# Threshold models in time series analysis — 30 years on

HOWELL TONG

> *Re-visiting the past can lead to new discoveries.*
> – Confucius (551 B.C.–479 B.C.)

This paper is a selective review of the development of the threshold model in time series analysis over the past 30 years or so. First, the review re-visits the motivation of the model. Next, it describes the various expressions of the model, highlighting the underlying principle and the main probabilistic and statistical properties. Finally, after listing some of the recent offsprings of the threshold model, the review finishes with some on-going research in the context of threshold volatility.

KEYWORDS AND PHRASES: Bi-modality, Canadian lynx, Chaos, Conditional least squares, Drift criterion, Heteroscedasticity, Ecology, Ergodicity, Indicator time series, Invertibility, Limit cycles, Markov chain, Markov switching, Nonlinear oscillations, Piecewise linearity, River-flow, Seiche record, Skeleton, Stability, Test for linearity, Threshold models, Threshold principle, Volatility, Yule's pea-shooter.

## 1. INTRODUCTION

On the 19th March 1980, the paper entitled *Threshold Autoregression, Limit Cycles and Cyclical Data* was read to the Royal Statistical Society at a meeting organized by the Research Section of the Society and chaired by Professor Peter Whittle. It presented the first comprehensive account of the research that I had been conducting on nonlinear time series analysis till that date, sometimes assisted by my research students on the numerical simulation. Before this paper, I had already presented rudiments of the approach based on piecewise linearization, with references given in the above paper.

In the Confucian spirit, let me start by recalling some of the background of the approach. To avoid repetition of Tong (2007a), I shall now reduce reminiscences but increase technical details. Nevertheless, I shall try to maintain a relaxed style. To keep the paper within reasonable length, some of the topics will have to be treated sparingly or not at all.

Cycles have been among the key objects of attention since the beginning of time series analysis. Of course, models for periodic oscillation have a much longer history in science. School physics has taught many that possibly the simplest model that can generate periodic oscillations is Simple Harmonic Motion (SHM):

$$\frac{d^2x(t)}{dt^2} = -\omega^2 x(t),$$

where $\omega$ is a real constant. This is an elementary *linear* differential equation. It admits the solution

$$x(t) = A\cos(\omega t) + B\sin(\omega t),$$

where the arbitrary constants $A$ and $B$ are fixed by the initial condition, e.g. $x(t)$ and $dx(t)/dt$ at $t = 0$. However, not so well-known is the fact that this highly idealized model is not practically useful, because different periodic oscillations result from different initial values. Thus, SHM does not produce a period that is robust against small disturbances to the initial value. Further, SHM assumes the absence of friction. In the presence of the latter, SHM is usually modified by including a linear function of the velocity. In that case, a stable solution is a damped sinusoid and will tend to a constant, the steady state, as $t$ tends to infinity. Evidence suggests that the British statistician, George Udny Yule, was well acquainted with the above physics and mathematics in his search for a statistical model for the sunspot cycles, although he worked in discrete time. Therefore to generate sustained (quasi-)periodic oscillations, he introduced in Yule (1927) random disturbances (now more commonly called the innovations) leading to the celebrated linear autoregressive model, or the AR model for short. The associated physical set-up has been affectionately referred to as Yule's (heavy) pendulum bombarded by some pea-shooters. Curiously, I have not been able to locate the pea-shooters in Yule (1927); the earliest reference that I can trace is Jeffreys (1940).

As I have said elsewhere (Tong, 1990, p. 20–21), it was perhaps not surprising that Yule chose a linear model because it was unlikely that he could be aware of nonlinear oscillations. After all, the theory of non-linear oscillations was still at its early stages of development around his time. In fact, it was only in 1925 that A. A. Andronov, the prominent Russian oscillation theorist, recognized the role of Poincaré's geometric notion of *limit cycle*. This important notion relates to the existence of sustained periodic oscillations *independent of initial conditions*. To exhibit a simple example

of existence, let me first recall the simplest first order linear differential equation:

$$\text{(1)} \qquad \frac{dx(t)}{dt} + \alpha x(t) = k,$$

where $\alpha$ and $k$ are positive real constants. Trivially, the solution is $x(t) = Ce^{-\alpha t} + k/\alpha$, which tends to $k/\alpha$, a constant, as $t$ tends to $\infty$. Similarly, the same equation with $k = 0$ has the solution $x(t) = Ce^{-\alpha t}$, which tends to zero as $t$ tends to $\infty$. Thus, each of the above stable equations admits the steady state solution of a constant, which is a general property of all stable linear differential equations. (The constant is called a limit point, which represents a state of *static equilibrium*.) Let me next consider a nonlinear differential equation of the following piecewise linear form

$$\text{(2)} \qquad \left(\frac{d}{dt} + \alpha\right)x(t) = \left\{ \begin{array}{ll} 0 & \text{if } x(t) \geq h, \\ k & \text{if } x(t) < h. \end{array} \right.$$

Here, $k, h$ and $\alpha$ are positive real constants with $h > k/\alpha$. By integrating the differential equation piece by piece, it is not difficult to show that the solution of the piecewise linear differential equation will quickly converge to a periodic function of a curved saw-tooth shape with maximum value $h + (k/\alpha)$, which decays to the minimum value $h$, exponentially fast with rate $\alpha$, over a period of duration $-\frac{1}{\alpha}\ln(1 - \frac{k}{\alpha h})$ and independent of the initial value $x(0)$. In other words, the steady state is a limit cycle. Note that the steady state for the dynamics inside each regime (i.e. above or below $h$) is a limit point; yet by dividing the state space into two regimes, each regime being governed by different simple linear dynamics, a new steady state of a fundamentally different character can be created. This is the magic of nonlinearity! For further exposition of piecewise linear differential equations, see Andronov and Khaikin (1937). The essential point is that a limit cycle can be generated by a dynamical system only if there is a balance between energy absorption and energy dissipation. Specifically, the supply of energy $k$ in the lower regime balances the exponential energy dissipation in the upper regime; this balance of energy results in a steady periodic oscillation after some initial transients. The system is then in a state of *dynamic equilibrium*.

The latter part of the last century witnessed the rapid development of an even more exotic form of stable oscillation called chaos. Chaos manifests itself in a time series that is almost undistinguishable from a stochastic process. The apparent randomness is caused by sensitivity of the dynamical system to initial conditions. Of course, limit cycles and chaos also exist in nonlinear difference equations; piecewise linear difference equations are especially important as pseudo-random number generators, which are actually chaos generators. However, the analysis as well as the topology is more complicated. See, e.g., Tong (1990) and Chan and Tong (2001).
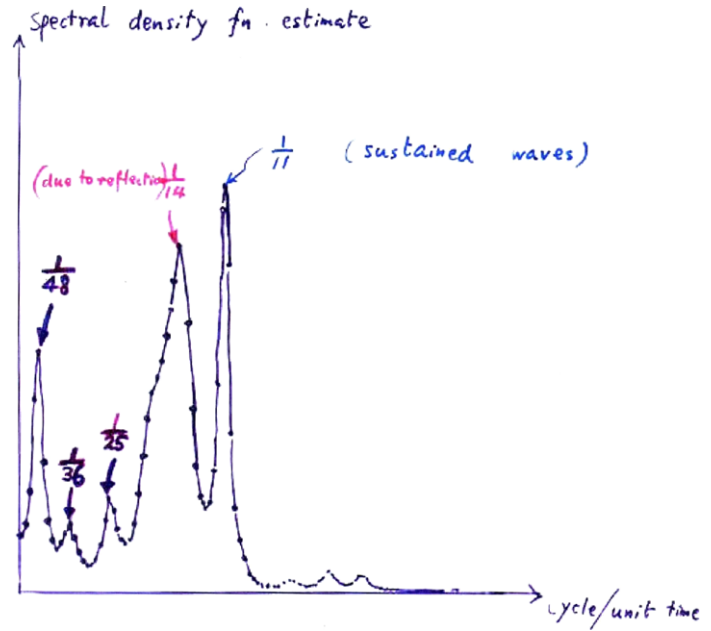
Figure 1. Estimated spectral density function via AR modelling.

