RESEARCH REPORT

Serial No. 507

November 2013

THRESHOLD MODELS IN TIME SERIES ANALYSIS-SOME REFLECTIONS

by

Howell Tong

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Pokfulam Road, Hong Kong. E-Mail: saas@hku.hk
Tel.: (852) 2859 2466/7 Fax: (852) 2858 9041

Threshold models in time series analysis-some reflections

Howell Tong

London School of Economics & Political Science

Abstract: In this paper, I reflect on the developments of the threshold model in time series analysis since its birth in 1978, with particular reference to econometrics.

Key words and phrases: all-step-ahead prediction; asymmetry; Bayesian decision; business cycle; catastrophe; conditionally heteroscedastic autoregressive models with thresholds; jump resonance; mis-specified model; non-likelihood approach; nonlinear unit root; non-stationarity; open-loop system; panel threshold model; smooth threshold autoregressive models; splines; structural breaks; threshold autoregressive models; threshold moving average models; threshold principle; threshold unit root; volatility; wrong model.

Classification code: JEL:C22

1 Introduction

This paper focuses on univariate time series, although many of the key ideas are also relevant for multivariate time series.

The initial idea of threshold models in time series analysis was conceived around 1976 and the conception was announced in my contribution to the discussion of the paper read by Drs (now Professors) Lawrance and Kottegoda to the Royal Statistical Society in London in 1977. The baby's birth was certified in Tong (1978). I read Tong and Lim (1980)[1] to the Royal Statistical Society at the discussion session organised by the Research Committee on 19th March 1980. The paper has distinguished itself by having no serious theorems but being perhaps rich in ideas, some of which are yet to be explored. In particular, it addresses the important issues of 'WHY' and 'HOW': (1) Why is a nonlinear time series model needed? The paper listed deficiencies of linear Gaussian time series models in respect of limit cycles, time irreversibilty, amplitude-frequency dependency, phase transition, chaos, deeper insights and others. (2) How to do it?

[1]The paper states on p.245 that Sections 6 (simulations) and 9 (real data) are due to both authors while the other sections are due to the first author.

Recognizing the infinitude of nonlinear models, the paper proposed the threshold approach and listed the following objectives: (i) statistical identification of an appropriate model should not entail excessive computation; (ii) the model should be general enough to capture some of the nonlinear phenomena mentioned previously; (iii) one-step-ahead predictions should be easily obtained from the fitted model and, if the adopted model is nonlinear, its overall prediction performance should be an improvement upon the linear model; (iv) the fitted model should preferably reflect to some extent the structure of the mechanism generating the data based on theories outside statistics; (v) the model should preferably possess some degree of generality and be capable of generalization to the multivariate case, not just in theory but in practice. Although the paper attracted 17 discussants at its reading, it did not attract many followers for the next 17 years. In fact, even with the publication of Tong (1983) and Tong (1990), the threshold approach had to wait till the late 1990s before it started its exponential growth.

On looking back, evidence suggests that it has achieved all the objectives, with the exception of the second part of objective (v); the generalization to multivariate time series remains an unconquered challenge. To-date, the threshold approach has been adopted, sometimes with enthusiasm, in many branches of social, natural and medical sciences. Hansen (2011) has given an extensive review of threshold autoregression in economics by reference to 75 papers published in the econometrics and economics literatures, many of which are themselves highly cited. Stenseth (2009) has summarized the importance of the threshold autoregressive model for understanding the structure of ecological dynamics. Unfortunately, exchanges of experiences between disciplines have not been as widespread as they should be.

My reflections will focus on several specific issues concerning the development of threshold models in their various forms, namely the decision theoretic underpinnings of the threshold approach in Section 2; conditional distribution formulation versus stochastic difference equation formulation in Section 3; smooth threshold models versus (unsmooth) threshold models in Section 4; change points over time and over state in Section 5; threshold unit root and catastrophe in Section 6. I conclude in Section 7.

