Continuous-time Gaussian Autoregression

Peter Brockwell, Richard Davis and Yu Yang

Statistics Department, Colorado State University

Abstract: The problem of fitting continuous-time autoregressions (linear and non-linear) to closely and regularly spaced data is considered. For the linear case Jones (1981) and Bergstrom (1985) used state-space representations to compute exact maximum likelihood estimators and Phillips (1959) did so by fitting an appropriate discrete-time ARMA process to the data. In this paper we use exact conditional maximum likelihood estimators for the continuously-observed process to derive approximate maximum likelihood estimators based on the closely-spaced discrete observations. We do this for both linear and non-linear autoregressions and indicate how the method can be modified also to deal with non-uniformly but closely-spaced data. Examples are given to indicate the accuracy of the procedure.

Key words and phrases: Continuous-time autoregression, threshold autoregression, maximum likelihood, sampled process, Wiener measure, Radon-Nikodym derivative, Cameron-Martin-Girsanov formula.

1. Introduction

This paper is concerned with estimation for continuous-time Gaussian autoregressions, both linear and non-linear, based on observations made at closely-spaced times. The idea is to use the exact conditional probability density of the $(p-1)^{st}$ derivative of an autoregression of order p with respect to Wiener measure in order to find exact conditional maximum likelihood estimators of the parameters under the assumption that the process is observed continuously. The resulting estimates are expressed in terms of stochastic integrals which are then approximated using the available discrete-time observations.

In Section 2 we define the continuous-time AR(p) (abbreviated to CAR(p)) process driven by Gaussian white noise and briefly indicate the relation between the CAR(p) process $\{Y(t), t \geq 0\}$ and the sampled process $\{Y_n^{(h)} := Y(nh), n = 0, 1, 2, \ldots\}$. The process $\{Y_n^{(h)}\}$ is a discrete-time ARMA process, a result employed by Phillips (1959) to obtain maximum likelihood estimates of the parameters of the continuous-time process based on observations of $\{Y_n^{(h)}, 0 \leq nh \leq T\}$. From the state-space representation of the CAR(p) process it is also possible to express the likelihood of observations of $\{Y_n^{(h)}\}$ directly in terms of the parameters of the CAR(p) process and thereby to compute maximum likelihood estimates of the parameters as in Jones (1981) and Bergstrom (1985). For a CAR(2) process we use the asymptotic distribution of the maximum likelihood estimators of the coefficients of the ARMA process $\{Y_n^{(h)}\}$ to derive the asymptotic distribution, as first $T \to \infty$ and then $h \to 0$, of the estimators of the coefficients of the underlying CAR process.

In Section 3 we derive the probability density with respect to Wiener measure of the $(p-1)^{st}$ derivative of the (not-necessarily linear) autoregression of order p. This forms the basis for the inference illustrated in Sections 4, 5 and 6. In the non-linear examples considered we restrict attention to continuous-time threshold autoregressive (CTAR) processes, which are continuous-time analogues of the discrete-time threshold models of Tong (1983).

In Section 4 we apply the results to (linear) CAR(p) processes, deriving explicit expressions for the maximum likelihood estimators of the coefficients and illustrating the performance of the approximations when the results are applied to a discretely observed CAR(2) process. In Section 5 we consider applications to CTAR(1) and CTAR(2) processes with known threshold and in Section 6 we show how the technique can be adapted to include estimation of the threshold itself. The technique is also applied to the analysis of the Canadian lynx trappings, 1821 - 1934.

2. The Gaussian CAR(p) and corresponding sampled processes

A continuous-time Gaussian autoregressive process of order $p > 0$ is defined symbolically to be a stationary solution of the stochastic differential equation,

$$a(D)Y(t) = bDW(t), \tag{2.1}$$

where $a(D) = D^p + a_1 D^{p-1} + \cdots + a_p$, the operator D denotes differentiation with respect to t and $\{W(t), t \geq 0\}$ is standard Brownian motion. Since $DW(t)$ does not exist as a random function, we give meaning to equation (2.1) by rewriting it in state-space form,

$$Y(t) = (b, 0, \ldots, 0)\mathbf{X}(t), \tag{2.2}$$

where the state vector $\mathbf{X}(t) = (X_0(t), \dots, X_{p-1}(t))^T$ satisfies the Itô equation,

$$dX(t) = A\mathbf{X}(t)dt + \mathbf{e}dW(t), \tag{2.3}$$

with

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{bmatrix} \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

From equation (2.3) we see that $X_j(t)$ is the j^{th} mean-square and pathwise derivative $D^j X_0(t), j = 0, \dots, p-1$. We are concerned in this paper with inference for the autoregressive coefficients, $a_1, \dots, a_p$, based on observations of the process Y at times $0, h, 2h, \dots, h[T/h]$, where h is small and $[x]$ denotes the integer part of x.

One approach to this problem, due to Phillips (1959), is to estimate the coefficients of the discrete-time ARMA process $\{Y_n^{(h)} := Y(nh), n = 0, 1, 2, \dots\}$ and from these estimates to obtain estimates of the coefficients $a_1, \dots, a_p$ in equation (2.1). The sampled process $\{Y_n^{(h)}\}$ is a stationary solution of the Gaussian ARMA(p', q') equations,

$$\phi(B)Y_n^{(h)} = \theta(B)Z_n, \quad \{Z_n\} \sim \mathrm{WN}(0, \sigma^2), \tag{2.4}$$

where $\phi(B)$ and $\theta(B)$ are polynomials in the backward shift operator B of orders p' and q' respectively, where $p' \leq p$ and $q' < p'$. (For more details see, e.g., Brockwell (1995).)

An alternative approach is to use equations (2.2) and (2.3) to express the likelihood of observations of $\{Y_n^{(h)}\}$ directly in terms of the parameters of the CAR(p) process and then to compute numerically the maximum likelihood estimates of the parameters as in Jones (1981) and Bergstrom (1985).

In this paper we take a different point of view by assuming initially that the process Y is observed continuously on the interval $[0, T]$. Under this assumption it is possible to calculate exact (conditional on $\mathbf{X}(0)$) maximum likelihood estimators of $a_1, \dots, a_p$. To deal with the fact that observations are made only at times $0, h, 2h, \dots,$ we approximate the exact solution based on continuous observations using the available discrete-time observations. This approach has the advantage that for very closely spaced observations it performs well and is extremely simple to implement.