## 2. LIMITATIONS OF LINEAR TIME SERIES MODELS: EXAMPLES

**Example 2.1** (Seiche Record at Island Bay, Wellington, New Zealand)**.** Whittle (1954) analyzed the seiche time series of 660 observations at 15 second intervals of the water level in a rock channel at Island Bay on the Wellington coast in his native country, New Zealand.

In Fig. 1, I have reproduced the spectral density function estimate obtained by Akaike (private communication) using a linear $AR(k)$ model with $k$ determined by AIC. (Whittle's original estimate is a smoothed periodogram using the Bartlett window.) Whittle noted a significant arithmetical relationship among the periods of the prominent peaks: $25 = 1 \times 11 + 14; 36 = 2 \times 11 + 14; 48 \approx 3 \times 11 + 14$. Such a relationship is beyond the scope of linear models. In fact, he pointed out that this relationship is due to nonlinearity and cited none other than Andronov and Chaikin (1937). He went on to give some detailed analysis based on a piecewise linear differential equation, of which (2) is the simplest special case. As far as I know, Whittle (1954) was the earliest recognition of the importance of the threshold idea in time series modelling.

**Example 2.2** (Jokulsa River System, Iceland)**.** Tong et al. (1985) studied the Jokulsa river system, consisting of three time series in 1972: river-flow, precipitation and temperature. Figure 2 gives a nonparametric regression of river-flow on temperature. It shows clearly the effect of the melting of glaciers in the catchment area of Jukulsa River on the latter's flow. The nonlinearity is a result of the phase change
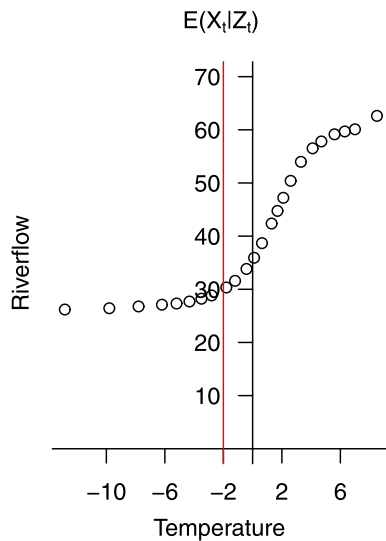
Figure 2. Nonparametric regression of Jokulsa river-flow on temperature.



Figure 4. Histogram of logarithmically transformed lynx data.

from ice to water. The inadequacy of linear models is self-evident in this case.

**Example 2.3** (MacKenzie River Lynx Data). In the early 1950s, the Australian statistician, Pat Moran, spent many of his working hours at the library of the Department of Zoology, Oxford, which became his office. As a result, he became interested in ecology and met the Oxford ecologist, Charles Elton. In particular, he was interested in the famous 10-year lynx cycle, which was and still is of immense interest to the ecologists. In Moran (1953a), among the many available annual records of lynx trappings, he chose the longest one, namely the 1821–1934 record of the MacKenzie River district in Canada. He remarked on the asymmetry of the lynx cycle. The asymmetry can be seen very clearly if the vertical scale is appropriately chosen as in Fig. 3.

He fitted the following linear AR model to the data:

$$X_t = 1.4101X_{t-1} - 0.7734X_{t-2} + \varepsilon_t,$$

where $X_t = \log_{10}$(Number of lynx trapped in year $t$) $-$ 2.9036. In his diagnostic checking, he noticed the '*curious feature*' that the sum of squares of the residuals corresponding to $X_t$ values above zero is 1.781 while that corresponding to $X_t$ values below zero is 4.007. He considered the difference significant. It seems to me that this was one of the
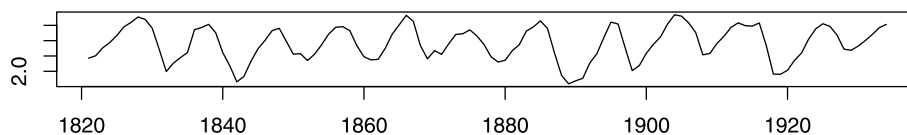
earliest observations of conditional heteroscedasticity in real time series analysis. In fact, he remarked that lynx dynamics '*would have to be represented by nonlinear equations.*' (Moran, 1953b, p. 292). I have presented several significant signatures of non-normality and nonlinearity of the lynx dynamics in Tong (1990). For example, a clear signature of non-normality is the bi-modality of the marginal distribution (Fig. 4) of the data, which is present even after the logarithmic transformation.

In fact, the table below shows that the test for multi-modality due to Chan and Tong (2004), which extends Silverman's test to dependent data, lends support to bi-modality.

| No. of modes | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| P-values | 0.03 | 0.45 | 0.72 | 0.67 |

The shortcomings highlighted in the above examples and elsewhere (e.g. Tong, 1990) pressed home the case for non-linear time series modelling with irresistible force in the late 1970s and early 1980s. The force has remained to the present time.

## 3. THE THRESHOLD MODEL

There are many different but equivalent ways to express a threshold model, each having its advantages, depending on the context and purpose. Let me start with a general form of a threshold autoregressive model or a TAR model



Figure 3. Annual trappings of lynx in the MacKenzie River district on $\log_{10}$ scale.

for short:

$$(3) \qquad X_t = a_0^{(J_t)} + \sum_{i=1}^{p} a_i^{(J_t)} X_{t-i} + b^{(J_t)} \varepsilon_t,$$

where $\varepsilon_t$s are iid $(0, \sigma^2)$ and $\{J_t\}$ is an (indicator) time series taking values in $\{1, 2, \ldots, J\}$. The indicator time series acts as the *switching* mechanism. Note that the TAR model can be easily extended to a threshold autoregressive moving average model (TARMA model for short) by replacing $b^{(J_t)} \varepsilon_t$ by $\sum_{j=0}^{q} b_j^{(J_t)} \varepsilon_{t-j}$. Further extension to include some exogenous time series is obvious and may be assigned the acronym TARMAX.

The basic idea of a threshold model is piecewise linearization through the introduction of the indicator time series, $\{J_t\}$. I have called this idea the *Threshold Principle* for easy reference. Numerous special cases are immediate and I shall describe some of them later. First, I should, however, stress that a special case is just that and should not be confused with the general case. It is unfortunate that perhaps because of its popularity, the self-exciting threshold autoregressive model (to be described later) has sometimes been presented as if it constitutes the entire family of TAR models, which it does not.

To simplify the description, I shall usually exhibit only the simplest/low order cases, namely $p = 1$ or $p = 2$, noting that the case with higher $p$ can be described similarly. I shall also minimize the use of super-scripts. Clearly, the time series $\{J_t\}$ could be either observable or hidden. In fact, it could also be a mixture of the two. Let me start with the case with observable $J_t$.

(O1) *SETAR model*: Suppose that $J = 2$. Let $J_t = 1$ if $X_{t-d} \le r$ and $J_t = 2$ if $X_{t-d} > r$ for some real threshold $r$ and some positive integer $d$, called the delay parameter. The acronym stands for self-exciting TAR model. This SETAR model can also be expressed in the following equivalent form:

$$(4) \qquad X_t = \begin{cases} \alpha + \beta X_{t-1} + \varepsilon_t & \text{if } X_{t-d} \le r; \\ \gamma + \delta X_{t-1} + \phi \varepsilon_t & \text{if } X_{t-d} > r, \end{cases}$$

where $\alpha, \beta, \gamma, \delta$, and $\phi$ are real constants.

(O2) *TARX model*: This is a special case of the TARMAX model, without the moving average component. Let $J_t$ be a measurable function of one or more exogenous and observable time series. Previously, Tong and Lim (1980) coined the acronym TARSO to stand for a TAR system (open loop); they were thinking in terms of control systems.

In the following models, $\{J_t\}$ is *independent of* $\{X_t\}$ *and not observed*.