2 Decision theoretic background

Let $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$ denote a time series in discrete time and for simplicity of discussion assume that the 'true' model is

$$E(X_t|X_{t-1} = x) = \mu(x)x,$$

where μ is a 'smooth' function. From a purely deterministic perspective, we can approximate the function μ arbitrarily closely by a series of step functions on invoking the Weierstrass theorem, as described in Tong and Lim (1980). Operationally speaking, we can consider at least two different ways to approximate μ. Splines built on pre-fixed knots are an obvious candidate. Despite the many desirable properties of the spline approach, the knots (i.e. the change points) and the sub-intervals are generally a numerical device without substantive interpretation. Moreover, there is the question of model parsimony. An alternative is to let the observed time series inform us on the number of knots/change points. The threshold approach (sometimes called the threshold principle) advocated by me is precisely one such alternative. In this approach, the knots/change points are called thresholds and the sub-intervals regimes. Tong (1982) argued that we usually approximate μ with some purpose in mind, e.g. forecasting, control, filtering, etc. A natural setting to proceed is to apply the Bayesian decision theory by starting with an approximation in the form of a Bayesian linear model with Gaussian belief:

$$E(X_t|X_{t-1} = x) = \theta x,$$
$$\theta \sim N(c, V).$$

The closeness of the approximating linear model to the 'true' model is measured by the loss function that is conjugate to the Gaussian distribution, namely

$$L(\theta) = h[1 - \exp\{-\frac{1}{2k}(\theta - \mu)^2\}],$$

where h and k are positive real constants. Note that to decide whether or not the approximating linear model is acceptable we need to evaluate the expected loss $E_V(\delta)$ of making the decision δ (of shifting from c to $c + \delta$) from the class of possible decisions D. In many practical situations, it is not unreasonable to suppose that

$$V(\delta) = \alpha + \beta|\delta|,$$

where $\alpha, \beta > 0$, to reflect the assumption that a bold decision increases the uncertainty of belief. In other words, we would not expect to have to make *drastic* adjustment to the value of θ for a 'smooth' function μ. Define the expected loss function $E_V(\delta)$ by

$$E_V(\delta) = \int_{-\infty}^{\infty} L(\theta) dF_V(\theta|\delta), \quad \delta \in D,$$

where $F_V(\theta|\delta)$ denotes the distribution of θ given that the decision δ is employed. Here, $F_V(\theta|\delta)$ is $N(c+\delta, V)$, and

$$E_V(\delta) = h[1 - (\frac{k}{k+V})^{\frac{1}{2}} \exp\{-[2(k+V)]^{-1}(\delta - \mu + c)^2\}].$$

Smith *et al.* (1981) showed that the minimizer of $E_V(\delta)$ with respect to δ, the Bayes decision, is uniformly zero, meaning that no adjustment is needed[2] , for $0 < \mu(x) - c < \{(1+\gamma^2)^{\frac{1}{2}} - 1\}\gamma^{-1}$, where $\gamma = \beta(k+\alpha)^{-\frac{1}{2}}$.

The above simple application provides the Bayesian decision theoretic underpinnings of the threshold approach to nonlinear time series analysis. It is curious that the Bayesian underpinnings have gone totally unnoticed in the time series literature as well as the econometrics literature.

