Statistica Sinica 1(1991), 411-430

# THRESHOLD AUTOREGRESSIVE MODELLING IN CONTINUOUS TIME

Howell Tong and Iris Yeung

*University of Kent and City Polytechnic of Hong Kong*

*Abstract:* We have developed a procedure for identifying continuous time, self-exciting, threshold, autoregressive models and applied the procedure to several real data sets. The performance of the fitted threshold models to real data is discussed and compared with that of the fitted linear models.

*Key words and phrases:* Continuous time models, dissolved oxygen content, Hang Seng Index, IBM stock price, Kalman filter, lynx, self-exciting threshold autoregression, state space, unequally spaced data.

## 1. Introduction

During the last ten years or so, various classes of nonlinear time series models have been studied and a number of tests for nonlinearity have been developed. One particular class of the nonlinear time series models is threshold autoregressive models (TAR) proposed by Tong (1978, 1983a). There are various forms of TAR models. Basically they are piecewise linear autoregressive models in which the linear relationship varies over regimes delineated by the threshold values. If the regime is determined by the past values of the time series, the model is described as self-exciting. Petruccelli and Davies (abbreviated PD (1986)), Petruccelli (1988) and Tsay (1989) have designed tests specifically to detect self-exciting, threshold, autoregressive (SETAR)-type nonlinearity.

So far, the study on TAR models and tests of threshold nonlinearity are mainly concerned with discrete time processes observed at equally spaced time points. However, unequally spaced data are common in practice. They may be partially observed with some missing observations or irregularly observed with an arbitrary sampling interval. Considering partially and irregularly observed data as being sampled from a discrete and a continuous time process respectively, we have recently extended the tests of PD, Petruccelli and Tsay to detect general threshold nonlinearity in these two situations of unequally spaced data (see Tong and Yeung (1988,1990)). The tests have been applied to several real data sets which are, in fact, or can be treated as, unequally spaced. As threshold nonlinearity has been suggested for some of them and these data can also be considered as being sampled from a continuous time process, continuous time TAR models seem appropriate for analysing such data.

In this paper, we shall discuss identification and estimation procedures of only the continuous time SETAR model. The identification procedure is similar to that proposed by Tong and Lim (1980) for the discrete time case. However, at the estimation stage, we shall use the Kalman filter algorithm, just as we have done in tests of nonlinearity, in order to cope with unequally spaced observations. We then apply the modelling procedure to three real data sets and compare the performance of the fitted TAR models with that of the continuous time, linear, autoregressive time series models fitted by a method due to Jones (1981).

## 2. Model Identification

Consider a two-regime continuous time, self-exciting, threshold, autoregressive model of order $p$, or SETAR$(2;p,p)$, with threshold $r$,

$$\begin{aligned} y^{(p)}(t) + a_{1,p-1}y^{(p-1)}(t) + \cdots + a_{1,0}y(t) &= \varepsilon_1(t) \quad \text{if } y(t) \leq r \\ y^{(p)}(t) + a_{2,p-1}y^{(p-1)}(t) + \cdots + a_{2,0}y(t) &= \varepsilon_2(t) \quad \text{if } y(t) > r, \end{aligned} \tag{2.1}$$

where $y^{(i)}(t)$ denotes the $i$th derivative of $y(t)$ with respect to time $t$, and $\varepsilon_1(t)$, $\varepsilon_2(t)$ are independent continuous time Gaussian white noise with instantaneous variances $\sigma_1^2$, $\sigma_2^2$ respectively.

Clearly, a delay term $s$ has not been introduced in Equation (2.1) unlike the discrete time case. It would be impractical to use $y(t-s)$, $s > 0$, as the threshold variable since this would lead us to a delayed-stochastic differential equation, an area which is not well charted by the probabilists. Even delayed-deterministic differential equations are quite tricky. (See, e.g., Tong (1983b).) However, the possibility of using $\dot{y}(t)$ (i.e. $\frac{dy(t)}{dt}$) and higher derivatives as the threshold variables would be interesting and is currently under investigation.

Suppose we have $n$ equally or *unequally* spaced observations $y(t_1)$, $y(t_2), \ldots,$ $y(t_n)$, $t_1 < t_2 < \cdots < t_n$, at which points only estimates are required; then the model (2.1) can be cast in the following state space form with system equation

$$\frac{d}{dt}\begin{pmatrix} x(t) \\ x^{(1)}(t) \\ \vdots \\ x^{(p-1)}(t) \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_0(t) & -a_1(t) & -a_2(t) & \cdots & -a_{p-1}(t) \end{pmatrix} \begin{pmatrix} x(t) \\ x^{(1)}(t) \\ \vdots \\ x^{(p-1)}(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \sigma(t) \end{pmatrix} w(t) \quad (2.2)$$

where $a_j(t) = a_{ij}$, $\sigma(t) = \sigma_i$, $i = 1, 2$; $j = 0, 1, \ldots, p-1$ during the time when the observation lies in regime $i$, $w(t)$ is a continuous time Gaussian white noise with instantaneous variance equal to unity, and observation equation

$$y(t_i) = (1\,0 \cdots 0) \begin{pmatrix} x(t_i) \\ x^{(1)}(t_i) \\ \vdots \\ x^{(p-1)}(t_i) \end{pmatrix}, \qquad i = 1, \ldots, n. \quad (2.3)$$

In matrix notation, Equations (2.2) and (2.3) can be written respectively as

$$\dot{X}(t) = F(t)X(t) + G(t)w(t) \quad (2.4)$$

$$y(t_i) = HX(t_i), \qquad i = 1, \ldots, n, \quad (2.5)$$

where $X(t)$ is a $p \times 1$ column vector representing the state of the process at time $t$,
$F(t)$ is a $p \times p$ system dynamics companion matrix at time $t$,
$w(t)$ is the random input to the state equation at time $t$,
$G(t)$ is a $p \times 1$ column vector defining how the random inputs are propagated into the state at time $t$,
$H$ is a $1 \times p$ row vector defining linear combinations of the state that are observed at time $t$,
$y(t)$ is the observation at time $t$,
and the overhead dot denotes the time derivative.

Note that here $F(t)$ and $G(t)$ are time dependent (and random) and for the time instant when the observation lies in regime $i$ ($i = 1, 2$),

$$F(t) = F_i = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_{i0} & -a_{i1} & -a_{i2} & \cdots & -a_{i,p-1} \end{pmatrix}, \quad (2.6)$$

$$G(t) = G_i = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \sigma_i \end{pmatrix}. \quad (2.7)$$

Now, to evaluate the integral of Equation (2.4) over the given time points $t_1 < t_2 < \cdots < t_n$, we need to determine the length of time during which the process lies in a certain regime and the time for the process to switch from one regime to another. For simplicity, we make the assumption that if $y(t_i)$ and $y(t_{i+1})$, $i = 1, \ldots, n-1$, lie in the same regime, then there is no switching between the regimes during the period $(t_i, t_{i+1})$. But if $y(t_i)$ and $y(t_{i+1})$ lie in different regimes, then we assume that the integral path of $y(t_i)$ crosses the threshold once and only once over the time interval $(t_i, t_{i+1})$; the time of crossing $T_r$ is approximated by a linear interpolation between the observation times as

$$T_r = \frac{\delta_i[r - y(t_i)]}{y(t_{i+1}) - y(t_i)} + t_i, \tag{2.8}$$

where $\delta_i = t_{i+1} - t_i$. Whether the above assumption is valid or not clearly depends on the data at hand. However, this assumption is at least one step closer to reality than the assumption of linearity under which everything always falls within one and only one regime.