(H1) *EAR model*: Let $\{J_t\}$ be a sequence of independent and identically distributed random variables such that $J_t =$ 1 with probability $1 - \alpha$, $J_t = 2$ with probability $\alpha$, and $(0 < \alpha < 1)$. Consider the TAR model with $a_0^{(J_t)} \equiv 0$ and $b^{(J_t)} \equiv 1$,

$$(5) \qquad X_t = a_1^{(J_t)} X_{t-1} + a_2^{(J_t)} X_{t-2} + \varepsilon_t,$$

where $0 \le a_1^{(1)} < 1, a_1^{(2)} = 0, a_2^{(1)} = 0$ and $0 \le a_2^{(2)} < 1$. This special case of the TAR model is an example of the EAR model proposed and first studied by Lawrance and Lewis (1980), who were interested in identifying appropriate distributions for $\varepsilon_t$ so that $X_t$ has a (negative) exponential (marginal) distribution. The acronym stands for exponential AR. The connection between the EAR model and the TAR model was first noted by Tong (1983, p. 63) and later exploited by Chan (1986, Ch. 4; 1988) to extend the EAR model to higher orders.

(H2) *Markov Chain Driven TAR model or Hidden Markov Switching model*: Let $\{J_t\}$ be a finite state Markov chain. This model was first proposed by Tong and Lim (1980, p. 285 line -12). It includes (H1) as a special case. Tyssedal and Tjøstheim (1988) was probably the first paper that studied this model in some depth, including an application to stock prices. Later Hamilton (1989) introduced the model, which he called the Markov switching model, to the econometric community. He did not refer to Tyssedal and Tjøstheim (1988) but referred to the Markov chain driven TAR model of Tong (1983, p. 62). However, he overlooked the fact that the Markov chain for the above $\{J_t\}$ was, as in a general TAR model, allowed to be hidden and was hidden. In fact, the overlooking could have been avoided if he had noticed Tong (1983, p. 63) or Chan (1986, 1988), to which I have referred in (H1), or Tong (1983, 276–277), to which I shall refer in model (P1).

The case with $\{J_t\}$ partially hidden has been explored only partially.

(P1) Suppose that $J = 2$ and $J_t$ is a measurable function of $X_{t-d}$ and $U_t$, where $\{U_t\}$ is a hidden (strict) white noise process, independent of $\{X_t\}$ and has distribution $F_u$, so that for some delay $d$, $Prob[J_t = 2 | x_s, s < t] = F_u(x_{t-d})$, and $Prob[J_t = 1 | x_s, s < t] = 1 - F_u(x_{t-d})$. Also, conditional on $\{X_t\}$, $\{J_t\}$ is a sequence of independent random variables. I formerly called a model of this form a *fuzzy extension* of the SETAR model in Tong (1983, p. 276–277). I shall return to a closely related extension in the next paragraph. I think it is worthwhile to explore other measurable functions for $J_t$, in which, e.g., the hidden $\{U_t\}$ can be more general than a white noise process. The resulting models can be perhaps better called *partially-hidden switching models*; the switching can be Markovian or non-Markovian. The stochastic unit root model of Gourieroux and Robert (2006), the TD-SAR model of Wu and Chen (2007) and the ACR model of Bec, Rahbek and Shephard (2008) are examples of recent development in this direction.

Model ([4](navigation)) can also be written in the equivalent form

$$(6) \quad X_t = (\alpha + \beta X_{t-1} + \varepsilon_t)I(X_{t-d} \leq r)$$
$$+ (\gamma + \delta X_{t-1} + \phi \varepsilon_t)(1 - I(X_{t-d} \leq r)),$$

where $I(A)$ is an indicator function such that $I(A) = 1$ if $A$ occurs and $I(A) = 0$ otherwise. In some applications, it might be more appropriate to replace the 'hard thresholding' due to $I$ by some 'soft thresholding'. One way to effect the latter is to replace the step function $I(X_{t-d} \leq r)$ by a sufficiently smooth function (Chan and Tong, 1986a, esp. p. 187), e.g. a continuous cumulative distribution function, say $F(\{X_{t-d} - r\}/\sigma)$; the latter includes the former as a limiting case upon taking $\sigma$ to zero. Chan and Tong (1986a) were the first to propose and develop these soft thresholding models under the name of *smooth threshold autoregressive* (STAR) models. They have given quite a comprehensive account of the probabilistic structure and statistical inference of these models. The STAR models have attracted many followings in econometrics, e.g., Franses and van Dijk (2000). Note that econometricians have changed the label 'threshold' to 'transition'.

Clearly, the right-hand side of ([6](navigation)), with $\varepsilon_t$ suppressed, is discontinuous. By judiciously constraining the parameters, Chan and Tsay (1998) have studied a *continuous SETAR model* of a form first mentioned in Tong (1983, p. 276). More recently, Xia and Tong (2010) have considered, among other nonlinear state-space models, an innovation free SETAR model, which is observed with observation error:

$$(7) \quad X_t = (\alpha + \beta X_{t-1})I(X_{t-d} \leq r)$$
$$+ (\gamma + \delta X_{t-1})(1 - I(X_{t-d} \leq r)),$$

and $Y_t = X_t + \varepsilon_t$, where $\{\varepsilon_t\}$ is a white noise process independent of the unobservable $\{X_t\}$.

Although my review is mostly concerned with discrete-time time series analysis, I think it appropriate to make a few remarks on the continuous-time case. Of course, a natural setting for the continuous-time case is the stochastic differential equation (SDE). There is much literature on SDEs and I shall leave the topic to the experts, except to mention Tong and Yeung (1991) and Brockwell and Williams (1997), who studied TAR models in continuous time, being primarily motivated by the analysis of unequally spaced time series data. Recent interests in high frequency data analysis add to the importance of the SDEs.

## 4. SOME PROBABILISTIC PROPERTIES

### 4.1 Ergodicity

The first question for any time series models is the existence of stationary distributions. The case of linear ARMA models is now well understood. For nonlinear time series models, the situation is much more complicated and the results are incomplete to-date. Of course, it is clear that threshold MA models are strictly stationary. For nonlinear AR models, some general results are available by treating them as Markov chains with states in a Euclidean space. A powerful tool is the drift criterion initiated by Foster (1953) in queueing theory and later more fully developed by Tweedie (1975) and Nummelin (1978, 1984). The basic idea is to first introduce the so-called small sets, which play the role of discrete states, then study their irreducibility and finally check if there is a drift towards the 'centre' of the state space. For more details of the drift criterion with special reference to nonlinear time series, see e.g. Tjøstheim (1990), and Tong (1990, esp. Appendix 1 by K. S. Chan).

A key notion in the drift criterion approach is the so-called $g$-function, the specific choice of which is not always obvious. However, by interpreting it as a 'generalized' energy function, Chan and Tong (1985) has established a link between the stability of a deterministic system and the ergodicity of a stochastic system. Essentially, they have shown how the existence of a Lyapunov function for the former can be exploited to provide the $g$-function to effect the drift criterion for the latter. The deterministic system is obtained by reducing the variance of the innovation of the nonlinear AR model to zero. They call the resulting deterministic system the 'skeleton' of the nonlinear AR model. (There is an interesting connection between the skeleton and the Kolmogorov construction in the chaos literature. A deterministic dynamical system typically admits more than one invariant measure. To obtain the 'physically meaningful' invariant measure, the Kolmogorov construction puts stochastic noise into the deterministic dynamical system. This produces a unique invariant measure if the stochastic system is ergodic. Then, the noise is reduced to zero and the limiting invariant measure is taken as the physically meaningful invariant measure.) The exploitation can sometimes facilitate the checking of stationarity of a complex nonlinear AR model because of the existence of a wealth of knowledge on stability in the applied mathematics and engineering literature. For recent developments of the skeleton approach with special reference to TAR models, see Cline (2009). For recent developments of the related area of stochastic stability, see Meyn et al. (2009).

To give a flavour of the approach, I recall the following theorem due to Chan and Tong (1985). First, let me state Condition A. Note that the first two parts concern the skeleton and its stability while the last part concerns the stochastic noise.

### Condition A

A1. $T$ is Lipschitz continuous over $R^m$, that is $\exists M > 0$, such that $\forall \mathbf{x}, \mathbf{y} \in R^m, \|T(\mathbf{x}) - T(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\|$.