This idea can be extended to non-linear (in particular threshold) continuous-time autoregressions. We illustrate this in Sections 4, 5 and 6. The assumption of uniform spacing, which we make in all our examples, can also be relaxed providing the maximum spacing between observations is small.

Before considering this alternative approach, we first examine the method of Phillips as applied to CAR(2) processes. This method has the advantage of requiring only the fitting of a discrete-time ARMA process to the discretely observed data and the subsequent transformation of the estimated coefficients to continuous-time equivalents. We derive the asymptotic distribution of these estimators as first $T \to \infty$ and then $h \to 0$.

Example 1. For the CAR(2) process defined by

$$(D^2 + a_1 D + a_2)Y(t) = bDW(t),$$

the sampled process $\{Y_n^{(h)} = Y(nh), n = 0, 1, \dots\}$ satisfies

$$Y_n^{(h)} - \phi_1^{(h)} Y_{n-1}^{(h)} - \phi_2^{(h)} Y_{n-2}^{(h)} = Z_n + \theta^{(h)} Z_{n-1}, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2(h)).$$

For fixed h, as $T \to \infty$, the maximum likelihood estimator of $\beta = (\phi_1^{(h)}, \phi_2^{(h)}, \theta^{(h)})^T$ based on observations $Y_1^{(h)}, \dots, Y_{[T/h]}^{(h)}$ satisfies (see Brockwell and Davis (1991), p.258)

$$\sqrt{T/h}(\hat{\beta} - \beta) \Rightarrow N(\mathbf{0}, M(\beta)). \tag{2.5}$$

where

$$M(\beta) = \sigma^2 \begin{bmatrix} E\mathbf{U}_t\mathbf{U}_t^T & E\mathbf{V}_t\mathbf{U}_t^T \\ E\mathbf{U}_t\mathbf{V}_t^T & E\mathbf{V}_t\mathbf{V}_t^T \end{bmatrix}^{-1}, \tag{2.6}$$

and the random vectors $\mathbf{U}_t$ and $\mathbf{V}_t$ are defined as $\mathbf{U}_t = (U_t, \dots, U_{t+1-p})^T$ and $\mathbf{V}_t = (V_t, \dots, V_{t+1-q})^T$, where $\{U_t\}$ and $\{V_t\}$ are stationary solutions of the autoregressive equations,

$$\phi(B)U_t = Z_t \qquad \text{and} \qquad \theta(B)V_t = Z_t. \tag{2.7}$$

In order to determine the asymptotic behaviour as $T \to \infty$ of the maximum likelihood estimators $(\hat{\phi}_1(h), \hat{\phi}_2(h))$, we consider the top left 2×2 submatrix M_2 of the matrix M. For small h we find that M_2 has the representation,

$$M_2 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} (2a_1h + \frac{2}{\sqrt{3}}(2-\sqrt{3})a_1^2h^2 + \frac{4}{3}(2-\sqrt{3})a_1^3h^3) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} a_1a_2h^3 + O(h^4) \quad \text{as } h \to 0. \tag{2.8}$$

The mapping from (ϕ_1, ϕ_2) to (a_1, a_2) is as follows:

$$a_1 = -\log(-\phi_2)/h,$$

$$a_2 = \frac{1}{h^2}\log\left(\frac{\phi_1}{2} + \sqrt{\frac{\phi_1^2}{4} + \phi_2}\right)\log\left(\frac{\phi_1}{2} - \sqrt{\frac{\phi_1^2}{4} + \phi_2}\right).$$

The matrix

$$C = \begin{bmatrix} \frac{\partial a_1}{\partial \phi_1} & \frac{\partial a_1}{\partial \phi_2} \\ \frac{\partial a_2}{\partial \phi_1} & \frac{\partial a_2}{\partial \phi_2} \end{bmatrix}$$

therefore has the asymptotic expansion

$$C = \begin{bmatrix} 0 & \frac{1}{h}\left(1 + a_1h + \frac{a_1^2}{2}h^2 + \cdots\right) \\ -\frac{1}{h^2}\left(1 + \frac{a_1}{2}h + \frac{a_1^2+2a_2}{12}h^2 + \cdots\right) & -\frac{1}{h^2}\left(1 + \frac{a_1}{2}h + \frac{a_1^2-4a_2}{12}h^2 + \cdots\right) \end{bmatrix}. \tag{2.9}$$

From (2.8) and (2.9) we find that

$$CM_2C^T = \frac{1}{h}\begin{bmatrix} 2a_1 & 0 \\ 0 & 2a_1a_2 \end{bmatrix}(1 + o(1)) \text{ as } h \to 0. \tag{2.10}$$

and hence, from (2.5) that the maximum likelihood estimator $\hat{\mathbf{a}}$ of $\mathbf{a} = (a_1, a_2)^T$ based on observations of Y at times $0, h, 2h, \ldots, h[T/h]$, satisfies

$$\sqrt{T}(\hat{\mathbf{a}} - \mathbf{a}) \Rightarrow N(\mathbf{0}, V), \text{ as } T \to \infty,$$

where

$$V = \begin{bmatrix} 2a_1 & 0 \\ 0 & 2a_1a_2 \end{bmatrix}(1 + o(1)) \text{ as } h \to 0. \tag{2.11}$$

Remark 1. Since the moving average coefficient $\theta^{(h)}$ of the sampled process is also a function of the parameters a_1 and a_2, and hence of $\phi_1^{(h)}$ and $\phi_2^{(h)}$, the question arises as to whether the discrete-time likelihood maximization should be carried out subject to the constraint imposed by the functional relationship between $\phi_1^{(h)}, \phi_2^{(h)}$ and $\theta^{(h)}$. However, as we shall see, the unconstrained estimation which we have considered in the preceding example leads to an asymptotic distribution of the estimators which, as $h \to 0$, converges to that of the maximum likelihood estimators based on the process observed continuously on the interval $[0, T]$. This indicates, at least asymptotically, that there is no gain in using the more complicated constrained maximization of the likelihood, so that widely available standard ARMA fitting techniques can be used.

Remark 2. As the spacing h converges to zero, the autoregressive roots $\exp(-\lambda_j h)$ converge to 1, leading to numerical difficulties in carrying out the discrete-time maximization. For this reason we consider next an approach which uses exact results for the continuously observed process to develop approximate maximum likelihood estimators for closely-spaced discrete-time observations. The same approach can be used not only for linear continuous-time autoregressions, but also for non-linear autoregressions such as continuous-time analogues of the threshold models of Tong (1983).