There are 4 cases to describe:

*Case 1:* 2 adjacent observations in regime 1 and Equation (2.4) becomes

$$\dot{X}(t) = F_1 X(t) + G_1 w(t), \quad t_i \le t \le t_{i+1}. \tag{2.9}$$

*Case 2:* 2 adjacent observations in regime 2 and Equation (2.4) becomes

$$\dot{X}(t) = F_2 X(t) + G_2 w(t), \quad t_i \le t \le t_{i+1}. \tag{2.10}$$

*Case 3:* $t_i$th observation in regime 1 and $t_{i+1}$th observation in regime 2 and Equation (2.4) becomes

$$\begin{aligned} \dot{X}(t) &= F_1 X(t) + G_1 w(t) \quad \text{if } t_i \le t \le T_r \\ &= F_2 X(t) + G_2 w(t) \quad \text{if } T_r < t \le t_{i+1}. \end{aligned} \tag{2.11}$$

*Case 4:* $t_i$th observation in regime 2 and $t_{i+1}$th observation in regime 1 and Equation (2.4) becomes

$$\begin{aligned} \dot{X}(t) &= F_2 X(t) + G_2 w(t) \quad \text{if } t_i \le t \le T_r \\ &= F_1 X(t) + G_1 w(t) \quad \text{if } T_r < t \le t_{i+1}. \end{aligned} \tag{2.12}$$

In each of the four cases, the solution of $\dot{X}(t)$ over a finite time step $\delta_i$ from time $t_i$ to time $t_{i+1}$ can be written as

$$X(t_{i+1}) = \phi(\delta_i) X(t_i) + g(t_i), \tag{2.13}$$

where

*Case 1:*

$$\phi(\delta_i) = \exp(F_1\delta_i), \quad g(t_i) = \int_0^{\delta_i} \exp[F_1(\delta_i - t)]G_1 w(t)dt; \tag{2.14}$$

*Case 2:*

$$\phi(\delta_i) = \exp(F_2\delta_i), \quad g(t_i) = \int_0^{\delta_i} \exp[F_2(\delta_i - t)]G_2 w(t)dt; \tag{2.15}$$

*Case 3:*

$$\begin{gathered}\phi(\delta_i) = \exp(F_1\delta_{1i})\exp(F_2\delta_{2i}), \quad \delta_{1i} = T_r - t_i, \quad \delta_{2i} = t_{i+1} - T_r, \\ g(t_i) = \exp(F_2\delta_{2i})\int_0^{\delta_{1i}} \exp[F_1(\delta_{1i} - t)]G_1 w(t)dt + \int_0^{\delta_{2i}} \exp[F_2(\delta_{2i} - t)]G_2 w(t)dt;\end{gathered} \tag{2.16}$$

*Case 4:*

$$\begin{gathered}\phi(\delta_i) = \exp(F_2\delta_{1i})\exp(F_1\delta_{2i}), \\ g(t_i) = \exp(F_1\delta_{2i})\int_0^{\delta_{1i}} \exp[F_2(\delta_{1i} - t)]G_2 w(t)dt + \int_0^{\delta_{2i}} \exp[F_1(\delta_{2i} - t)]G_1 w(t)dt.\end{gathered} \tag{2.17}$$

Cases 1 and 2 are obvious. To see Case 3, we use the usual superscripts "+", "−" to denote "from the right" and "from the left" respectively; the solution of Equation (2.11) over a finite time step $\delta_i$ from time $t_i$ to time $t_{i+1}$ is

$$X(T_r^-) = \exp(F_1\delta_{1i})X(t_i^+) + \int_0^{\delta_{1i}} \exp[F_1(\delta_{1i} - t)]G_1 w(t)dt, \tag{2.18}$$

$$X(t_{i+1}^-) = \exp(F_2\delta_{2i})X(T_r^+) + \int_0^{\delta_{2i}} \exp[F_2(\delta_{2i} - t)]G_2 w(t)dt. \tag{2.19}$$

If $X(t_i^-) = X(t_i^+)$, $X(T_r^-) = X(T_r^+)$ and $X(t_{i+1}^-) = X(t_{i+1}^+)$, then

$$\begin{aligned}X(t_{i+1}) &= \exp(F_2\delta_{2i})\Big\{\exp(F_1\delta_{1i})X(t_i) + \int_0^{\delta_{1i}} \exp[F_1(\delta_{1i} - t)]G_1 w(t)dt\Big\} \\ &\quad + \int_0^{\delta_{2i}} \exp[F_2(\delta_{2i} - t)]G_2 w(t)dt \\ &= \exp(F_1\delta_{1i})\exp(F_2\delta_{2i})X(t_i) + \Big\{\exp(F_2\delta_{2i})\int_0^{\delta_{1i}} \exp[F_1(\delta_{1i} - t)]G_1 w(t)dt \\ &\quad + \int_0^{\delta_{2i}} \exp[F_2(\delta_{2i} - t)]G_2 w(t)dt\Big\} \\ &= \phi(\delta_i)X(t_i) + g(t_i),\end{aligned}$$

where $\phi(\delta_i)$ and $g(t_i)$ are as defined in Equation (2.16). The same argument applies to Case 4.

Assuming distinct eigenvalues for $F_1$ and $F_2$, we have

$$F_1 = U_1\Lambda_1 U_1^{-1},$$
$$F_2 = U_2\Lambda_2 U_2^{-1},$$

where $\Lambda_i$ is a diagonal matrix of the eigenvalues, $\lambda_{ik}$, of $F_i$ and $U_i$ is a $p \times p$ matrix, the columns of which are the right eigenvectors of $F_i$, $i = 1,2$; $k = 1,\dots,p$. Since $F_i$ is not symmetric, $\Lambda_i$ and $U_i$ are complex and can be used to evaluate $\phi(\delta_i)$ in each of the four cases as follows:

*Case 1:* $$U_1 e^{\Lambda_1\delta_i}(U_1^{-1}) \tag{2.20}$$

*Case 2:* $$U_2 e^{\Lambda_2\delta_i}(U_2^{-1}) \tag{2.21}$$

*Case 3:* $$(U_1 e^{\Lambda_1\delta_{1i}} U_1^{-1})(U_2 e^{\Lambda_2\delta_{2i}} U_2^{-1}) \tag{2.22}$$

*Case 4:* $$(U_2 e^{\Lambda_2\delta_{1i}} U_2^{-1})(U_1 e^{\Lambda_1\delta_{2i}} U_1^{-1}). \tag{2.23}$$