A2. $T(\mathbf{0}) = \mathbf{0}$ and $\exists K, c > 0$ such that $\forall t \geq 0$, and starting with $\mathbf{x}_0 \in R^m, \|\mathbf{x}_t\| \leq Ke^{-ct}\|\mathbf{x}_0\|$, where $\|.\|$ denotes the Euclidean norm in $R^m$.

**A3.** *Either* $\{\varepsilon_t\}$ are independent identically distributed random variables, the marginal distributions of which are absolutely continuous (with respect to the Lebesgue measure) and have each an everywhere positive probability density function over $R^m$ with $E\|\varepsilon_t\| < \infty$.

*Or* $\varepsilon_t = (e_t, 0, \ldots, 0)'$ with $\{e_t\}$ being independent identically distributed random variables, each having an absolutely continuous distribution function with an everywhere positive probability density function over $R$ and $E|e_t| < \infty$.

**Theorem 4.1.** *Under condition A, the Markov chain with state space $R^m$ and given by*

$$\mathbf{X}_t = T(\mathbf{X}_{t-1}) + \varepsilon_t, \ \ t \geq 1, \ \ T : R^m \to R^m$$

*is geometrically ergodic.*

The example below re-visits the case studied by Petruccelli and Woolford (1984) and highlights the facility that the above theorem can sometimes provide.

**Example 4.1.**

$$x_t = \begin{cases} \alpha x_{t-1} & \text{if } x_{t-1} > 0 \\ \beta x_{t-1} & \text{if } x_{t-1} \leq 0. \end{cases}$$

Clearly, $|\alpha| < 1$ and $|\beta| < 1 \Rightarrow x_t \to 0$ as $t \to \infty$. As it will be shown later, stationarity region can be considerably bigger on the $(\alpha, \beta)$-space.

*Case (i):* $\alpha > 0, \beta > 0$
If $x_0 > 0$, then $x_k = \alpha^k x_0$, $k = 1, 2, \ldots$. Stability requires $\alpha < 1$ so that $x_k \to 0$ as $k \to \infty$. If $x_0 < 0$, then $x_k = \beta^k x_0$, $k = 1, 2, \ldots$. Stability requires $\beta < 1$ so that $x_k \to 0$ as $k \to \infty$.

*Case (ii):* $\alpha > 0, \beta < 0$
If $x_0 > 0$, then $x_k = \alpha^k x_0$, $k = 1, 2, \ldots$. Stability requires $\alpha < 1$ so that $x_k \to 0$ as $k \to \infty$. If $x_0 < 0$, then $x_1 = \beta x_0(> 0)$; $x_2 = \alpha x_1(> 0)$; $x_3 = \alpha^2 x_1(> 0)$;... So $x_k = \alpha^{k-1} x_0$, $k = 2, 3, \ldots$. Stability requires $\alpha < 1$ so that $x_k \to 0$ as $k \to \infty$.

*Case (iii):* $\alpha < 0, \beta > 0$
By symmetry with case (ii), it is required that $\beta < 1$.

*Case (iv):* $\alpha < 0, \beta < 0$
$x_k = \alpha \beta x_{k-2}$, $k = 2, 3, \ldots$ Stability requires $\alpha \beta < 1$ so that $x_k \to 0$ as $k \to \infty$.

Putting everything together, the condition can be relaxed for stability of the skeleton from $|\alpha| < 1$ and $|\beta| < 1$ to $\alpha < 1, \beta < 1$ and $\alpha \beta < 1$, the latter being considerably bigger. (Note that the above argument has effectively used

a Lyapunov function, $V(x)$, of the form: $V(x) = ax$ for $x \geq 0$ and $b|x|$ for $x \leq 0$, where $a > 0, b > 0, 1 > \alpha > -a/b, 1 > \beta > -b/a$.) Thus, it can be concluded that, under the same condition on $\alpha$ and $\beta$, the SETAR model

$$X_t = \begin{cases} \alpha X_{t-1} + \varepsilon_t & \text{if } X_{t-1} > 0 \\ \beta X_{t-1} + \varepsilon_t & \text{if } X_{t-1} \leq 0, \end{cases}$$

where $\varepsilon_t \sim IID(0, \sigma^2)$ with an everywhere positive probability density function is geometrically ergodic. The model is then strictly stationary if $X_0$ has the unique invariant measure. (Note: $\sigma^2$ need not be finite.) Using the so-called inverse theorems in stability theory, the necessity of the condition can also be established.

For more complicated nonlinear AR models, sometimes ingenuity may have to be exercised or searched in the applied mathematics and engineering literature to produce one or more appropriate Lyapunov functions. Even if this fails, running the noise-free skeleton with a variety of initial values can often be quite informative.

## 4.2 Stationary distributions

Under strict stationarity and conditionally homoscedastic noise, the probability density function (pdf), say $\Pi(x)$, (assumed to exist), of a nonlinear AR model of order 1 satisfies the following integral equation:

$$(8) \qquad \Pi(x) = \int_{-\infty}^{\infty} \Pi(y) p_\varepsilon(x - f(y)) dy,$$

where $p_\varepsilon(\cdot)$ denotes the pdf of the innovation.

Analytic solutions of such equations rarely exist in closed form; numerical solutions are usual. Even for strictly stationary linear AR models, the situation can be non-trivial: although the Gaussian solution is well known for the case with Gaussian innovation, an analytic solution may not always be possible for the case with non-Gaussian innovation. For nonlinear AR models, analytic solutions are even rarer. However, analytic solutions are useful when they do exist. They can be used, among others, to check the efficacy of numerical techniques suggested for solving the above integral equation and to provide stochastic models and insights for some of the newer distributions proposed in the literature.

**Example 4.2** (A special SETAR model)**.**

$$X_t = -\alpha|X_{t-1}| + \varepsilon_t; \ \ |\alpha| < 1,$$

where $p_\varepsilon(\cdot)$ is assumed to be symmetric about the origin. Exploiting the symmetry of the autoregressive function, Chan and Tong (1986b) showed how the nonlinear integral equation for the stationary distribution could be solved by ref-
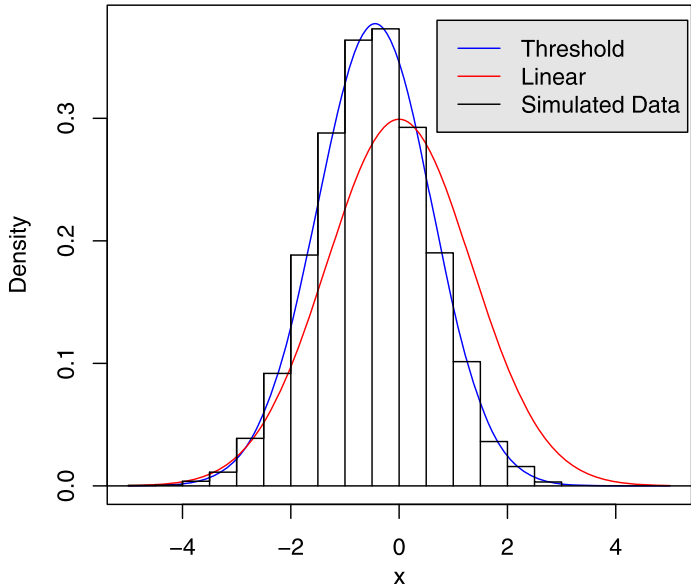
*Figure 5. Stationary distribution of Example 4.2.*

erence to a linear AR model:

$$\Pi(y) = \int_{-\infty}^{\infty} \Pi(x) p_\varepsilon(y + \alpha|x|) dx$$

$$= \int_{0}^{\infty} \Pi(x) p_\varepsilon(y + \alpha x) dx + \int_{-\infty}^{0} \Pi(x) p_\varepsilon(y - \alpha x) dx$$

$$= \int_{0}^{\infty} \Pi(x) p_\varepsilon(y + \alpha x) dx + \int_{0}^{\infty} \Pi(-x) p_\varepsilon(y + \alpha x) dx$$

$$= 2 \int_{0}^{\infty} \bar{\Pi}(x) p_\varepsilon(y + \alpha x) dx,$$

where

$$2\bar{\Pi}(y) = \Pi(y) + \Pi(-y) = 2 \int_{-\infty}^{\infty} \bar{\Pi}(x) p_\varepsilon(y + \alpha x) dx,$$

corresponding to the integral equation for the stationary *linear* AR(1) model with parameter $\alpha$.