3. Inference for Continuously Observed Autoregressions

We now consider a more general form of (2.1), i.e.

$$(D^p + a_1 D^{p-1} + \cdots + a_p) Y(t) = b(DW(t) + c), \tag{3.1}$$

in which we allow the parameters $a_1, \ldots, a_p$ and c to be bounded measurable functions of $Y(t)$ and assume that b is constant. In particular if we partition the real line into subintervals, $(-\infty, y_1]$, $(y_1, y_2]$, $\ldots$, (y_m, ∞), on each of which the parameter values are constant, then we obtain a continuous-time analogue of the threshold models of Tong (1983) which we shall refer to as a CTAR(p) process. Continuous-time threshold models have been used by a number of authors (e.g. Tong and Yeung (1991), Brockwell and Williams (1997)) for the modelling of financial and other time series).

The equation (3.1) has a state space representation analogous to (2.2) and (2.3), namely

$$Y(t) = bX_0(t), \tag{3.2}$$

where

$$\begin{aligned} dX_0 &= X_1(t)dt, \\ dX_1 &= X_2(t)dt, \\ &\vdots \\ dX_{p-2} &= X_{p-1}(t)dt, \\ dX_{p-1} &= [-a_p X_0(t) - \cdots - a_1 X_{p-1}(t) + c]dt + dW(t), \end{aligned} \tag{3.3}$$

and we have abbreviated $a_i(Y(t))$ and $c(Y(t))$ to a_i and c respectively. We show next that (3.3) with initial condition $\mathbf{X}(0) = \mathbf{x} = (x_0, x_1, \cdots, x_{p-1})^T$ has a unique (in law) weak solution $\mathbf{X} = (\mathbf{X}(t), 0 \le t \le T)$ and determine the probability density of the random function $X_{p-1} = (X_{p-1}(t), 0 \le t \le T)$ with respect to Wiener measure. For parameterized functions a_i and c, this allows the possibility of maximization of the likelihood, conditional on $\mathbf{X}(0) = \mathbf{x}$, of $\{X_{p-1}(t), 0 \le t \le T\}$. Of course a complete set of observations of $\{X_{p-1}(t), 0 \le t \le T\}$ is not generally available unless X_0 is observed continuously. Nevertheless the parameter values which maximize the likelihood of $\{X_{p-1}(t), 0 \le t \le T\}$ can be expressed in terms of observations of $\{Y(t), 0 \le t \le T\}$ as described in subsequent sections. If Y is observed at discrete times, the stochastic integrals appearing in the solution for continuously observed autoregressions will be approximated by corresponding approximating sums. Other methods for dealing with the problem of estimation for continuous-time autoregressions based on discrete-time observations are considered by Stramer and Roberts (2004) and by Tsai and Chan (1999, 2000).

Assuming that $\mathbf{X}(0) = \mathbf{x}$, we can write $\mathbf{X}(t)$ in terms of $\{X_{p-1}(s), 0 \le s \le t\}$ using the relations, $X_{p-2}(t) = x_{p-2} + \int_0^t X_{p-1}(s)ds, \ldots, X_0(t) = x_0 + \int_0^t X_1(s)ds$. The resulting functional relationship will be denoted by

$$\mathbf{X}(t) = \mathbf{F}(X_{p-1}, t). \tag{3.4}$$

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

$$dX_{p-1} = G(X_{p-1}, t)dt + dW(t), \tag{3.5}$$

where $G(X_{p-1},t)$, like $\mathbf{F}(X_{p-1},t)$, depends on $\{X_{p-1}(s), 0 \le s \le t\}$.

Now let B be standard Brownian motion (with $B(0) = x_{p-1}$) defined on the probability space $(C[0,T], \mathcal{B}[0,T], P_{x_{p-1}})$ and, for $t \le T$, let $\mathcal{F}_t = \sigma\{B(s), s \le t\} \vee \mathcal{N}$, where $\mathcal{N}$ is the sigma-algebra of $P_{x_{p-1}}$-null sets of $\mathcal{B}[0,T]$. The equations

$$\begin{aligned} dZ_0 &= Z_1 dt, \\ dZ_1 &= Z_2 dt, \\ &\vdots \\ dZ_{p-2} &= Z_{p-1} dt, \\ dZ_{p-1} &= dB(t), \end{aligned} \tag{3.6}$$

with $\mathbf{Z}(0) = \mathbf{x} = (x_0, x_1, \cdots, x_{p-1})^T$, clearly have the unique strong solution, $\mathbf{Z}(t) = \mathbf{F}(B,t)$, where $\mathbf{F}$ is defined as in (3.4). Let G be the functional appearing in (3.5) and suppose that $\hat{W}$ is the Ito integral defined by $\hat{W}(0) = x_{p-1}$ and

$$d\hat{W}(t) = -G(B,t)dt + dB(t) = -G(Z_{p-1},t)dt + dZ_{p-1}(t). \tag{3.7}$$

For each T, we now define a new measure $\hat{P}_{\mathbf{x}}$ on $\mathcal{F}_T$ by

$$d\hat{P}_{\mathbf{x}} = M(B,T) dP_{x_{p-1}}, \tag{3.8}$$

where

$$M(B,T) = \exp\left[-\frac{1}{2}\int_0^T G^2(B,s)ds + \int_0^T G(B,s)dW(s)\right]. \tag{3.9}$$

Then by the Cameron-Martin-Girsanov formula (see e.g. Øksendal (1998), p.152), $\{\hat{W}(t),\ 0 \le t \le T\}$ is standard Brownian motion under $\hat{P}_{\mathbf{x}}$. Hence we see from (3.7) that the equations (3.5) and (2.3) with initial condition $\mathbf{X}(0) = \mathbf{x}$ have, for $t \in [0,T]$, the weak solutions $(Z_{p-1}(t), \hat{W}(t))$ and $(\mathbf{Z}(t), \hat{W}(t))$ respectively. Moreover, by Proposition 5.3.10 of Karatzas and Shreve (1991), the weak solution is unique in law, and by Theorem 10.2.2 of Stroock and Varadhan (1979) it is non-explosive.

