721–732.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, Springer, New York.

Kushner, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes*, The MIT Press, Cambridge, Massachusetts.

Nicholls, D. F. and Quinn, B. G. (1982). Random coefficient autoregressive models: an introduction, *Lecture Notes in Statist.*, **11**, Springer, New York.

Oksendal, B. (1989). *Stochastic Differential Equations*, 2nd ed., Springer, Berlin.

Petruccelli, J. D. (1990). A comparison of tests for SETAR-type non-linearity in time series, *International Journal of Forecasting*, **6**, 25–36.

Stroock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes*, Springer, Berlin.

Subba-Rao T. and Gabr, M. M. (1984). An introduction to bispectral analysis and bilinear time series models, *Lecture Notes in Statist.*, **24**, Springer, New York.

Tong, H. (1983). Threshold models in non-linear time series analysis, *Lecture Notes in Statist.*, **21**, Springer, New York.

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

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

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes, *J. Amer. Statist. Assoc.*, **84**, 231–240.