We write $$Q(\delta_i) = E[g(t_i)g'(t_i)]. \tag{2.24}$$

To discuss the evaluation of $Q(\delta_i)$, we use the overhead symbol "$*$" to denote the complex conjugate transposed matrix and an overhead "$-$" to denote the complex conjugate. For Cases 1 and 2 ($m = 1,2$ respectively),

$$Q(\delta_i) = \int_0^{\delta_i} \exp[F_m(\delta_i - t)]G_m G_m' \exp[F_m'(\delta_i - t)]dt \tag{2.25}$$

$$= \int_0^{\delta_i} U_m \exp[\Lambda_m(\delta_i - t)]U_m^{-1} G_m G_m' (U_m^*)^{-1} \exp[\Lambda_m^*(\delta_i - t)]U_m^* dt$$

$$= U_m\Big\{\int_0^{\delta_i} \exp[\Lambda_m(\delta_i - t)]L_{mm} \exp[\Lambda_m^*(\delta_i - t)]dt\Big\}U_m^*$$

$$\big(\text{where} \quad L_{mm} = U_m^{-1} G_m G_m' (U_m^*)^{-1}\big)$$

$$= U_m Q_{mm}(\delta_i) U_m^*, \tag{2.26}$$

where $Q_{mm}(\delta_i) = \int_0^{\delta_i} \exp[\Lambda_m(\delta_i - t)]L_{mm} \exp[\Lambda_m^*(\delta_i - t)]dt$. The $(j,k)$th element of $Q_{mm}(\delta_i)$ is

$$Q_{mm,jk}(\delta_i) = \begin{cases} \dfrac{L_{mm,jk}[\exp\{(\lambda_{mj} + \bar\lambda_{mk})\delta_i\} - 1]}{\lambda_{mj} + \bar\lambda_{mk}}, & \lambda_{mj} + \bar\lambda_{mk} \neq 0 \\ L_{mm,jk}\delta_i, & \lambda_{mj} + \bar\lambda_{mk} = 0 \end{cases} \tag{2.27}$$

where $L_{mm,jk}$ is the $(j,k)$th element of $L_{mm}$.

For Cases 3 and 4, the covariance matrix of $g(t_i)$ can be expressed as

$$Q(\delta_i) = T_1 + T_2 + T_3 + T_3^*. \tag{2.28}$$

For Case 3, the three terms $T_1, T_2, T_3$ are evaluated as follows:

$$T_1 = E\Big(\exp(F_2\delta_{2i})\Big\{\int_0^{\delta_{1i}}\int_0^{\delta_{1i}}\exp[F_1(\delta_{1i}-s)]G_1w(s)w(t)G_1'\exp[F_1'(\delta_{1i}-t)]dsdt\Big\}$$
$$\exp(F_2'\delta_{2i})\Big) \tag{2.29}$$
$$= \exp(F_2\delta_{2i})\Big\{\int_0^{\delta_{1i}}\exp[F_1(\delta_{1i}-t)]G_1G_1'\exp[F_1'(\delta_{1i}-t)]dt\Big\}\exp(F_2'\delta_{2i}).$$

Note that the middle term is just the same as the expression of $Q(\delta_i)$ in Equation (2.25) except that $\delta_i$ is replaced by $\delta_{1i}$ and $m = 1$ here. So

$$T_1 = \exp(F_2\delta_{2i})[U_1Q_{11}(\delta_{1i})U_1^*]\ \exp(F_2'\delta_{2i}), \tag{2.30}$$

where $Q_{11}(\delta_{1i})$ and the $(j,k)$th element of $Q_{11}(\delta_{1i})$, $Q_{11,jk}(\delta_{1i})$, are as defined in Equations (2.25)–(2.27) with $\delta_i$ replaced by $\delta_{1i}$ and $m = 1$.

$$T_2 = E\Big\{\int_0^{\delta_{2i}}\int_0^{\delta_{2i}}\exp[F_2(\delta_{2i}-s)]G_2w(s)w(t)G_2'\exp[F_2'(\delta_{2i}-t)]dsdt\Big\}$$
$$= \int_0^{\delta_{2i}}\exp[F_2(\delta_{2i}-t)]G_2G_2'\exp[F_2'(\delta_{2i}-t)]dt \tag{2.31}$$

which is just the same as the expression of $Q(\delta_i)$ in Equation (2.25) except that $\delta_i$ is replaced by $\delta_{2i}$ and $m = 2$ here. So

$$T_2 = U_2Q_{22}(\delta_{2i})U_2^*, \tag{2.32}$$

where $Q_{22}(\delta_{2i})$ and the $(j,k)$th element of $Q_{22}(\delta_{2i})$, $Q_{22,jk}(\delta_{2i})$, are as defined in Equations (2.25)–(2.27) with $\delta_i$ replaced by $\delta_{2i}$ and $m = 2$.

$$T_3 = E\Big\{\int_0^{\delta_{2i}}\int_0^{\delta_{1i}}\exp[F_2(\delta_{2i}-s)]G_2w(s)w(t)G_1'\exp[F_1'(\delta_{1i}-t)]\exp(F_2'\delta_{2i})dsdt\Big\}. \tag{2.33}$$

If $\delta_{2i} > \delta_{1i}$,

$$T_3 = \int_0^{\delta_{1i}}\exp[F_2(\delta_{2i}-t)]G_2G_1'\exp[F_1'(\delta_{1i}-t)]\exp(F_2'\delta_{2i})dt$$
$$= \int_0^{\delta_{1i}}U_2\exp[\Lambda_2(\delta_{2i}-t)]U_2^{-1}G_2G_1'(U_1^*)^{-1}\exp[\Lambda_1^*(\delta_{1i}-t)]U_1^*\exp(F_2'\delta_{2i})dt$$
$$= U_2\Big\{\int_0^{\delta_{1i}}\exp[\Lambda_2(\delta_{2i}-t)]L_{21}\exp[\Lambda_1^*(\delta_{1i}-t)]dt\Big\}U_1^*\exp(F_2'\delta_{2i})$$
$$\Big(\text{where}\ \ L_{21} = U_2^{-1}G_2G_1'(U_1^*)^{-1}\Big)$$

$$= U_2 Q_{21}(\delta_{1i}) U_1^* \exp(F_2' \delta_{2i}) \tag{2.34}$$

where $Q_{21}(\delta_{1i}) = \int_0^{\delta_{1i}} \exp[\Lambda_2(\delta_{2i} - t)] L_{21} \exp[\Lambda_1^*(\delta_{1i} - t)]dt$. The $(j,k)$th element of $Q_{21}(\delta_{1i})$ is

$$\begin{aligned}
&Q_{21,jk}(\delta_{1i}) \\
&= L_{21,jk} \int_0^{\delta_{1i}} \exp[\lambda_{2j}(\delta_{2i} - t)] \exp[\bar{\lambda}_{1k}(\delta_{1i} - t)]dt \\
&= L_{21,jk} \exp[\lambda_{2j}\delta_{2i} + \bar{\lambda}_{1k}\delta_{1i}] \int_0^{\delta_{1i}} \exp[-(\lambda_{2j} + \bar{\lambda}_{1k})t]dt \\
&= \begin{cases} \dfrac{L_{21,jk} \exp[\lambda_{2j}\delta_{2i} + \bar{\lambda}_{1k}\delta_{1i}]}{-(\lambda_{2j} + \bar{\lambda}_{1k})} [\exp\{-(\lambda_{2j} + \bar{\lambda}_{1k})\delta_{1i}\} - 1], & \lambda_{2j} + \bar{\lambda}_{1k} \neq 0 \\ L_{21,jk} \exp[\lambda_{2j}\delta_{2i} + \bar{\lambda}_{1k}\delta_{1i}]\delta_{1i}, & \lambda_{2j} + \bar{\lambda}_{1k} = 0 \end{cases}
\end{aligned} \tag{2.35}$$

where $L_{21,jk}$ is the $(j,k)$th element of $L_{21}$.