3 Conditional distribution formulation versus stochastic difference equation

It is well known that there are two ways to write down a model for a stationary real-valued Markov chain in discrete time, say $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$. Besides the popular stochastic difference equation, there is the alternative of using a conditional probability formulation. For threshold modelling, Tong and others have used the former formulation. See, e.g., Tong (1990, 2011). Wong and Li (2002) adopted the latter formulation and proposed a mixture autoregressive (MAR) model. Let $F(x_{t+1}|x_s, s \leq t)$ denote the conditional distribution of X_{t+1} given the past history up to and including t. In its simplest form, an MAR model may be expressed as

$$F(x_{t+1}|x_s, s \leq t) = \alpha_1 \Phi(\frac{x_{t+1} - a_0^{(1)} - a_1^{(1)} x_t}{\sigma_1})$$

[2]There were two typographical errors below equation (2.6) in Tong (1982), which are corrected here.

$$+\alpha_2\Phi(\frac{x_{t+1} - a_0^{(2)} - a_1^{(2)}x_t}{\sigma_2}),$$

where $1 \geq \alpha_i \geq 0, \alpha_1 + \alpha_2 = 1$, $\sigma_i > 0$, $a_0^{(i)}, a_1^{(i)}, i = 1, 2$ are real constants, and Φ denotes the standard normal distribution.

Let $\{J_t : t = 0, \pm 1, \pm 2, \cdots\}$ be a sequence of two-valued independent and identically distributed (iid) random variables, independent of $\{X_t\}$, with $P[J_t = 1] = \alpha_1; P[J_t = 2] = 1 - \alpha_1$. Then the MAR model is equivalent to the following threshold autoregressive model:

$$X_{t+1} = a_0^{(J_{t+1})} + a_1^{(J_{t+1})}X_t + \sigma_{(J_{t+1})}\varepsilon_{t+1}.$$

Tong (1983, pp. 63,96,97; 2011) has stressed the versatility of the indicator time series $\{J_t\}$. Here, $\{J_t\}$ is a hidden process. The possibility of extending $\{J_t\}$ from a hidden iid to a hidden Markov chain was mentioned in Tong and Lim (1980, p.285 line-12). Hamilton (1989) had not referred to the above references when he gave his development of the Markov switching model. In fact, prior to Hamilton, Chan (1986, 1988) had used the hidden setup when he studied the exponential autoregressive model of Lawrance and Lewis (1980).

A related issue is the reformulation of the MAR model as a stochastic difference equation with $\{J_t\}$ observable. The following theorem is due to Rosenblatt (1971). (See also Chan and Tong, 2002.)

Theorem (1): For a Markov chain $\{X_t\}$, there exist iid random variables $\{\varepsilon_t\}$ such that ε_{t+1} is independent of $\sigma\{X_s : s \leq t\}$ and $\sigma\{X_s : s \leq t+1\} = \sigma\{\varepsilon_{t+1}, X_s : s \leq t\}$.

The proof involves the following construction, which is sometimes called the Rosenblatt construction. Denote the conditional distribution function of X_{t+1} given X_t by $F_t(x_{t+1})$, and $F_t(X_{t+1})$ by U_{t+1}. Obviously, U_{t+1} is uniform on $(0, 1)$. Now,

$$X_{t+1} = F_t^{-1}(U_{t+1}) = Q(U_{t+1}, X_t),$$

where $Q(U_{t+1}, X_t)$ is the conditional quantile corresponding to F given X_t. Note that the Us are independently generated random variables from $U(0, 1)$. To reformulate an MAR model, we therefore need to evaluate $F_t^{-1}(U_{t+1})$. Unfortunately, in general there is no closed-form expression for the quantile of a mixture of two

Gaussian distributions.

4 To smooth or not to smooth?

Chan and Tong (1986) proposed the smooth threshold autoregressive model and gave it the acronym *STAR* after the acronym TAR for the threshold autoregressive model. It takes the following form:

$$X_t = a_0 + \sum_{j=1}^{p} a_j X_{t-j} + (b_0 + \sum_{j=1}^{p} b_j X_{t-j}) F(\frac{X_{t-d} - r}{z}) + e_t,$$

where r is a real constant, $z \geq 0$, d (an integer) $\geq 1, p$ (an integer) $\geq 0, \{e_t\}$ is a sequence of iid random variables with zero mean and finite second moment, e_t is independent of $X_s, s < t$, and F is a continuous distribution function. Here, the real positive parameter z controls the smoothness of the function F such that a very small z corresponds to a steep switching of F at the threshold r. A popular choice of F, leading to the so-called LSTAR model in the econometrics literature, is, in its simplest form,

$$F(x) = (1 + e^{-x})^{-1}.$$

Chan and Tong (1986) proved essentially the following theorem, which shows that the STAR model includes the TAR model as a special case.