If $\varepsilon_t \sim N(0, 1)$, then $\bar{\Pi}(x)$ is the pdf for $N(0, \frac{1}{1-\alpha^2})$, and so

$$(9) \quad \Pi(x) = \sqrt{\frac{2(1 - \alpha^2)}{\pi}} \exp\left\{ -\frac{1}{2}(1 - \alpha^2) x^2 \right\} \Phi(-\alpha x).$$

Andel et al. (1984) obtained the same solution by postulating and verifying it.

Figure 5 shows the skewness of the solution. By varying the value of $\alpha$, a family of skew densities can be produced that includes the normal at $\alpha = 0$. In fact, this SETAR model provides a stochastic model that can be used to underpin the so-called family of 'skew-normal distributions' advocated by Azzalini (1985). These distributions were apparently first introduced by O'Hagan and Leonard (1976) in a Bayesian context. For multivariate time series, there are many more groups of symmetry available than the above reflexive symmetry. Chan and Tong (1986b) also showed how similar exploitation could be made of these. I think that some of their results are relevant to developing multivariate generalizations of skew-normal distributions, skew-t distributions and others, although they do not seem to have been followed up for the purpose.

**Example 4.3** (A piecewise constant AR model). This is a particularly simple SETAR model.

$$(10) \qquad X_t = b + (a - b)I(X_{t-1} < 0) + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, 1)$. The stationary pdf is a mixture of Guassian distributions (-I learnt this result from John Pemberton):

$$(11) \qquad \Pi(x) = A\phi(x - a) + (1 - A)\phi(x - b),$$

where $A = \Phi(-b)/\{1 - \Phi(-a) + \Phi(-b)\}$.

**Example 4.4** (A martingale difference model). This is another simple special case of the SETAR model (Tong, 2007b).

$$(12) \qquad X_t = \begin{cases} \varepsilon_t & \text{if } X_{t-1} \leq 0, \\ \beta\varepsilon_t & \text{if } X_{t-1} > 0, \end{cases}$$

where $\{\varepsilon_t\}$ is standard Gaussian white noise. Simple calculation yields the stationary marginal pdf

$$(13) \quad f(x) = \frac{1}{2}\left[\phi(x) + \frac{1}{\beta}\phi\left(\frac{x}{\beta}\right)\right], \quad -\infty < x < \infty,$$

where $\phi(\cdot)$ denotes the pdf of the standard normal, giving another mixture of standard Gaussian distributions. Clearly, it is leptokurtic for $\beta \neq 1$. I shall return to this model in Section 7.

Once the marginal distribution is found, then the next job would be the joint distributions and the conditional distributions. The conditional distributions are of particular relevance to prediction and are sometimes called the predictive distributions. In fact, the derivative of the m-step-ahead predictive distribution, with respect to the 'present' datum, can be used to form the basis of a stochastic extension of the notion of Lyapunov exponent that characterizes the initial-value sensitivity of deterministic dynamical systems. (Yao and Tong, 1994).

To evaluate more-than-one-step-ahead predictive distributions exactly is very challenging for a general nonlinear time series model. Fortunately, for SETAR models, the piecewise linearity coupled with the Markovian structure can be exploited to produce some exact solutions. Let me explain. Suppose a stationary SETAR model of the form (4) has been fitted to some real data and is now taken, for the exercise, as the true model. Then for $1 \leq m \leq d$, the $m$-step-ahead predictive distribution of $X_{t+m}$ given $X_s, s \leq t$

can be expressed exactly as an $m$-fold convolution of the distribution of $\varepsilon$ with itself, up to a suitable scaling factor by $\phi$. For $m > d$, the situation is more complicated, although there is always the Chapman-Kolmogorov equation as a last resort. However, for cycling data with an approximate period of say $\theta$, I would suggest that the SETAR model be refitted now with delay parameter fixed at $d + \theta$. Then, under stationarity, the above convolution solution could still be used to give a reasonable approximate solution for cases with $m \in [d+1, d+\theta]$. How good the approximation is depends on how strongly cyclical the data are. In principle, the method can be extended to cover further periods but it is expected that the approximation will deteriorate rapidly, unless the data are very strongly cyclical. The above suggestion rests heuristically on the notion of cyclically moving subsets in a Markov chain due to W. Doeblin (See, e.g., Doob, 1953), which I have exploited in Tong (1983) for point predictors.

### 4.3 Invertibility and irreversibility

Invertibility concerns the ability to express the innovation in terms of present and past observations. The issue for nonlinear ARMA models (including TARMA models) is non-trivial and has attracted some attention recently. For the following threshold MA model, Ling et al. (2007) have obtained sufficient and almost necessary conditions.

$$(14) \quad X_t = \left\{ \phi_0 + \sum_{j=1}^{k} \psi_j I(r_{j-1} < X_{t-1} \leq r_j) \right\} \varepsilon_{t-1} + \varepsilon_t,$$

where $-\infty = r_0 < r_1 < \cdots < r_k = \infty$. Let $F_x(\cdot)$ denote the distribution function of $X_t$. They proved the following theorem.

**Theorem 4.2.** *Model (14) is invertible if $\prod_{j=1}^{k}\{|\phi_0 + \psi_j|^{F_x(r_j)-F_x(r_{j-1})}\} < 1$ and is not invertible if $\prod_{j=1}^{k}\{|\phi_0 + \psi_j|^{F_x(r_j)-F_x(r_{j-1})}\} > 1$.*

More recently, Chan and Tong (2010) have considered the harder problem of the invertibility of nonlinear ARMA models. Under very general conditions, their local analysis shows that there is a generic dichotomy that the innovation reconstruction errors either diminish geometrically fast or grow geometrically fast. They derive a simple sufficient condition for a nonlinear ARMA model to be locally invertible.

Despite their linguistic similarity, the terms 'invertibility' and 'reversibility' refer to two different concepts. Intuitively, a time series is time-reversible if running it forward in time has the same probability distribution as running it backward in time, i.e. the direction of time is irrelevant. For a stationary univariate time series, reversibility is universal for the Gaussian case but rare for the non-Gaussian case. Even a *linear* AR model driven by non-Gaussian innovation is generally time irreversible. A striking and classic example is the AR(1) model: $X_t = \frac{1}{2}X_{t-1} + \varepsilon_t$, $t \geq 0$, where $\{\varepsilon_t\}$ is a sequence of iid random variables with a two-point

distribution residing at 0 or 1 with equal probability, and $X_0 \in [0, 1]$. In reversed time, the model is a noise-free SETAR model, which generates chaos. (Clearly this model is not strongly mixing.) This model has attracted the attention of many time series analysts, including Whittle (1963, p. 24), who referred to Moran, Rosenblatt (1964), who referred to Jamison, and Bartlett (1990), which was his last research paper. For a stationary multivariate time series, reversibility is rare for all cases. (Chan et al., 2006). I find that a lot can be learnt about a statistical technique for time series by first running the time series forward in time and then backward in time. For example, the forward-backward check will show that, for vector AR modelling, the naive uni-directional Levinson-Durbin algorithm has to be modified. In fact, a pair of algorithms running simultaneously is needed, one for the 'forward' AR model and another for the 'backward' AR model. (Whittle, 1963, p. 102). Another example is the so-called backforecasting technique due to Box and Jenkins (1970, Ch. 7). In fact, it is predicated on time-reversibility.