If $\delta_{2i} < \delta_{1i}$,

$$\begin{aligned}
T_3 &= \int_0^{\delta_{2i}} \exp[F_2(\delta_{2i} - t)] G_2 G_1' \exp[F_1'(\delta_{1i} - t)] \exp(F_2'\delta_{2i})dt \\
&= \int_0^{\delta_{2i}} U_2 \exp[\Lambda_2(\delta_{2i} - t)] U_2^{-1} G_2 G_1' (U_1^*)^{-1} \exp[\Lambda_1^*(\delta_{1i} - t)] U_1^* \exp(F_2'\delta_{2i})dt \\
&= U_2 \Big\{ \int_0^{\delta_{2i}} \exp[\Lambda_2(\delta_{2i} - t)] L_{21} \exp[\Lambda_1^*(\delta_{1i} - t)]dt \Big\} U_1^* \exp(F_2'\delta_{2i}) \\
&\qquad\qquad (\text{where } L_{21} = U_2^{-1} G_2 G_1' (U_1^*)^{-1}) \\
&= U_2 Q_{21}(\delta_{2i}) U_1^* \exp(F_2'\delta_{2i})
\end{aligned} \tag{2.36}$$

where $Q_{21}(\delta_{2i}) = \int_0^{\delta_{2i}} \exp[\Lambda_2(\delta_{2i} - t)] L_{21} \exp[\Lambda_1^*(\delta_{1i} - t)]dt$. The $(j,k)$th element of $Q_{21}(\delta_{2i})$ is

$$\begin{aligned}
&Q_{21,jk}(\delta_{2i}) = L_{21,jk} \int_0^{\delta_{2i}} \exp[\lambda_{2j}(\delta_{2i} - t)] \exp[\bar{\lambda}_{1k}(\delta_{1i} - t)]dt \\
&= L_{21,jk} \exp[\lambda_{2j}\delta_{2i} + \bar{\lambda}_{1k}\delta_{1i}] \int_0^{\delta_{2i}} \exp[-(\lambda_{2j} + \bar{\lambda}_{1k})t]dt \\
&= \begin{cases} \dfrac{L_{21,jk} \exp[\lambda_{2j}\delta_{2i} + \bar{\lambda}_{1k}\delta_{1i}]}{-(\lambda_{2j} + \bar{\lambda}_{1k})} [\exp\{-(\lambda_{2j} + \bar{\lambda}_{1k})\delta_{2i}\} - 1], & \lambda_{2j} + \bar{\lambda}_{1k} \neq 0 \\ L_{21,jk} \exp[\lambda_{2j}\delta_{2i} + \bar{\lambda}_{1k}\delta_{1i}]\delta_{2i}, & \lambda_{2j} + \bar{\lambda}_{1k} = 0 \end{cases}
\end{aligned} \tag{2.37}$$

where $L_{21,jk}$ is the $(j,k)$th element of $L_{21}$.

For Case 4, the three terms are determined similarly and the following is obtained:

$$T_1 = \exp(F_1\delta_{2i})[U_2Q_{22}(\delta_{1i})U_2^*]\exp(F_1'\delta_{2i}) \tag{2.38}$$

where $Q_{22}(\delta_{1i})$ and the $(j,k)$th element of $Q_{22}(\delta_{1i})$, $Q_{22,jk}(\delta_{1i})$, are as defined in Equations (2.25)–(2.27) with $\delta_i$ replaced by $\delta_{1i}$ and $m = 2$.

$$T_2 = U_1Q_{11}(\delta_{2i})U_1^* \tag{2.39}$$

where $Q_{11}(\delta_{2i})$ and the $(j,k)$th element of $Q_{11}(\delta_{2i})$, $Q_{11,jk}(\delta_{2i})$, are as defined in Equations (2.25)–(2.27) with $\delta_i$ replaced by $\delta_{2i}$ and $m = 1$.

$$T_3 = \begin{cases} U_1Q_{12}(\delta_{1i})U_2^*\exp(F_1'\delta_{2i}), & \delta_{2i} > \delta_{1i} \\ U_1Q_{12}(\delta_{2i})U_2^*\exp(F_1'\delta_{2i}), & \delta_{2i} < \delta_{1i} \end{cases} \tag{2.40}$$

where, for any $\delta$, $Q_{12}(\delta) = \int_0^\delta \exp[\Lambda_1(\delta_{2i} - t)]L_{12}\exp[\Lambda_2^*(\delta_{1i} - t)]dt$ and $L_{12} = U_1^{-1}G_1G_2'(U_2^*)^{-1}$. The $(j,k)$th element of $Q_{12}(\delta)$ is

$$Q_{12,jk}(\delta) = \begin{cases} \dfrac{L_{12,jk}\exp[\lambda_{1j}\delta_{2i} + \bar{\lambda}_{2k}\delta_{1i}]}{-(\lambda_{1j} + \bar{\lambda}_{2k})}[\exp\{-(\lambda_{1j} + \bar{\lambda}_{2k})\delta\} - 1], & \lambda_{1j} + \bar{\lambda}_{2k} \neq 0 \\ L_{12,jk}\exp[\lambda_{1j}\delta_{2i} + \bar{\lambda}_{2k}\delta_{1i}]\delta, & \lambda_{1j} + \bar{\lambda}_{2k} = 0 \end{cases} \tag{2.41}$$

where $L_{12,jk}$ is the $(j,k)$th element of $L_{12}$.

Let $X(t_{i+1}|t_i)$ denote the 'estimate' of the state vector at observation time $t_{i+1}$ given observations up to time $t_i$ and $P(t_{i+1}|t_i)$ denote the covariance matrix of the corresponding estimate. Similarly, let $X(t_{i+1}|t_{i+1})$ denote the 'estimate' of the state vector at observation time $t_{i+1}$ given observations up to time $t_{i+1}$ and $P(t_{i+1}|t_{i+1})$ denote the covariance matrix of the corresponding estimate. For each observed $y(t_{i+1})$, $i = 1, \dots, n-1$, the following Kalman filter algorithm is executed to evaluate $-2\ell n$ likelihood.

*Step 1:* Check to which case $y(t_{i+1})$ belongs and calculate the length(s) of the time step $\delta_i$ for Cases 1 and 2 but $\delta_{1i}$, $\delta_{2i}$ for Cases 3 and 4. Then the one-step prediction is $X(t_{i+1}|t_i) = \phi(\delta_i)X(t_i|t_i)$, where $\phi(\delta_i)$ is defined in Equations (2.20)–(2.23) in each of the four cases respectively.

*Step 2:* Calculate the covariance matrix of this prediction $P(t_{i+1}|t_i) = \phi(\delta_i) \cdot P(t_i|t_i)\phi^*(\delta_i) + Q(\delta_i)$, where $Q(\delta_i)$ is defined in Equations (2.24)–(2.41) for the four cases.

*Step 3:* Predict the next observation $y(t_{i+1}|t_i) = HX(t_{i+1}|t_i)$.

*Step 4:* Calculate the innovation $I(t_{i+1}) = y(t_{i+1}) - y(t_{i+1}|t_i)$.

*Step 5:* Calculate the innovation variance $V(t_{i+1}) = HP(t_{i+1}|t_i)H'$.

*Step 6:* Contribution to $-2\ell n$ likeihood is $I^2(t_{i+1})/V(t_{i+1}) + \ell nV(t_{i+1})$.

*Step 7:* Calculate the Kalman gain matrix $K(t_{i+1}) = P(t_{i+1}|t_i)H'/V(t_{i+1})$.

*Step 8:* Update the estimate of the state vector $X(t_{i+1}|t_{i+1}) = X(t_{i+1}|t_i) + K(t_{i+1})I(t_{i+1})$.

*Step 9:* Update the estimate of the state covariance matrix $P(t_{i+1}|t_{i+1}) = P(t_{i+1}|t_i) - K(t_{i+1})HP(t_{i+1}|t_i)$.