If f is a bounded measurable functional on $C[0,T]$,

$$\begin{aligned} \hat{E}_{\mathbf{x}} f(Z_{p-1}) &= E_{x_{p-1}}(M(B,T)f(B)) \\ &= \int f(\xi) M(\xi,T) dP_{x_{p-1}}(\xi). \end{aligned}$$

In other words, $M(\xi, T)$ is the density at $\xi \in C[0,T]$, conditional on $\mathbf{X}(0) = \mathbf{x}$, of the distribution of X_{p-1} with respect to the Wiener measure $P_{x_{p-1}}$ and, if we observed $X_{p-1} = \xi$, we could compute conditional maximum likelihood estimators of the unknown parameters by maximizing $M(\xi, T)$.

4. Estimation for CAR(p) Processes

For the CAR(p) process defined by (2.1), denoting the realized state process on $[0,T]$ by $\{\mathbf{x}(s) = (x_0(s), x_1(s), \dots, x_{p-1}(s))^T, 0 \le s \le T\}$, we have, in the notation of Section 3,

$$-2 \log M(x_{p-1}, s) = \int_0^T G^2 ds - 2 \int_0^T G dx_{p-1}(s), \tag{4.1}$$

where

$$G = -a_1 x_{p-1}(s) - a_2 x_{p-2}(s) - \cdots - a_p x_0(s). \tag{4.2}$$

Differentiating $\log M$ partially with respect to $a_1, \dots, a_p$ and setting the derivatives equal to zero gives the maximum likelihood estimators, conditional on $\mathbf{X}(0) = \mathbf{x}(0)$,

$$\begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{bmatrix} = - \begin{bmatrix} \int_0^T x_{p-1}^2 ds & \cdots & \int_0^T x_{p-1} x_0 ds \\ \vdots & \ddots & \vdots \\ \int_0^T x_{p-1} x_0 ds & \cdots & \int_0^T x_0^2 ds \end{bmatrix}^{-1} \begin{bmatrix} \int_0^T x_{p-1} dx_{p-1} \\ \vdots \\ \int_0^T x_0 dx_{p-1} \end{bmatrix}. \tag{4.3}$$

Note that this expression for the maximum likelihood estimators is unchanged if x is replaced throughout by y, where y_0 denotes the observed CAR(p) process and y_j denotes its j^{th} derivative.

Differentiating $\log M$ twice with respect to the parameters $a_1, \dots, a_p$, taking expected values and assuming that the zeroes of the autoregressive polynomial a all have negative real parts, we find that

$$-E \frac{\partial^2 \log M}{\partial \mathbf{a}^2} \sim T \Sigma \text{ as } T \to \infty, \tag{4.4}$$

where Σ is the covariance matrix of the limit distribution as $T \to \infty$ of the random vector $(X_{p-1}(t), X_{p-2}(t), \dots, X_0(t))^T$. It is known (see Arató (1982)) that

$$\Sigma^{-1} = \mathbf{2} \left[m_{ij} \right]_{i,j=1}^{p}, \tag{4.5}$$

where $m_{ij} = m_{ji}$ and for $j \geq i$,

$$m_{ij} = \begin{cases} 0 & \text{if } j - i \text{ is odd,} \\ \sum_{k=0}^{\infty} (-1)^k a_{i-1-k} a_{j+k} & \text{otherwise,} \end{cases}$$

where $a_0 := 1$ and $a_j := 0$ if $j > p$ or $j < 0$, and that the estimators given by (4.3) satisfy

$$\sqrt{T}(\hat{\mathbf{a}} - \mathbf{a}) \Rightarrow N(\mathbf{0}, \Sigma^{-1}), \tag{4.6}$$

where Σ^{-1} is given by (4.5). The asymptotic result (4.6) also holds for the Yule-Walker estimates of $\mathbf{a}$ as found by Hyndman (1993).

In the case $p = 1$, $\Sigma^{-1} = 2a_1$ and when $p = 2$, Σ^{-1} is the same as the leading term in the expansion of the covariance matrix V in (2.11).

In order to derive approximate maximum likelihood estimators for closely-spaced observations of the CAR(p) process defined by (2.1) we shall use the result (4.3) with the stochastic integrals replaced by approximating sums. Thus if observations are made at times $0, h, 2h, \ldots$, we replace, for example,

$$\int_0^T x'(s)^2 ds \quad \text{by} \quad \frac{1}{h} \sum_{i=0}^{[T/h]-1} (x((i+1)h) - x(ih))^2,$$

$$\int_0^T x'(s) dx'(s) \quad \text{by} \quad \frac{1}{h^2} \sum_{i=0}^{[T/h]-3} (x((i+1)h) - x(ih))$$
$$\times (x((i+3)h) - 2x((i+2)h) + x((i+1)h)),$$

taking care, as in the latter example, to preserve the non-anticipating property of the integrand in the corresponding approximating sum.

Example 2. For the CAR(2) process defined by

$$(D^2 + a_1 D + a_2) Y(t) = b DW(t),$$

Table 1 shows the result of using approximating sums for the estimators defined by (4.3) in order to estimate the coefficients a_1 and a_2.

As expected, the variances of the estimators are reduced by a factor of approximately 5 as T increases from 100 to 500 with h fixed. As h increases with T

Table 1. Estimated coefficients based on 1000 replicates on $[0,T]$ of the linear CAR(2) process with $a_1 = 1.8$ and $a_2 = 0.5$

h		T=100 Sample mean of estimators	T=100 Sample variance of estimators	T=500 Sample mean of estimators	T=500 Sample variance of estimators
0.001	a_1	1.8120	0.03585	1.7979	0.006730
	a_2	0.5405	0.02318	0.5048	0.003860
0.01	a_1	1.7864	0.03404	1.7727	0.006484
	a_2	0.5362	0.02282	0.5007	0.003799
0.1	a_1	1.5567	0.02447	1.5465	0.004781
	a_2	0.4915	0.01902	0.4588	0.003217

fixed, the variances actually decrease while the bias has a tendency to increase. This leads to mean squared errors which are quite close for $h = .001$ and $h = .01$. The asymptotic covariance matrix Σ^{-1} in (4.6), based on continuously observed data, is diagonal with entries 3.6 and 1.8. For $h = .001$ and $h = .01$, the variances $3.6/T$ and $1.8/T$ agree well with the corresponding entries in the table.