## 5. STATISTICAL MODELLING

### 5.1 Initial data analysis

The emergence of the threshold models and other non-linear time series models has led to the development and sometimes the re-discovery of numerous data analytic techniques. Many of these have found their way into book forms, e.g. Tong (1990), Tsay (2002), Fan and Yao (2003), Cryer and Chan (2008). An amusing example of re-discovery is the plot of $X_t$ against $X_s$. For an early exponent of correlation analysis such as Yule, such plots came very naturally or even routinely. Indeed, Yule (1927, p. 277) plotted $X_t$ against $X_{t-1} + X_{t+1}$. The choice of the latter is, however, intriguing. Could he have some vague inkling of time irreversibility in view of the asymmetry of the sunspot cycles? Sadly these plots dropped out of favour with time series practitioners in the 1960s and 1970s, most probably due to the pre-occupation with linear Gaussian time series of the period. In contrast, free from attachment to linearity and normality, population ecologists have continued to plot them, under the name of 'Moran diagrams'. I remember being fascinated in 1978 by such a plot as I was watching it evolving on the monitor of my IBM 386 computer; slow computers can sometimes be more helpful than fast ones! I joined the scattered points by straight lines as they were being plotted over time and called the resulting plot a (discrete-time) phase plot, following the terminology in dynamical systems.

### 5.2 Fitting a threshold model

A popular approach to fit a threshold model is the conditional least squares method. The first rigorous study of its properties for a SETAR model was by Chan (1993). He assumed that the SETAR model is geometric ergodic and stationary, with strictly positive and uniformly continuous distribution for the innovations, and both the innovations and

the marginal distribution of the time series have finite second order moment. Under these assumptions. he proved that all the estimators (i.e. for the threshold $r$, the delay $d$ and the 'slopes' and 'intercepts') are consistent. He further proved that if finite fourth moments exist, then (i) the threshold estimator is $N-$consistent; (ii) the threshold estimator is asymptotically independent of the slope estimators and the intercept estimators; and (iii) the slope estimators and the intercept estimators are $\sqrt{N}-$consistent and asymptotically normal. He also showed that the threshold estimator follows a compound Poisson process (CCP), due to the fact that the sum of squares of the residuals is a discontinuous function of the threshold parameter. He noted that $N-$consistency is not unusual by recalling a similar feature in the estimation of a 'jump' in a cumulative distribution function.

After Chan's important results, the estimation of the threshold parameter has continued to attract much attention since it is a key parameter of the SETAR model. For example, Hansen (1997) looked at the case in which the threshold effect (the difference in slopes between the two regimes) becomes small as the sample size increases and showed then the asymptotic distribution of the threshold estimate is free of nuisance parameters (up to scale). Chan and Tsay (1998) studied the estimation of the threshold parameter of a continuous SETAR model and showed $\sqrt{N}-$ asymptotic normality. Bayesian estimation has also been studied. See, e.g., Stramer and Lin (2002), Chen and Lee (2008) and Geweke (2009). For the related STAR model, although $\sqrt{N}-$ normality is obtained for the estimators of all the parameters (Chan and Tong, 1986), the imprecision of the estimator of the 'smoothness' parameter $\sigma$ is problematic unless the sample size is very large. This has a bearing on the issue of SETAR models versus STAR models for real data, to which I shall return shortly.

Fitting a time series model to real data is an art as well as a science. To assist practitioners to fit a TAR model, numerous statistical tools are now available and many of these have been described in book forms. See, e.g., Tong (1990), Tsay (2002), Fan and Yao (2003), Cryer and Chan (2008) and others. Computer softwares are also available, some of which come as companions to books. The R codes in Cryer and Chan (2008) are particularly helpful.

In practical applications, whether a STAR model or a SETAR model should be used depends on a number of considerations. However, there is always a place for sensible compromises. Let me elaborate. Take the (log transformed) MacKenzie lynx data for example. Ecological theory will probably support a smooth autoregressive function. Now, there is evidence to support that a threshold should lie in the vicinity of 3. However, Figure 4 shows that there are few observations to provide sufficient information on the precise shape of the autoregression around 3, which is the *anti-mode*. In this case, a SETAR model, of either the continuous or the discontinuous genre, is a reasonable and practical compromise/approximation because the 'true'

smooth autoregessive function simply cannot be estimated with any precision around the threshold. The best that can be hoped for is the estimation of the function away from the anti-mode, where linear dynamics would be quite an adequate approximation. On the other hand, a STAR model is clearly sensible by reference to ecological dynamics but the imprecision of the estimator of the 'smoothness' parameter $\sigma$ must be noted.

There are, of course, situations in which subject matter considerations would dictate an un-smooth or even discontinuous model. For example, a decision (e.g. an intervention) is often discontinuous, which may lead to a switch to a different dynamics after say $d$ units of time. The connection between discontinuous decision processes and TAR modelling was discussed in a rarely cited note (Tong, 1982).

Although a TARMAX model and a multivariate TAR model have been studied and fitted to real data, e.g. Tong (1990), the sampling properties of the estimators have not been studied in depth. Moreover, the estimation of parameters for more than two regimes is not fully developed.

In real applications of TAR models involving several time series, it is not often clear as to the best way to define the indicator time series $\{Z_t\}$, i.e. the switching mechanism. The principle of parsimony coupled with computational considerations suggests that a vectorial $\{Z_t\}$ is probably not a practical proposition. A practical approach could be based on some sort of (non-linear) principal components. See, e.g., Wu and Chen (2007) and Xia et al. (2007).

## 5.3 Testing for linearity in threshold models

In model (6), suppose that $\phi = 1$. Consider the null hypothesis $H_0 : \alpha = \gamma, \beta = \delta$. Under $H_0$, model (6) is a linear AR model. Thus, to test $H_0$, or its obvious extension to higher order cases, is equivalent to testing for linearity within the SETAR setup. This was a non-standard problem because, under $H_0$, the nuisance parameter $r$ is absent, implying that the conventional $\chi^2$ asymptotics for the likelihood ratio test would not apply. Chan (1990) and Chan and Tong (1990) laid the theoretical foundations, by recognizing that the problem is related to the maximum of a stochastic process. These led to usable percentage points (Chan, 1991) and extensions to threshold cointegration (Hansen and Seo, 2002), threshold MA models (Ling and Tong, 2005), and others. I should remark that the non-standard problem has serious implications on model selection. If the selection criterion is based on the likelihood, such as Akaike's information criterion or the Bayesian information criterion, then care should be exercised in counting the number of *independently adjusted* parameters. This remark applies to parametric nonlinear time series modelling, neural networks, and others.

A separate but related area is the goodness-of-fit test. Li's comprehensive monograph (2004) is very valuable. There have been some recent advances in the use of empirical processes, following the lead by An and Cheng (1991). For example, adopting a score-based approach, Ling and Tong

(2010) has developed an array of goodness-of-fit tests, each tailored for a specific model, e.g. an SETAR model. This approach can, in many cases, lead to greater power than residual based tests.

## 5.4 Real applications

The threshold models have been widely applied. Significant examples are in ecology (e.g. Stenseth et al., 1999), epidemiology (e.g. Stenseth et al., 2006; Samia et al., 2007), actuarial science (e.g. Chan et al., 2004), economics and finance (e.g. Tsay, 2002), water resources (e.g. Tong et al., 1985). Many other examples can be found in, e.g., Tong (1990) or by a Google search. Let me highlight just one example to give some general flavour. In ecology, key questions include (i) What gives rise to the lynx cycle across the whole of Canada (not restricted to just the MacKenzie River region)? Is it due to 'phase-dependency' or 'density-dependency', in the language of the ecologists? (ii) In what way does the lynx dynamics vary with climatic regions? Stenseth (2009) has given an authoritative summary of the many collaborative contributions made, over many years, by a team of ecologists, biologists, statisticians and climatologists gathered around him.

# 6. SOME OFF-SPRINGS

In Chan (2009), many of the important off-springs of the TAR model have been discussed. Some notable ones are threshold unit root (e.g. Chan et al. (1991), Enders and Granger (1998) and Caner and Hansen (2001)), threshold co-integration (e.g. Balke and Fomby, 1997), threshold GARCH models (e.g. Glosten et al. (1993), Zakoian (1994), Li (2009)), double threshold ARMA models (e.g. Li and Li, 1996).