Then the NAG subroutine E04JBF is applied to find the minimum of $-2\ell n$ likelihood which gives the maximum likelihood estimates of the parameters. For ease of minimization, we normalize the data and scale the time units as appropriate.

To compare the continuous time linear and threshold autoregressive models of various orders we use the normalized AIC criterion (NAIC):

$$\text{NAIC} = (-2\ell n M_j + 2k_j)/n^*,$$

where $M_j$ denotes the maximum likelihood value for the $j$th model under consideration, $k_j$ is the number of free parameters in the model and $n^*$ denotes the effective number of observations. For continuous time AR($p$) model, the "best" model refers to the one which has the smallest NAIC value. For the continuous time SETAR($2;p,p$) model, which involves the threshold value $r$ as well, we adopt an identification procedure similar to that for the discrete time case. Firstly, we fix the order of autoregression to be fitted for each regime and then compute the NAIC value over a preselected set of threshold values. The choice of this set is arbitrary. In this paper, we use $\{\overline{x} \pm qs, q = 0, \frac{1}{7}, \ldots, 1\}$ where $\overline{x}$ and $s$ denote the sample mean and sample standard deviation of the data respectively. The value of $r$, denoted by $\hat{r}$ which gives the minimum NAIC, will be adopted as our initial estimate of $r$. Then we allow the order to vary and repeat the above procedure of finding the threshold value. Finally the models of various orders with a chosen threshold value are compared and the one which gives the minimum NAIC is adopted. Apart from NAIC, we consider some other methods in choosing the final model, including e.g. the behavior of the parameter estimates in the contour plot of $-2\ell n$ likelihood function. We also use the normalized BIC criterion (NBIC):

$$\text{NBIC} = [-2\ell n M_j + k_j \ell n(n^*)]/n^*,$$

where $M_j$, $k_j$ and $n^*$ are as defined above. We consider the size of $n_1$ and $n_2$ as well in the TAR case.

The final linear and TAR models are then checked for adequacy. As some of the conventional diagnostic checking methods cannot be used for unequally spaced standardized innovations, we consider the following alternatives based on the predictive residuals:

(a) CUSUM and CUSUMSQ plots suggested by Brown, Durbin and Evans (1975) for detecting structural change over time in linear regression models.
(b) CUSUM and reverse CUSUM test statistics suggested by PD (1986) and Petruccelli (1988) for detecting threshold nonlinearity.

If there is no structural change or threshold nonlinearity in the data, the predictive residuals are independent and identically distributed (i.i.d.) normal variables. Then the above plots or test statistics constructed from them will fall within some limits. Similarly if our model is adequate, the standardized innovations are also i.i.d. normal variables. So, if there is any significant deviation of the above plots or test statistics from the limits, we suspect that the model is inadequate. For details, see Yeung (1989).

## 3. Applications

### 3.1. Stock price data

In Tong and Yeung (1988), we have considered three sets of daily closing stock price data: (i) IBM(1) covering the period Dec. 18, 1959 – May 12, 1960; (ii) IBM(2) covering the period May 18, 1961 – March 30, 1962; (iii) Hang Seng Index (HSI) covering the period June 20, 1985 – March 7, 1986. Commonly stock price data are treated as fully observed time series for analysis by ignoring the closing dates. (We retain this usage in later discussions.) As the data contain gaps corresponding to weekends, public holidays or other extraneous factors, they can also be treated as partially observed. The daily percentage relative changes in price $r'_t = (P_t - P_{t-1})/P_{t-1} \times 100$, where $P_t$ denotes the stock price at time $t$, are calculated for the three sets of price series and tested for nonlinearity. Using discrete time autoregressive models with order $p = 1$, we have obtained the following results with our tests for nonlinearity:

(a) IBM(1)—strong linearity, moderate closing-date effects
(b) IBM(2)—strong nonlinearity, almost no closing-date effects
(c) HSI—marginal nonlinearity, strong closing-date effects.

Now the $\{r'_t\}$ series are normalized and fitted by linear and threshold autoregressive models with order one in the following form:

*First-order linear autoregressive model* AR(1)

$$\frac{dy(t)}{dt} + ay(t) = \varepsilon(t), \quad \varepsilon(t) \sim N(0,\sigma^2). \tag{3.1}$$

*First-order threshold autoregressive model* TAR

$$\begin{aligned} \frac{dy(t)}{dt} + a_1 y(t) = \varepsilon_1(t) \quad \text{if} \quad y(t) \le r \\ \frac{dy(t)}{dt} + a_2 y(t) = \varepsilon_2(t) \quad \text{if} \quad y(t) > r, \end{aligned} \tag{3.2}$$

where $\varepsilon_1(t) \sim N(0, \sigma_1^2)$ and $\varepsilon_2(t) \sim N(0, \sigma_2^2)$.

The parameter estimates of the fitted models are given in Table 1 and some conclusions can be drawn:

1. In fitting linear models to IBM(1) data, the parameter estimates for the autoregressive coefficients are insignificant for both fully and partially observed case so that the fitted models correspond to a continuous time random walk. Hence, these results agree with those of the runs test for randomness as reported in Tong and Yeung (1988).
2. Previously, the tests of nonlinearity have suggested that there may be a threshold effect at the value zero for IBM(2) and HSI data such that different linear dynamics dominate according to the direction of price change. From continuous time threshold modelling, the suggested threshold values in non-normalized units for IBM(2) and HSI data are also fairly close to zero. Specifically, the threshold value is negative, around $-0.433$ to $-0.477$ for the former data set but positive, around 0.656 to 0.791 for the latter. Given that our computer package for fitting the TAR models is at present constrained to two regimes, our interpretation of the results must be quite tentative. It is suspected that the IBM firm, which has a large share in the computer market and a long standing reputation, is responsive to the negative rate of change, whereas the speculative Hong Kong stock market is responsive to the profitable positive rate of change. We hope to improve our fitting package so that we can explore the possibility of a three-regime model, with one positive threshold and one negative threshold. This would be quite similar to the model of Granger and Morgenstern (1970) on stock price data with two reflecting barriers due to the operation of limit orders.
3. TAR models for both fully and partially observed IBM(2) data have failed to reduce the residuals to normality ($p$-value of the Kolmogorov-D test statistic $< 0.1$), and perform poorly in respect to the CUSUM plots (see Figure 1). As for the fully observed HSI data, the likelihood function of the fitted TAR model is badly behaved (see Figure 2). It seems that we might need to allow non-normal errors or to consider TAR models with three regimes.
4. In fitting linear models, the parameter estimates for fully and partially observed IBM(2) data are closer to each other as compared with the IBM(1) and HSI series. This also suggests the absence of strong closing date effects

for the IBM(2) data as observed in the tests of nonlinearity.

5. In fitting the TAR models for each set of price data, both fully and partially observed series give similar parameter estimates and threshold values (in normalized units). On the other hand, for the IBM(2) and the HSI data, the fully observed series give a greater NAIC reduction ratio as compared with the linear models; whereas for the IBM(1) data, the partially observed series gives a greater NAIC reduction ratio. Hence, closing dates have not affected the parameter estimates and threshold values. However, they have reduced nonlinearity in nonlinear process such as IBM(2) and HSI data but introduced nonlinearity in linear process such as IBM(1) data.