5. Estimation for CTAR(p) Processes

The density derived in Section 3 is not restricted to linear continuous-time autoregressions as considered in the previous section. It applies also to non-linear autoregressions and in particular to CTAR models as defined by (3.2) and (3.3). In this section we illustrate the application of the continuous-time maximum likelihood estimators and corresponding approximating sums to the estimation of coefficients in CTAR(1) and CTAR(2) models.

Example 3. Consider the CTAR(1) process defined by

$$DY(t) + a_1^{(1)}Y(t) = bDW(t), \quad \text{if } Y(t) < 0,$$

$$DY(t) + a_1^{(2)}Y(t) = bDW(t), \quad \text{if } Y(t) \geq 0,$$

with $b > 0$ and $a_1^{(1)} \neq a_1^{(2)}$. We can write

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

where

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

and $a(x) = a_1^{(1)}$ if $x < 0$ and $a(x) = a_1^{(2)}$ if $x \geq 0$. Proceeding as in Section 4, $-2\log M$ is as in (4.1) with

$$G = -a_1^{(1)} x(s) I_{x(s)<0} - a_1^{(2)} x(s) I_{x(s)\geq 0}. \tag{5.1}$$

Maximizing $\log M$ as in Section 4, we find that

$$\hat{a}_1^{(1)} = -\frac{\int_0^T I_{x(s)<0}\, x(s)\, dx(s)}{\int_0^T I_{x(s)<0}\, x(s)^2 ds}$$

and

$$\hat{a}_1^{(2)} = -\frac{\int_0^T I_{x(s)\geq 0}\, x(s)\, dx(s)}{\int_0^T I_{x(s)\geq 0}\, x(s)^2 ds},$$

where, as in Section 4, x can be replaced by y in these expressions. For observations at times $0, h, 2h, \ldots$, with h small the integrals in these expressions were replaced by corresponding approximating sums and the resulting estimates are shown in Table 2.

Table 2. Estimated coefficients based on 1000 replicates on $[0, T]$ of the threshold AR(1) with threshold $r = 0$, $a_1^{(1)} = 6$, $a_1^{(2)} = 1.5$

h		T=100		T=500	
		Sample mean of estimators	Sample variance of estimators	Sample mean of estimators	Sample variance of estimators
0.001	$a_1^{(1)}$	6.0450	0.41207	5.9965	0.07185
	$a_1^{(2)}$	1.5240	0.04824	1.4986	0.00891
0.01	$a_1^{(1)}$	5.8978	0.39427	5.8472	0.06785
	$a_1^{(2)}$	1.5135	0.04771	1.4875	0.00883
0.1	$a_1^{(1)}$	4.7556	0.27969	4.7085	0.04506
	$a_1^{(2)}$	1.3891	0.03840	1.3682	0.00711

Again we see that as T increases from 100 to 500, the variances of the estimators are reduced by a factor of approximately 5. As h increases with T fixed, the variances decrease while the bias tends to increase, the net effect being (as expected) an increase in mean squared error with increasing h.

Example 4. Consider the CTAR(2) process defined by

$$D^2Y(t) + a_1^{(1)}DY(t) + \ a_2^{(1)}Y(t) = bDW(t), \quad \text{if } Y(t) < 0,$$

$$D^2Y(t) + a_1^{(2)}DY(t) + \ a_2^{(2)}Y(t) = bDW(t), \quad \text{if } Y(t) \geq 0,$$

with $a_1^{(1)} \neq a_1^{(2)}$ or $a_2^{(1)} \neq a_2^{(2)}$, and $b > 0$. We can write

$$Y(t) = (b, 0)\,\mathbf{X}(t),$$

where

$$d\mathbf{X}(t) = A\mathbf{X}(t)dt + \mathbf{e}\ dW(t),$$

and $A = A^{(1)}$ if $x < 0$ and $A = A^{(2)}$ if $x \geq 0$, where

$$A^{(1)} = \begin{bmatrix} 0 & 1 \\ -a_2^{(1)} & -a_1^{(1)} \end{bmatrix}, \ A^{(2)} = \begin{bmatrix} 0 & 1 \\ -a_2^{(2)} & -a_1^{(2)} \end{bmatrix}, \ \mathbf{e} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Proceeding as in Section 4, $-2\log M$ is as in (4.1) with

$$G = \left(-a_1^{(1)}x_1(s) - a_2^{(1)}x(s)\right) I_{x(s)<0} + \left(-a_1^{(2)}x_1(s) - a_2^{(2)}x(s)\right) I_{x(s)\geq 0}. \tag{5.2}$$

Maximizing $\log M$ as in Section 4, we find that

$$\begin{bmatrix} \hat{a}_1^{(1)} \\ \hat{a}_2^{(1)} \end{bmatrix} = -\begin{bmatrix} \int_0^T I_{x(s)<0}x_1^2(s)ds & \int_0^T I_{x(s)<0}x_1(s)x_0(s)ds \\ \int_0^T I_{x(s)<0}x_1(s)x_0(s)ds & \int_0^T I_{x(s)<0}x_0^2(s)ds \end{bmatrix}^{-1} \times$$

$$\begin{bmatrix} \int_0^T I_{x(s)<0}x_1(s)dx_1(s) \\ \int_0^T I_{x(s)<0}x_0(s)dx_1(s) \end{bmatrix},$$

while $[\hat{a}_1^{(2)}, \hat{a}_2^{(2)}]^T$ satisfies the same equation with $I_{x(s)<0}$ replaced throughout by $I_{x(s)\geq 0}$.

As in Section 4, x can be replaced by y in these expressions. For observations at times $0, h, 2h, \ldots$, with h small, the integrals in these expressions were replaced by corresponding approximating sums and the resulting estimates are shown in Table 3.