# 7. THRESHOLD VOLATILITY MODELS

In the application of SETAR to the MacKenzie lynx data, Tong and Lim (1980) addressed the issue raised by Moran referred to in Section 2 by having different estimates for the innovation variance in different regimes. However, it is clear that the full potential of the TAR models for the modelling of the conditional variance has not been exploited. I shall now report some on-going research conducted by Chan, Ling and myself. I have referred to the leptokurticity of model (12). Now, consider a multi-regime generalization of the model.

$$(15) \qquad X_t = \sigma(X_{t-1})\varepsilon_t.$$

where $R = \bigcup_{i=1}^{m} R_i$, $R_i$s are mutually disjoint subsets of the real line, $\sigma(x) = \sigma_i > 0$ for $x \in R_i$, and $\varepsilon_t$ are independent and identically distributed random variables with zero mean and unit variance, independent of past $X$'s. The above model is called the *threshold volatility model*, or TVM for short, by Chan et al. (2010), in which the following results are established.

FACT 1: TVM is always strictly stationary and ergodic.

FACT 2: $E[X_t|X_{t-1}] \equiv 0$; $\{X_t\}$ is a martingale difference sequence.

FACT 3: Under general conditions, any continuous instantaneous nonlinear transformation of $\{X_t\}$ is an $ARMA(m-1, m-1)$ process.

**Theorem 7.1.** *Let* $P = (p_{ij})$, *where* $p_{ij} = Pr(X_{t+1} \in R_j | X_t \in R_i)$. *Let* $Y_t = h(X_t)$, *where* $h$ *is a continuous function. Assume that* $P$ *is irreducible,* $\{Y_t\}$ *admits finite second order moments and that* $E(h(\varepsilon_t)) \neq 0$. *Let* $\gamma_k = \gamma_{k,Y}$ *be the lag-k auto-covariance of* $\{Y_t\}$. *Then, there exist constants* $c_i, i = 1, \ldots, c_{m-1}$ *such that* $\{\gamma_k\}$ *satisfies the Yule-Walker equation:*

$$(16) \qquad \gamma_k = c_1\gamma_{k-1} + \cdots + c_{m-1}\gamma_{k-m+1},$$

*for* $k \geq m$.

The above theorem indicates the possibility of clustering effects. The order of the ARMA model is potentially useful for the determination of the number of regimes for the TVM in practical applications.

Note that by replacing the argument $X_{t-1}$ of the piecewise constant function $\sigma(\cdot)$ in (15) by $U_{t-1}$, where $\{U_t\}$ is a hidden time series independent of $\{X_t\}$, a hidden TVM results. It parallels the well-known stochastic volatility model.

# 8. CONCLUDING REMARKS

At the dawn of time series modelling, masters like Udny Yule (1927) were aware that linear models are unrealistic. Whittle (2009) recounted that 'nonlinearity was forced on him observationally when a seiche study (Whittle, 1954) revealed the existence of subharmonics.' He took a decisive step and suggested a piecewise linear differential model in 1954, referring to the master piece by Andronov and Khaikin (1937) for the necessary theory. In a different context, Tukey (1961) proposed the regressogram, which is piecewise constant. Akaike told me in the 1970s that the 'secret' of his success in controlling a complex cement kiln is a piecewise linear filter rather than the multivariate linear AR models. He alluded to this in his discussion of Tong and Lim (1980). I have never doubted the inevitability of the threshold principle in time series analysis just as I have never doubted non-linearity. The proliferation of the threshold family, over the past 30 years, merely illustrates once again the universal truth that simplicity is not only beautiful but also productive.

Looking ahead, I can see that the threshold principle will probably continue to make worthwhile contributions in time series analysis over the next 30 years. In particular, I am optimistic that it will lead to advances, for example, in nonstationary-nonlinear time series modelling, multivariate time series modelling, spatial-temporal series modelling, panel time series modelling and others.

## REFERENCES

AN, H. and CHENG, B. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *Int. Stat. Rev.*, **59**, 287–307.

ANDEL, J., NETUKA, I. and ZVARA, K. (1984). On threshold autoregressive processes. *Kybernetika*, **20**, 89–106. MR0747062

ANDRONOV, A. A. and KHAIKIN, S. E. (1937). *Theory of Oscillations* (in Russian). (English trans. by S. Lefschetz, 1947, Princeton Univ. Press.)

AZZALINI, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist*, **12**, 171–178. MR0808153

BALKE, N. and FOMBY, T. (1997). Threshold cointegration. *Int. Econ. Rev.*, **38**, 627–645. MR1467826

BARTLETT, M. S. (1990). Chance or chaos? (with discussions) *J. Roy. Stat. Soc.*, Series A, **153**, 321–347.

BEC, F., RAHBEK, A. and SHEPHARD, N. (2008). The ACR Model: A Multivariate Dynamic Mixture Autoregression. *Oxf. Bull. Econ. & Stat.*, **70**,583–618.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day. MR0272138

BROCKWELL, P. J. and WILLIAM, R. J. (1997). On the existence and application of continuous-time threshold autoregressions of order two. *Adv. Appl. Prob.*, **29**, 205–227. MR1432937

CANER, M. and HANSEN, E. B. (2001). Threshold autoregression with a unit root. *Econometrica*, **69**, 1555–1596. MR1865221

CHAN, K. S. (1986). *Topics in nonlinear time series analysis*. Doctoral thesis, Princeton University.

CHAN, K. S. (1988). On the existence of the stationary and ergodic NEAR (p) model. *J. Time Series Anal.*, **9**, 319–328. MR0995610

CHAN, K. S. (1990). Testing for threshold autoregression. *Ann. Stat.*, **18**, 1886–1894. MR1074443

CHAN, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *J. Roy. Stat. Soc. Ser. B*, **53**, 691–696. MR1125726

CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Stat.*, **21**, 520–533. MR1212191

CHAN, K. S. (2009). *Exploration of a Nonlinear World*. Singapore: World Scientific.

CHAN, K. S. and TONG, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Prob.*, **17**, 666–678. MR0798881

CHAN, K. S. and TONG, H. (1986a). On estimating thresholds in autoregressive models. *J. Time Series Anal.*, **7**, 179–190. MR0857248

CHAN, K. S. and TONG, H. (1986b). A note on certain integral equations associated with non-linear time series analysis. *Prob. Th. Rel. Fields*, **73**, 153–159. MR0849071

CHAN, K. S. and TONG, H. (1990). On likelihood ratio tests for threshold autoregression. *J. Roy. Stat. Soc., Ser. B*, **52**, 469–476. MR1086798

CHAN, K. S. and TONG, H. (2001). *Chaos: A Statistical Perspective*. New York: Springer. MR1851668

CHAN, K. S. and TONG, H. (2004). A note on testing for multi-modality with dependent data. *Biometrika*, **91**, 113–123. MR2050463

CHAN, K. S. and TONG, H. (2010). A note on the invertibility of nonlinear ARMA models. *J. Stat. Plan. & Inf.*, **140**, 3709–3714.

CHAN, K. S. and TSAY, R.S. (1998). Limiting properties of the conditional least squares estimator of a continuous TAR model. *Biometrika*, **85**, 413–426. MR1649122

CHAN, K. S., HO, L-H. and TONG, H. (2006). A note on time-reversibility of multivariate linear processes. *Biometrika*, **93**, 221–227. MR2277753

CHAN, K. S., LING, S. and TONG, H. (2010). On threshold volatility models. *under preparation.*

CHAN, K. S., PHAM, D. T. and TONG, H. (1991). Strong consistency of least squares estimator for non-ergodic threshold autoregressive model. *Stat. Sinica*, **1**, 361–369. MR1130124

CHAN, W. S., TONG, H. and WONG, A. C. S. (2004). Some nonlinear threshold autoregressive time series models for actuarial use. *North Amer. Act. J.*, **8**, 37–61. MR2100929

CHEN, C. W. S. and LEE J. C. (2008). Bayesian inference of threshold autoregressive models. *J. Time Series Anal.*, **16**, 483–492. MR1365643

CLINE, D. B. H. (2009). Thoughts on the connection between threshold time series models and dynamical systems. In *Exploration of a Nonlinear World*, ed. K. S. Chan, Singapore: World Scientific, 165–181.

CRYER, J. D. and CHAN, K. S. (2008). *Time Series Analysis with Applications in R*. New York: Springer.