### 3.2. Beach water quality data

In Hong Kong, water quality data are taken from the selected beaches between one to three times a month, which are not equally spaced. In Tong and Yeung (1990), we carried out tests of nonlinearity on dissolved oxygen content (DO) measured in percentage of saturation for six beaches shown in Figure 3 for the period 1980–1985. The data are divided by 10 and then log transformed to enhance data stability. Using continuous time autoregressive models with order $p = 1$, the test suggested linearity for the heavily polluted Anglers and Butterfly beaches, nonlinearity for the moderately clean Repulse Bay and Shek O beaches and linearity for the relatively clean Cheung Sha and Pui O beaches. In other words, the extents of linearity for the beaches are similar in pairs but vary with the pollution level.

Now, as in the tests of nonlinearity, the unit of time is taken as 10 days and outliers are removed. The data are normalized and then fitted by linear and threshold autoregressive models with order one as in Section 3.1. The parameter estimates of the fitted models are given in Table 2 and some conclusions can be drawn:

1. For the Anglers and the Butterfly beaches, the fitted linear models have failed to reduce the residuals to normality. Though the TAR model for Anglers Bay can reduce the residuals to normality and achieve quite small NAIC and NBIC values, the reverse CUSUM statistic crosses the boundary 7 times at the 10% significance level. The TAR model for the Butterfly beach performs badly by reference to the Kolmogorov-D test statistic, CUSUM statistic and CUSUM plot.
2. For the Repulse Bay and the Shek O beaches, the fitted linear models seem inadequate by reference to the Kolmogorov-D test statistic, the reverse CUSUM statistic and the CUSUMSQ plots. Similar problems apply to the fitted TAR models but the TAR model for Shek O performs slightly

better and has a threshold value of −0.285 (i.e. 2.230 in non-normalized units) which is roughly comparable to the value of −0.143 (i.e. 2.130 in non-normalized units) obtained from the TAR model for the Anglers Bay. Usually, less than 50% saturation level of DO (i.e. 1.609 in our non-normalized units) is considered unsatisfactory and the decomposition mechanism changes from aerobic to anaerobic. The above estimated threshold values, which are smaller than the sample means of the two beach series, tend towards this low level.

3. For the Cheung Sha and the Pui O beaches, both linear and threshold autoregressive models perform badly by reference to the Kolmogorov-D test statistic and CUSUMSQ plots.
4. Overall, the modelling results for the six beaches are not good. The data themselves seem to be very noisy and might be affected by irregular factors such as tidal flow and remedial works carried out by the government to improve the beach water quality, which we have not taken into account. Moreover our data are less than adequate to reflect the known periodicities of diurnal effect (24hr.) in the DO series caused by photosynthesis and respiration of aquatic plants. As a result, the assumptions about the number of switchings and the times of switching in threshold modelling are probably not appropriate.

### 3.3. Lynx data

This data set (1821–1934) has usually been treated as fully observed for tests of nonlinearity and model fitting. However, the data for the years 1892–1896 and 1914 were apparently obtained from private records kept by some of the older fur trade factories. Therefore, if we treat the data for these six years as missing, the lynx data set becomes partially observed. In Yeung (1989), we consider four cases for the logged lynx data: fully observed and then partially observed for the whole period 1821–1934 and the early period 1821–1910. In each case, the data have been tested for nonlinearity using discrete and continuous time models respectively with order $p = 2$; and nonlinearity has been detected.

Now, the data in each case are normalized and fitted by linear and threshold autoregressive models with order up to two in the following forms:

*Second-order linear autoregressive model* AR(2)

$$y^{(2)}(t) + a_1 y^{(1)}(t) + a_0 y(t) = \varepsilon(t), \quad \varepsilon(t) \sim N(0, \sigma^2). \tag{3.3}$$

*Second-order threshold autoregressive model* TAR

$$\begin{aligned} y^{(2)}(t) + a_{11} y^{(1)}(t) + a_{10} y(t) &= \varepsilon_1(t) \quad \text{if} \quad y(t) \le r \\ y^{(2)}(t) + a_{21} y^{(1)}(t) + a_{20} y(t) &= \varepsilon_2(t) \quad \text{if} \quad y(t) > r, \end{aligned} \tag{3.4}$$

where $\varepsilon_1(t) \sim N(0,\sigma_1^2)$ and $\varepsilon_2(t) \sim N(0,\sigma_2^2)$.

The parameter estimates of the fitted models are given in Table 3 and some conclusions can be drawn:

1. For all four cases, the fitted linear and threshold autoregressive models are satisfactory, based on the diagnostic checking statistics and various plots. Furthermore, TAR models have a smaller NAIC but a larger NBIC value than the linear models.
2. Similar parameter estimates are obtained irrespective of whether the data for the whole period or early period are used for fitting both linear and threshold autoregressive models. However, in terms of NAIC and NBIC values, the performance of TAR models is slightly better if the data for the whole period are used.
3. All fitted TAR models suggest a threshold value at about 3 which is comparable with the previous results reported in Tong and Lim (1980) and Tong (1983a).
4. The missing observations have not caused much difference in the parameter estimates. Based on NBIC values, the missing observations in the early period case seem to mask nonlinearity. However, based on NAIC values, the missing observations in the above case seem to induce further nonlinearity.

## 4. Concluding Remarks and Unsolved Problems

We have fitted continuous time threshold autoregressive models to three unequally spaced data sets and found that these models can in some cases improve the goodness of fit over linear models. In Yeung (1989), we have also fitted discrete time models to the "fully observed" case of stock price series and lynx data and obtained similar results as in the continuous time case. The main contributions of using continuous time models are (i) they are more natural for modelling continuous time processes and (ii) they are particularly useful if the processes are irregularly observed. Provided that the data are of reasonable quality and the sampling rate is comparatively short relative to the frequency of oscillations of the underlying continuous time realisation, continuous time TAR models fitted to unequally spaced discrete time data can provide useful interpretations. Since we have not included a delay term in the model, least squares prediction may be obtained by the standard method as described in Liptser and Shiryayev (1978, Chapter 12).

At present, the application of continuous time TAR models is somewhat limited due to the unknown switching conditions. To overcome this problem, we have suggested the switching assumption described after Equation (2.7). We have tried previously to approximate the time of crossing in Equation (2.8) by a weighted average of the observation times. The results were not satisfactory. To explore the various possible assumptions we are planning to carry out analogue simulation of some continuous time processes. Furthermore, we are currently studying the effect of including a model error term in Equation (2.13) to account for the switching assumption. We may then apply the error analysis methods commonly described in the engineering and optimal control literature (see, e.g., Anderson and Moore (1979), pp. 130–132) to study the sensitivity of the switching assumption to model fitting.

## Acknowledgment

I. Y. wishes to thank the Croucher Foundation for financial support for her Ph.D. study at the University of Kent at Canterbury.