The pattern of results is more complicated in this case. As T is increased from 100 to 500 with h fixed, the sample variances all decrease, but in a less regular fashion than in Tables 1 and 2. As h increases with T fixed, the variances also

Table 3. Estimated coefficients based on 1000 replicates on $[0,T]$ of the threshold AR(2) with threshold $r=0$, $a_1^{(1)}=1.5$, $a_2^{(1)}=0.4$, $a_1^{(2)}=4.6$, $a_2^{(2)}=2$

h		T=100		T=500	
		Sample mean of estimators	Sample variance of estimators	Sample mean of estimators	Sample variance of estimators
0.001	$a_1^{(1)}$	1.5187	0.05441	1.5071	0.01128
	$a_2^{(1)}$	0.4763	0.04119	0.4163	0.00480
	$a_1^{(2)}$	4.6084	0.21224	4.5755	0.03995
	$a_2^{(2)}$	2.3186	0.72069	2.0456	0.08881
0.01	$a_1^{(1)}$	1.5262	0.05234	1.5163	0.01095
	$a_2^{(1)}$	0.4729	0.04056	0.4135	0.00473
	$a_1^{(2)}$	4.3819	0.19823	4.3480	0.03746
	$a_2^{(2)}$	2.2697	0.68915	2.0025	0.08509
0.1	$a_1^{(1)}$	1.5091	0.04177	1.4928	0.00805
	$a_2^{(1)}$	0.4402	0.03489	0.3851	0.00411
	$a_1^{(2)}$	2.7053	0.11312	2.7014	0.01874
	$a_2^{(2)}$	1.7654	0.41380	1.5599	0.05221

decrease. The mean squared errors for $h=.001$ and $h=.01$ are again quite close.

6. Estimation when the threshold is unknown

In the previous section we considered the estimation of the autoregressive coefficients only, under the assumption that the threshold r is known. In this section we consider the corresponding problem when the threshold also is to be estimated. The idea is the same, that is to maximize the (conditional) likelihood of the continuously-observed process, using the closely spaced discrete observations to approximate what would be the exact maximum likelihood estimators if the continuously-observed data were available. We illustrate first with a CTAR(1) process. The goal is to use observations $\{y(kh), k=1,2,\dots;0<kh\le T\}$, with h small, to estimate the parameters $a_1^{(1)}, a_1^{(2)}, c_1^{(1)}, c_1^{(2)}, b$ and r in the following model.

$$\begin{aligned} D\left(Y(t)-r\right)+a_1^{(1)}\left(Y(t)-r\right)+c_1^{(1)} &= bDW_t, \quad Y(t)<r, \\ D\left(Y(t)-r\right)+a_1^{(2)}\left(Y(t)-r\right)+c_1^{(2)} &= bDW_t, \quad Y(t)\ge r. \end{aligned} \tag{6.1}$$

The process $Y^* = Y - r$, satisfies the threshold autoregressive equations,

$$\begin{aligned} DY^*(t) + a_1^{(1)}Y^*(t) + c_1^{(1)} &= bDW_t, \quad Y^*(t) < 0, \\ DY^*(t) + a_1^{(2)}Y^*(t) + c_1^{(2)} &= bDW_t, \quad Y^*(t) \geq 0, \end{aligned}$$

with state-space representation,

$$Y^*(t) = bX(t),$$

where

$$dX(t) = G(X,t)dt + dW(t),$$

as in equation (3.5), and

$$G(x,s) = -\left(a_1^{(1)}x(s) + \frac{c_1^{(1)}}{b}\right) I_{x(s)<0} - \left(a_1^{(2)}x(s) + \frac{c_1^{(2)}}{b}\right) I_{x(s)\geq 0}.$$

Substituting for G in the expression (4.1), we obtain

$$\begin{aligned} -2\log M(x(s),s) &= \int_0^T G^2 ds - 2\int_0^T G dx(s) \\ &= \int_0^T \left(a_1^{(1)}x(s) + \frac{c_1^{(1)}}{b}\right)^2 I_{x(s)<0} ds + \int_0^T \left(a_1^{(2)}x(s) + \frac{c_1^{(2)}}{b}\right)^2 I_{x(s)\geq 0} ds \\ &\quad +2\int_0^T \left(a_1^{(1)}x(s) + \frac{c_1^{(1)}}{b}\right) I_{x(s)<0} dx(s) + 2\int_0^T \left(a_1^{(2)}x(s) + \frac{c_1^{(2)}}{b}\right) I_{x(s)\geq 0} dx(s) \\ &= \frac{1}{b^2}\left[\int_0^T \left(a_1^{(1)}y^* + c_1^{(1)}\right)^2 I_{y^*<0} ds + \int_0^T \left(a_1^{(2)}y^* + c_1^{(2)}\right)^2 I_{y^*\geq 0} ds \right. \\ &\quad \left. +2\int_0^T \left(a_1^{(1)}y^* + c_1^{(1)}\right) I_{y^*<0} dy^* + 2\int_0^T \left(a_1^{(2)}y^* + c_1^{(2)}\right) I_{y^*\geq 0} dy^*\right]. \end{aligned}$$

Minimizing $-2\log M(x(s),s)$ with respect to $a_1^{(1)}$, $a_1^{(2)}$, $c_1^{(1)}$, and $c_1^{(2)}$ with b fixed gives,

$$\begin{aligned} &\hat{a}_1^{(1)}(r)\left[\int_0^T y^{*2} I_{y^*<0} ds \int_0^T I_{y^*<0} ds - \left(\int_0^T y^* I_{y^*<0} ds\right)^2\right] \\ &\quad = -\left[\int_0^T y^* I_{y^*<0} dy^* \int_0^T I_{y^*<0} ds - \int_0^T I_{y^*<0} dy^* \int_0^T y^* I_{y^*<0} ds\right] \end{aligned}$$

(6.2)
$$\hat{c}_1^{(1)}(r)\left[\int_0^T y^{*2} I_{y^*<0} ds \int_0^T I_{y^*<0} ds - \left(\int_0^T y^* I_{y^*<0} ds\right)^2\right]$$
$$= -\left[\int_0^T I_{y^*<0} dy^* \int_0^T y^{*2} I_{y^*<0} ds - \int_0^T y^* I_{y^*<0} dy^* \int_0^T y^* I_{y^*<0} ds\right],$$

with analogous expressions for $\hat{a}_1^{(2)}$ and $\hat{c}_1^{(2)}$. An important feature of these equations is that they involve only the values of $y^* = y - r$ and not b.

For any fixed value of r and observations y, we can therefore compute the maximum likelihood estimators $\hat{a}_1^{(1)}(r)$, $\hat{a}_1^{(2)}(r)$, $\hat{c}_1^{(1)}(r)$ and $\hat{c}_1^{(2)}(r)$ and the corresponding minimum value, $m(r)$, of $-2b^2 \log M$. The maximum likelihood estimator $\hat{r}$ of r is the value which minimizes $m(r)$ (this minimizing value also being independent of b). The maximum likelihood estimators of $a_1^{(1)}, a_1^{(2)}, c_1^{(1)}$ and $c_1^{(2)}$ are the values obtained from (6.2) with $r = \hat{r}$. Since the observed data are the discrete observations $\{y(h), y(2h), y(3h), \dots\}$, the calculations just described are all carried out with the integrals in (6.2) replaced by approximating sums as described in Section 4.