DOOB, J. L. (1953). *Stochastic Processes*, New York: John Wiley. MR0058896

ENDERS, W. and GRANGER, C. W. J. (1998). Unit root tests and asymmetric adjustment with an example using the term structure of interest rates. *J. Bus. & Econ. Stat.*, **16**, 304–311.

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer. MR1964455

FOSTER, F. G. (1953). On the stochastic matrices associated with certain queueing processes. *Ann. Math. Stat.*, **24**, 355–360. MR0056232

FRANSES, P. H. and VAN DIJK, D. (2000). *Non-linear Time Series Models in Empirical Finances.* Cambridge: Cambridge University Press.

GEWEKE, J. (2009). The SETAR model of Tong and Lim and advances in computation. In *Exploration of a Nonlinear World*, ed. K. S. Chan, 85–94. Singapore: World Scientific.

GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. E. (1993). On the relation between representation, estimation and testing. *Econometrika*, **55**, 251–276.

GOURIEROUX C. and ROBERT C. (2006). Stochastic unit root models. *Econ. Th.*, **22**, 10521090. MR2328528

HAMILTON, J. D. (1989). A new approach to economic analysis of nonstationary time series and the business cycle. *Econometrika*, **57**, 357–384. MR0996941

HANSEN, E. B. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics*, **2**, 1–14. MR1467458

HANSEN, E. B. and SEO, B. (2002). Testing for two-regime threshold cointegration in vector error correction models. *J. Econometrics*, **110**, 293–318. MR1928307

JEFFREYS, H. (1940). The variation of latitude. *Mon. Not. Roy. Astro. Soc.*, **100**, 139–155.

LI, W. K. (2004). *Diagnostic Checks in Time Series*. London: Chapman & Hall/CRC.

LI, W. K. (2009). The threshold approach to volatility modelling. In *An Exploration of a Nonlinear World*, ed. K. S. Chan, Singapore: World Scientific, 95–100.

LI, C. W. and LI, W. K. (1996). On a double threshold heteroskedastic autoregressive time series model. *J. Appl. Econ.*, **11**, 253–274.

LING, S. and TONG, H. (2005). Testing for a linear MA model against threshold MA models. *Ann. Stat.*, **33**, 2529–2552. MR2253094

LING, S. and TONG, H. (2010). Score based goodness-of-fit tests for time series. *To appear in Stat. Sinica.*

LING, S., TONG, H. and LI, D. (2007). Ergodicity and invertibility of threshold moving-average models. *Bernoulli*, **13**, 161–168. MR2307400

MEYN, S., TWEEDIE, R. L. and GLYNN, P. (2009). *Markov Chains and Stochastic Stability.*, Cambridge: Cambridge Univ. Press. MR2509253

MORAN, P. A. P. (1953a). The statistical analysis of the Canadian lynx cycle. *Aust. J. Zoo.*, **1**, 163–173.

MORAN, P. A. P. (1953b). The statistical analysis of the Canadian lynx cycle, II. synchronization and meterology. *Aust. J. Zoo.*, **1**, 291–298.

NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeit. Wahr. Ver. Gebiete*, **43**, 309–318. MR0501353

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge: Cambridge Univ. Press. MR0776608

O'HAGAN, A. and LEONARD, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **63**, 201–202. MR0428571

PETRUCCELLI, J. D. and WOOLFORD, S. W. (1984). A threshold AR(1) model. *J. Appl. Prob.*, **21**, 207–286. MR0741130

ROSENBLATT, M. (1964). Some nonlinear problems arising in the study of random processes. *Radio Sci. J. Research*, NBS (USNC-RRSI), **68D**, 933–936. MR0165570

SAMIA, N. I., CHAN, K. S. and STENSETH, N. C. (2007). A generalized threshold mixed model for analyzing nonnormal nonlinear time series; with application to plague in Kazakhstan. *Biometrika*, **94**, 101–118. MR2367827

STENSETH, N. C., CHAN, K. S., TONG, H., BOONSTRA, R., BOUTIN, S., KREBS, C. J., POST, E., O'DONOGHUE, M., YOCCOZ, N. G., FORCHHAMMER, M. C. and HURRELL, J. W. (1999). Common dynamic structure of Canadian lynx populations within three climatic regions. *Science*, **285**, 1071–1073.

STENSETH, N. C. (2009). The importance of TAR-modelling for understanding the structure of ecological dynamics: the hare-lynx population cycles as an example. In *Exploration of a Nonlinear World*, ed. K. S. Chan. Singapore: World Scientific, 365–374.

STENSETH, N. C., SAMIA, N. I., VILJUGREIN, H., KAUSRUD, K., BEGON, M., DAVIS, S., LEIRS, H., DUBYANSKIY, V. M., ESPER, J., AGEYEV, V. S., KLASSOVSKIY, N. L., POLE, S. B. and CHAN, K. S. (2006). Plague dynamics are driven by climate variation. *Proceedings of the National Academy of Sciences*, **103**, 13110–13115.

STRAMER, O. and LIN, Y. (2002). On inference for threshold autoregressive models. *Test*, **11**, 51–71. MR1915777

TJØSTHEIM, D. (1990). Nonlinear time series and Markov chains. *Adv. Appl. Prob.*, **22**, 587–611. MR1066965

TONG, H. (1982). Discontinuous decision processes and threshold autoregressive time series modelling. *Biometrika*, **69**, 274–276. MR0655696

TONG, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. New York: Springer. MR0717388

TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press. MR1079320

TONG, H. (2007a). Birth of the threshold time series model. *Stat. Sinica*, **17**, 8–14.

TONG, H. (2007b). Exploring volatility from a dynamical system perspective. *Proc. 56th Sess. Int. Stat. Inst.*, **Invited Paper session 64**.

TONG, H. and LIM, K.S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Sco.*, Series B, **42**, 245–292.

TONG, H. and YEUNG, I. (1991). Threshold autoregressive modelling in continuous time. *Statist. Sinica*, **1**, 411–430. MR1130127

TONG, H., THANOON, B. and GUDMUNDSSON, G. (1985). Threshold time series modelling of two Icelandic riverflow systems. *Water Res. Bull.*, **21**, 651–661.

TSAY, R. (2002). *Analysis of Financial Time Series*. New York: Wiley. MR2162112

TUKEY, J. W. (1961). Curves as parameters and touch estimation. *Proc 4th Berkeley Symposium*, 681–694. MR0132677

TWEEDIE, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stoch. Proc. & Appl.*. **3**, 385–403. MR0436324

TYSSEDAL, J. S. and TJØSTHEIM, D. (1988). An autoregressive model with suddenly changing parameters and an application to stock market prices. *Appl. Stat.*, **37**, 353–369. MR0966805

WHITTLE, P. (1954). The statistical analysis of a seiche record. *Sears Foun. J. Mar. Res.*, **13**, 76–100. MR0078600

WHITTLE, P. (1963). *Prediction and Regulation*, London: English Univerity Press. MR0157416

WHITTLE, P. (2009). Foreword. In *Exploration of a Nonlinear World*, ed. K. S. Chan. Singapore: World Scientific.

WU, S. and CHEN, R. (2007) Threshold variable determination and threshold variable driven switching autoregressive models. *Stat. Sinica*, **17**, 241–264.

XIA, Y. and TONG, H. (2010). Feature matching in time series modelling. *To appear in Stat. Science.*

XIA, Y., LI, W. K. and TONG, H. (2007). Threshold variable selection using nonparametric methods. *Stat. Sinica*, **17**, 265–287. MR2352510

YAO, Q. and TONG, H. (1994). Quantifying the influence of initial values on non-linear prediction. *J. Roy. Stat. Soc.*, Series B, **56**, 701–725. MR1293241

YULE, U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. Roy. Soc. (London)*, Series A, **226**, 267–298.

ZAKOIAN, J. M. (1994). Threshold heteroscedastic models. *J. Econ. Dyn. & Cont.*, **18**, 931–955.

Howell Tong

London School of Economics and Political Science

London, UK

E-mail address: howell.tong@gmail.com