Table 1. Parameter estimates of the fitted models to stock price data with standard errors in parentheses

(a) Linear Model

| | Fully observed | | Partially observed | |
|---|---|---|---|---|
| | $a$ | $\sigma$ | $a$ | $\sigma$ |
| IBM(1) | 5.539<br>(7.522) | 3.314<br>(2.422) | 9.046<br>(19.163) | 4.237<br>(5.113) |
| IBM(2) | 1.561<br>(0.235) | 1.764<br>(0.157) | 1.684<br>(0.396) | 1.834<br>(0.213) |
| HSI | 2.601<br>(0.819) | 2.255<br>(0.380) | 3.282<br>(1.995) | 2.525<br>(0.810) |

(b) TAR Model

| | Fully observed | | | | | Partially observed | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $a_1$ | $\sigma_1$ | $a_2$ | $\sigma_2$ | $r^*$ | $a_1$ | $\sigma_1$ | $a_2$ | $\sigma_2$ | $r^*$ |
| IBM(1) | 10.234<br>(4.417) | 3.783<br>(0.843) | —<br>— | 0.900<br>(0.211) | 0.571<br>(0.599) | 11.058<br>(6.769) | 4.014<br>(1.193) | —<br>— | 0.589<br>(0.136) | 0.715<br>(0.848) |
| IBM(2) | —<br>— | 0.662<br>(0.091) | 10.941<br>(4.082) | 4.169<br>(0.849) | −0.571<br>(−0.477) | —<br>— | 0.475<br>(0.086) | 8.269<br>(2.301) | 3.710<br>(0.546) | −0.571<br>(−0.433) |
| HSI | 8.271<br>(2.510) | 3.735<br>(0.597) | —<br>— | 0.482<br>(0.089) | 0.571<br>(0.656) | 6.497<br>(2.026) | 3.275<br>(0.522) | —<br>— | 0.579<br>(0.161) | 0.715<br>(0.791) |

* the threshold values are given in normalized units (with non-normalized units enclosed in parentheses).

Table 2. Parameter estimates of the fitted models to beach water quality data with standard errors in parentheses

| | (a) Linear Model | | (b) TAR Model | | | | |
|---|---|---|---|---|---|---|---|
| | $a$ | $\sigma$ | $a_1$ | $\sigma_1$ | $a_2$ | $\sigma_2$ | $r^*$ |
| Anglers (mean=2.15) | 3.054 (1.189) | 2.433 (0.464) | 21.025 (13.622) | 7.669 (2.451) | 0.495 (0.108) | 0.687 (0.064) | −0.143 (2.130) |
| Butterfly (mean=2.25) | 2.425 (0.644) | 2.199 (0.302) | 8.694 (3.868) | 3.653 (0.912) | — — | 0.867 (0.206) | 0.715 (2.339) |
| Repulse (mean=2.27) | 37.361 (49.822) | 8.540 (6.411) | 16.295 (9.541) | 4.897 (1.528) | — — | 1.042 (0.271) | 0.715 (2.346) |
| Shek O (mean=2.26) | 2.754 (0.811) | 2.286 (0.352) | 24.456 (10.734) | 8.228 (1.835) | 0.466 (0.132) | 0.826 (0.075) | −0.285 (2.230) |
| Cheung Sha (mean=2.29) | 2.411 (0.730) | 2.203 (0.371) | — — | 0.943 (0.187) | 13.310 (8.019) | 4.390 (1.462) | −0.715 (2.230) |
| Pui O (mean=2.29) | 2.198 (0.686) | 2.100 (0.363) | — — | 0.250 (0.087) | 15.804 (8.971) | 5.198 (1.575) | −0.571 (2.233) |

* the threshold values are given in normalized units (with non-normalized units enclosed in parentheses).

Table 3. Parameter estimates of the fitted models to lynx data with standard errors in parentheses
(a) Linear Model

| | Fully observed | | | Partially observed | | |
|---|---|---|---|---|---|---|
| | $a_1$ | $a_0$ | $\sigma$ | $a_1$ | $a_0$ | $\sigma$ |
| early period (1821–1920) | 0.482 (0.156) | 0.447 (0.085) | 0.740 (0.078) | 0.553 (0.171) | 0.442 (0.093) | 0.774 (0.088) |
| whole period (1821–1934) | 0.491 (0.142) | 0.433 (0.078) | 0.736 (0.072) | 0.552 (0.155) | 0.427 (0.085) | 0.764 (0.079) |

* the threshold values are given in normalized units (with non-normalized units enclosed in parentheses).

(b) TAR Model

— — — Fully observed

| | $a_{11}$ | $a_{10}$ | $\sigma_1$ | $a_{21}$ | $a_{20}$ | $\sigma_2$ | $r^*$ |
|---|---|---|---|---|---|---|---|
| early period (1821–1920) | 0.334 (0.157) | 0.547 (0.111) | 0.706 (0.083) | 1.782 (0.884) | 0.259 (0.146) | 0.852 (0.280) | 0.857 (3.374) |
| whole period (1821–1934) | 0.354 (0.146) | 0.521 (0.104) | 0.707 (0.077) | 1.877 (0.860) | 0.247 (0.144) | 0.870 (0.272) | 0.857 (3.382) |

— — — Partially observed

| | $a_{11}$ | $a_{10}$ | $\sigma_1$ | $a_{21}$ | $a_{20}$ | $\sigma_2$ | $r^*$ |
|---|---|---|---|---|---|---|---|
| early period (1821–1920) | 0.407 (0.172) | 0.579 (0.142) | 0.745 (0.091) | 2.005 (1.130) | 0.234 (0.157) | 0.914 (0.361) | 0.857 (3.361) |
| whole period (1821–1934) | 0.414 (0.155) | 0.539 (0.129) | 0.739 (0.083) | 2.017 (1.071) | 0.217 (0.150) | 0.902 (0.339) | 0.857 (3.372) |

* the threshold values are given in normalized units (with non-normalized units enclosed in parentheses).

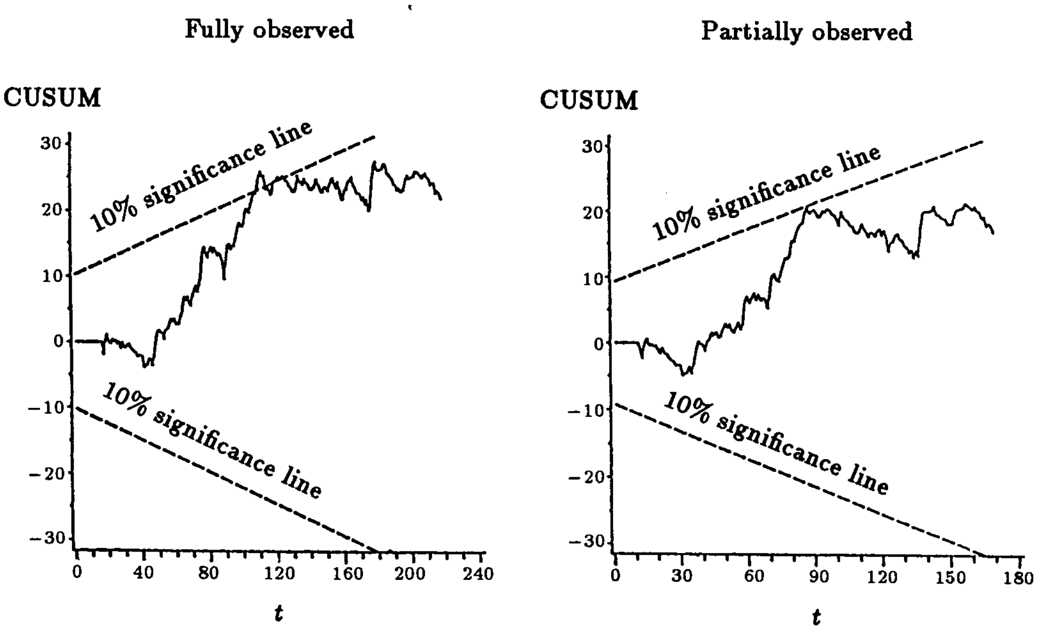


Figure 1. CUSUM plots of the TAR models fitted to IBM(2) data. The 10% significance lines are given by $C = \pm[0.85\sqrt{n-k} + 1.70(t-k)/\sqrt{n-k}]$ where $n = 218$, $k = 16$ for fully observed case and $n = 169$, $k = 11$ for partially observed case.