If the data y are observed continuously, the quadratic variation of y on the interval $[0, T]$ is *exactly* equal to b^2T. The discrete approximation to b based on $\{y(h), y(2h), \dots\}$ is

$$\hat{b} = \sqrt{\sum_{k=1}^{[T/h]-1} (y((k+1)h) - y(kh))^2/T}. \tag{6.3}$$

Example 5. Table 4 shows the results obtained when the foregoing estimation procedure is applied to a CTAR(1) process defined by (6.1) with $a_1^{(1)} = 6$, $c_1^{(1)} = .5$, $a_1^{(2)} = 1.5$, $c_1^{(2)} = .4$, $b = 1$ and $r = 10$.

The pattern of results is again rather complicated. As expected however there is a clear reduction in sample variance of the estimators as T is increased with h fixed. For $T = 1000$ the mean squared errors of the estimators all increase as h increases, with the mean squared errors when $h = .001$ and $h = .01$ being rather close and substantially better than those when $h = .1$.

Table 4. The sample mean and sample variance of the estimators of the parameters of the model (6.1) based on 1000 replicates of the process on $[0, T]$. The parameters of the simulated process are $a_1^{(1)} = 6$, $c_1^{(1)} = .5$, $a_1^{(2)} = 1.5$, $c_1^{(2)} = .4$, $b = 1$ and $r = 10$.

h		T=100 Sample mean	T=100 Sample variance	T=500 Sample mean	T=500 Sample variance	T=1000 Sample mean	T=1000 Sample variance
0.001	$a_1^{(1)}$	5.9179	1.5707	5.9758	0.1950	5.9835	0.0904
	$c_1^{(1)}$	0.3787	0.7780	0.3448	0.1561	0.3832	0.0753
	$a_1^{(2)}$	1.7149	0.4105	1.5370	0.0511	1.5178	0.0224
	$c_1^{(2)}$	0.2891	0.1476	0.3415	0.0273	0.3601	0.0133
	b	0.9996	5.00×10^{-6}	0.9991	4.78×10^{-7}	0.9991	4.84×10^{-7}
	r	9.9963	0.0244	9.9769	0.0041	9.9818	0.0020
0.01	$a_1^{(1)}$	5.7201	1.3175	5.7507	0.1834	5.7614	0.0910
	$c_1^{(1)}$	0.4271	0.7524	0.3235	0.1699	0.3535	0.0705
	$a_1^{(2)}$	1.7373	0.4248	1.5567	0.0538	1.5360	0.0239
	$c_1^{(2)}$	0.2877	0.1598	0.3227	0.0357	0.3407	0.0162
	b	0.9914	4.82×10^{-5}	0.9913	4.55×10^{-6}	0.9907	4.91×10^{-6}
	r	10.011	0.0278	9.9807	0.0058	9.984	0.0024
0.1	$a_1^{(1)}$	4.1166	0.7861	4.1087	0.1587	4.1115	0.0720
	$c_1^{(1)}$	0.3953	0.5638	0.2834	0.2287	0.2708	0.0944
	$a_1^{(2)}$	1.7308	0.5109	1.5924	0.0805	1.5851	0.0324
	$c_1^{(2)}$	0.2636	0.2391	0.2658	0.1003	0.2666	0.0472
	b	0.9191	5.00×10^{-4}	0.9208	4.60×10^{-5}	0.9160	4.79×10^{-5}
	r	10.074	0.0425	10.038	0.0191	10.030	0.0086

Example 6. Although the procedure described above is primarily intended for use in the modelling of very closely spaced data, in this example we illustrate its performance when applied to the natural logarithms of the annual Canadian lynx trappings, 1821 - 1934 (see e.g. Brockwell and Davis (1991), p.559). Linear and threshold autoregressions of order two were fitted to this series by Tong and Yeung (1991) and a linear CAR(2) model using a continuous-time version of the Yule-Walker equations by Hyndman (1993).

The threshold AR(2) model fitted by Tong and Yeung (1991) to this series was

$$
(6.4)\qquad \begin{aligned} &D^2Y(t) + a_1^{(1)}DY(t) + \ a_2^{(1)}Y(t) = b_1DW(t), \quad \text{if } Y(t) < r, \\ &D^2Y(t) + a_1^{(2)}DY(t) + \ a_2^{(2)}Y(t) = b_2DW(t), \quad \text{if } Y(t) \geq r, \end{aligned}
$$

with

$$
(6.5)\qquad \begin{aligned} &a_1^{(1)} = .354, \quad a_1^{(2)} = .521, \ b_1 = .707, \\ &a_1^{(2)} = 1.877, \ a_2^{(2)} = .247, \ b_2 = .870, \end{aligned}
$$

and threshold $r = 0.857$.

An argument exactly parallel to that for the CTAR(1) process at the beginning of this section permits the estimation of the coefficients and threshold of a CTAR(2) model of this form with $b_1 = b_2 = b$, $h = 1$ and with time measured in years. It leads to the coefficient estimates,

$$
(6.6)\qquad \begin{aligned} &a_1^{(1)} = .3163, \quad a_1^{(2)} = .1932, \ b_1 = 1.150, \\ &a_1^{(2)} = 1.2215, \ a_2^{(2)} = .9471, \ b_2 = 1.150, \end{aligned}
$$

with estimated threshold $r = 0.478$. (Because of the large spacing of the observations in this case it is difficult to obtain a good approximation to the quadratic variation of the derivative of the process. The coefficient b was therefore estimated by a simple one-dimensional maximization of the Gaussian likelihood (GL) of the original discrete observations (computed as described by Brockwell(2001)), with the estimated coefficients fixed at the values specified above.)

In terms of the Gaussian likelihood of the original data, the latter model (with $-2\log(GL) = 220.15$) is considerably better than the Tong and Yeung model (for which $-2\log(GL) = 244.41$). Using our model as an initial approximation for maximizing the Gaussian likelihood of the original data, we obtain the following more general model, which has higher Gaussian likelihood than both of the preceding models ($-2\log(GL) = 161.06$).