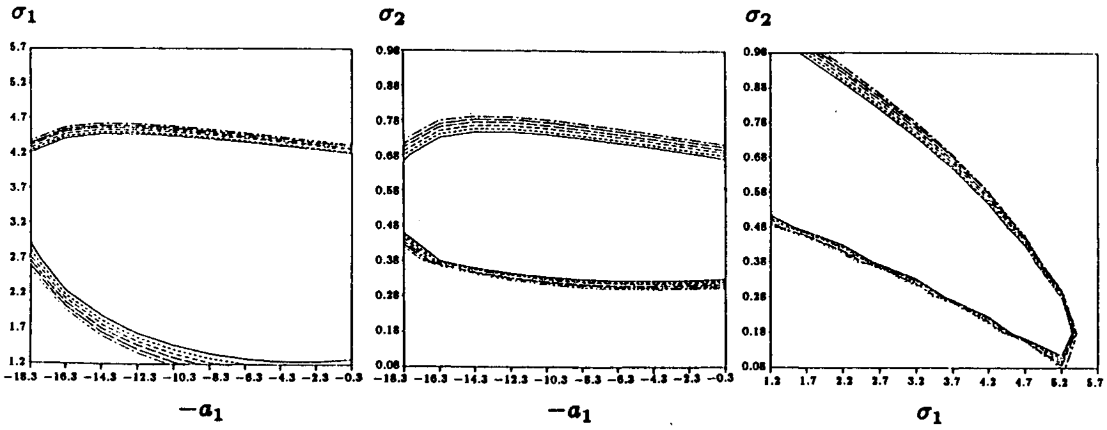


Figure 2. Contour plot of $-2\ell n$ likelihood of the TAR models fitted to fully observed HSI data. The broad contours around the maximum likelihood estimates of the parameters indicate that the model is inadequate.

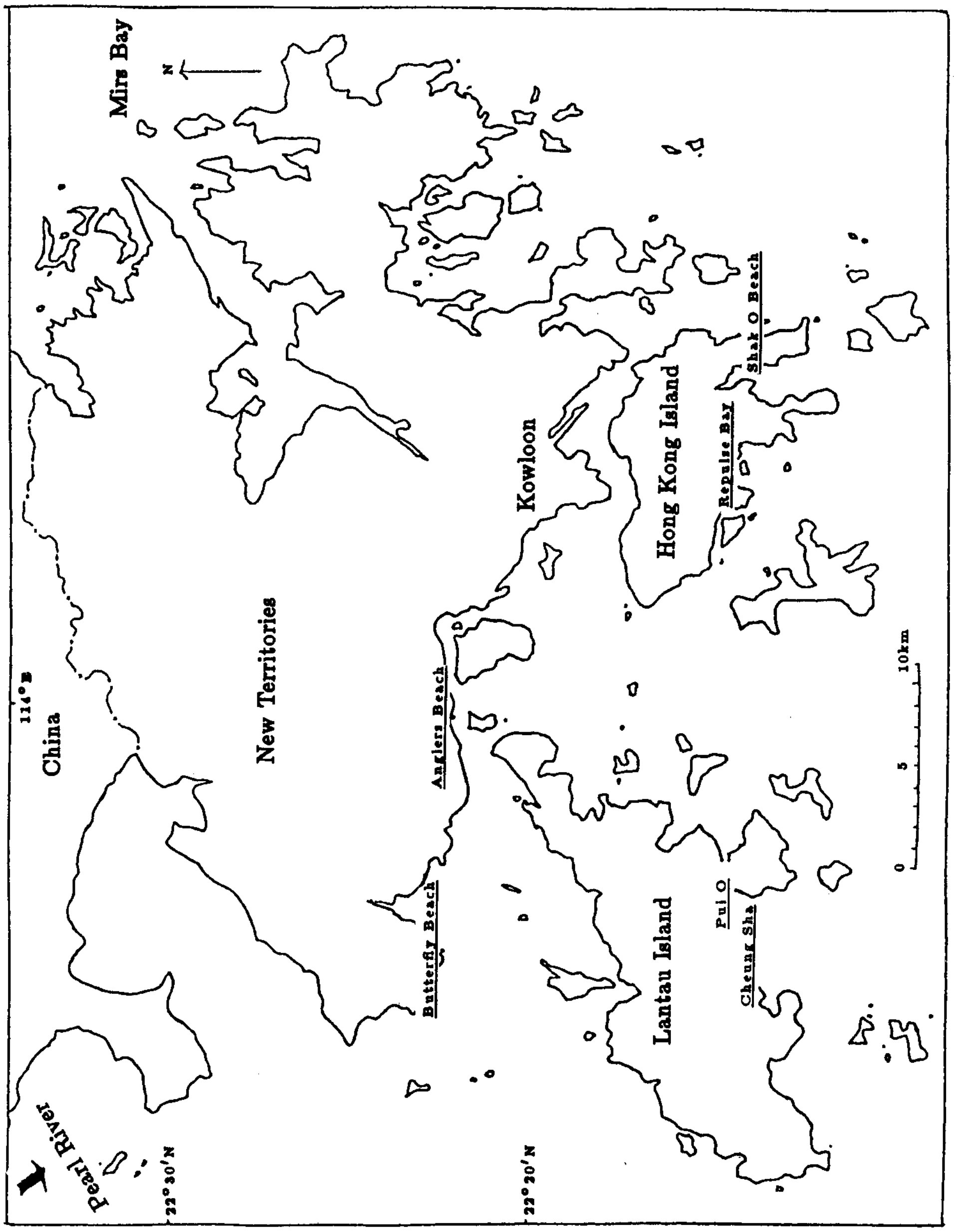


Figure 3. Location of study beaches in Hong Kong (The names of the beaches are underlined)

## References


Anderson, B. D. A. and Moore, J. B. (1979). *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, NJ.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationship over time (with discussion). *J. Roy. Statist. Soc. Ser.B* **37**, 149–192.

Granger, C. W. J. and Morgenstern, O. (1970). *Predictability of Stock Market Prices.* Massachusetts, Heath-Lexington.

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

Liptser, R. S. and Shiryayev, A. N. (1978). *Statistics of Random Processes II.* Springer-Verlag, New York.

Petruccelli, J. D. and Davies, N. (1986). A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika* **73**, 687–694.

Petruccelli, J. D. (1988). A comparison of tests for SETAR-type nonlinearity in time series. Technical Report, Worcester Polytechnic Institute.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (Edited by C. H. Chen). Sijthoff and Noordhoff, Amsterdam.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc. Ser.B* **42**, 245–292.

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

Tong, H. (1983b). A note on a delayed autoregressive process in continuous time. *Biometrika* **70**, 710–712.

Tong, H. and Yeung, I. (1990). On tests for threshold type nonlinearity in irregularly spaced time series. *J. Statist. Comput. Simulation* **34**, 177–194.

Tong, H. and Yeung, I. (1991). On tests for SETAR-type nonlinearity in partially observed time series. *Applied Statistics* **40**, 43–62.

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

Yeung, I. (1989). Continuous time threshold autoregressive models. Unpublished Ph.D. thesis, University of Kent at Canterbury, U.K.



Mathematical Institute, University of Kent at Canterbury, Canterbury CT2 7NF, U.K.
Department of Applied Mathematics, City Polytechnic of Hong Kong, Kowloon, H.K.


(Received November 1989; accepted January 1991)