(6.7)

$$
\begin{aligned} &D^2Y(t) + 1.181DY(t) + \ 0.308Y(t) - 0.345 = 1.050DW(t), \quad \text{if } Y(t) < -0.522, \\ &D^2Y(t) + 0.0715DY(t) + \ 0.452Y(t) + 0.500 = 0.645DW(t), \quad \text{if } Y(t) \geq -0.522, \end{aligned}
$$

Simulations of the model (6.4) with parameters as in (6.5) and (6.6) and of the model (6.7) are shown together with the logged and mean-corrected lynx data in Figure 1. As expected, the resemblance between the sample paths and the data appears to improve with increasing Gaussian likelihood.

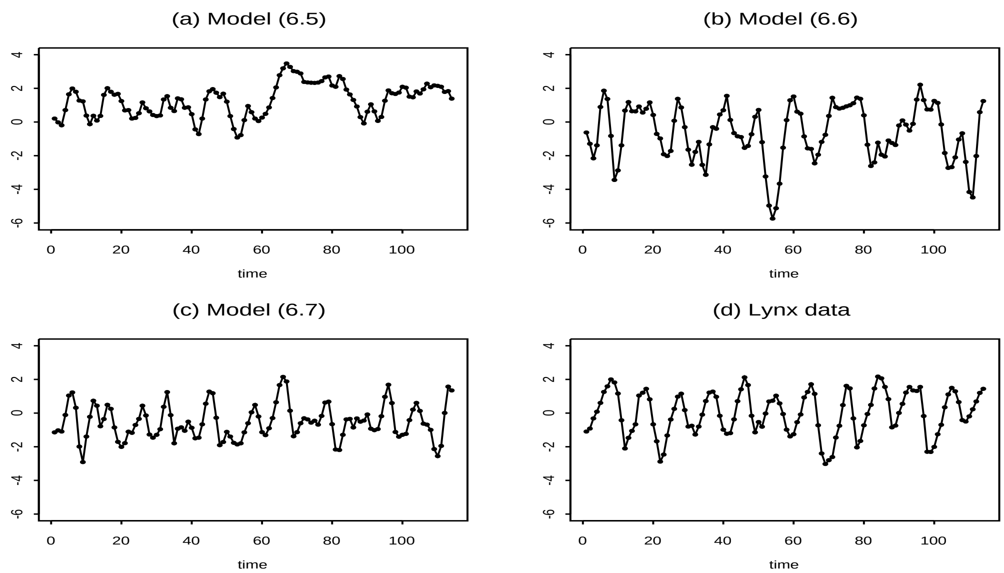

Figure 6.1: Figures (a) and (b) show simulations of the CTAR model (6.4) for the logged and mean-corrected lynx data when the parameters are given by (6.5) and (6.6) respectively. Figure (c) shows a simulation (with the same driving noise as in Figures (a) and (b)) of the model (6.7). Figure (d) show the logged and mean-corrected lynx series itself.

7. Conclusions

From the Radon-Nikodym derivative with respect to Wiener measure of the distribution of the $(p-1)^{\text{th}}$ derivative of a continuous-time linear or non-linear autoregression, observed on the interval $[0,T]$, we have shown how to compute maximum likelihood parameter estimators, conditional on the initial state vector. For closely-spaced discrete observations, the integrals appearing in the estimators are replaced by approximating sums.

The examples illustrate the accuracy of the approximations in special cases. If the observations are not uniformly spaced but the maximum spacing is small, appropriately modified approximating sums can be used in order to approximate the exact solution for the continuously observed process.

Acknowledgments

The authors are indebted to the National Science Foundation for support of this work under Grant DMS-0308109 and PB for the additional support of Deutsche Forschungsgemeinschaft, SFB 386, at Technische Universität München. We are also indebted to two referees for valuable comments and suggestions.

References

Arató, M. (1982). *Linear Stochastic Systems with Constant Coefficients*, Springer Lecture Notes in Control and Information Systems, **45**, Springer-Verlag, Berlin.

Bergstrom, A.R. (1985). The estimation of parameters in non-stationary higher-order continuous-time dynamic models. *Econometric Theory*, **1**, 369–385.

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

Brockwell, P.J. (1995). A note on the embedding of discrete-time ARMA processes. *J. Time Series Analysis* **16**, 451–460.

Brockwell, P.J. and Williams, R.J. (1997). On the existence and application of continuous-time threshold autoregressions of order two. *Adv. Appl. Prob.* **29**, 205–227.

Brockwell, P.J. (2001). Continuous-time ARMA Processes, in *Handbook of Statistics 19; Stochastic Processes: Theory and Methods*, D.N. Shanbhag and C.R. Rao, (eds), 249–276, Elsevier, Amsterdam.

Hyndman, R.J. (1993). Yule-Walker estimates for continuous-time autoregressive models. *J. Time Series Analysis* **14**, 81–296.

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

Karatzas, Ioannis and Steven E. Shreve (1991). *Brownian Motion and Stochastic Calculus* Springer-Verlag, New York.

Øksendal, Bernt (1998). *Stochastic Differential Equations*, 5^{th} ed. Springer-Verlag, Berlin.

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

Stramer, O. and G.O. Roberts (2004). On Bayesian analysis of non-linear continuous-time autoregression models. Technical Report 327, Department of Statistics and Actuarial Science, The University of Iowa, downloadable at http://www.stat.uiowa.edu/techrep/.

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

Tsai, H. and K.S. Chan (1999). A new EM method for estimating continuous-time autoregressive models. Technical Report 285, Department of Statistics and Actuarial Science, The University of Iowa.

Tsai, H. and K.S. Chan (2000). Comparison of two discretization methods for estimating continuous-time autoregressive models, in *Statistics and Finance: An Interface*, Wai-Sum Chan, Wai Keung Li and Howell Tong (eds.). Imperial College Press, London, 68–85.

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.

Peter Brockwell, Statistics Department, Colorado State University, Fort Collins, Colorado, 80523-1877, U.S.A., email: pjbrock@stat.colostate.edu.

Richard Davis, Statistics Department, Colorado State University, Fort Collins, Colorado, 80523-1877, U.S.A. email: rdavis@stat.colostate.edu

Yu Yang, Statistics Department, Colorado State University, Fort Collins, Colorado, 80523-1877, U.S.A., email: yangy@stat.colostate.edu.