Theorem(2): Suppose $\{e_t\}$ is absolutely continuous with a bounded and uniformly continuous density function. Suppose $\{X_t\}$ is ergodic and strictly stationary. Then, for each $z \geq 0$, the invariant distribution of $\{X_{t,z}\}$ is absolutely continuous with density function denoted by g_z. Moreover, $g_z(y)$ is continuous in y, uniformly bounded and equicontinuous over z. Also, $g_z \to g_0$ everywhere as $z \to 0$.

They noted that the proof of Theorem 2 can easily be adapted for any sufficiently smooth function F with a rapidly decaying tail. This covers the other popular choice of F in econometrics, namely the so-called ESTAR model:

$$F(x) = 1 - e^{-zx^2}, z > 0.$$

They also proved the central limit theorem and the law of the iterated logarithm for the least squares estimates of unknown parameters of the STAR model, and studied the forecasting of the STAR model. In STAR modelling, the importance of giving the standard error of z cannot be over-emphasized. Chan and Tong (1986) gave the value 0.084 as their least squares estimate of the z parameter in their STAR model for the Australian blowfly data. They were conscious of the implications of its standard error which they gave as 0.045. As pointed out in Tong (2013), the implications are that 'unless there is *a priori* reason (e.g., from some underlying economics theory) or there is a large amount of data in the neighbourhood of the slope change, the choice between a STAR model and a TAR model is often a matter of convenience. The crux of the matter is that while regime change may be discernible in practice, the precise functional form of the change is a different story.'

Since the turn of the 21st century, a few econometricians have taken up the STAR ideas with much vigor, partial comprehension, frequently no acknowledgement and mysterious re-labeling of the letter T from *threshold* to *transition*. In their vigorous push of STAR modelling to the econometric community, sufficient caution has not always been exercised.

Recently, Ekner and Nejstgaard (2013) have re-examined two published applications of STAR models in the literature. After a careful re-examination of the profile likelihood function of z^{-1} of the STAR model fitted by Teräsvirta *et. al.* (2010) to the Wolf's sunspot numbers (1710-1979), Ekner and Nejstgaard (2013) found that 'the global maximum is actually the TAR model' whereas the STAR model adopted by Teräsvirta *et al.* is only a local maximum. Ekner and Nejstgaard also re-examined the model fitted by van Dijk *et al.* (2002) to the U.S. male unemployment rate (1968:6-1989:12). They found that for the STAR model, the profile likelihood of the γ parameter (equivalent to z^{-1}) is rather flat and the maximum occurs at a rather large value of γ. They suggested that 'a large and imprecise estimate of γ implies that the LSTAR model is effectively a TAR model.'

In reality, both the TAR model and the STAR model are most likely wrong models. A more relevant issue is how to estimate the parameters of a wrong model because the likelihood approach is typically predicated on the model being true. Although results are available concerning maximum likelihood of mis-specified models, e.g., White (1982), I would argue that a systematic *non-likelihood* based approach to fitting a wrong model deserves much more serious attention. For some recent attempts, see, e.g., Davies (2008) and Xia and Tong (2011). Such an approach can be very rewarding. For example, using a simple TAR model of order 3 for the Wolf's sunspot numbers, Xia and Tong (*op. cit.*) have demonstrated substantial improvement in terms of the fit of the sunspot cycle as well as the prediction performance for varying horizons, if the conventional likelihood function, which is a functional of one-step-ahead prediction errors, is replaced by a non-likelihood approach based on a functional of all-step-ahead prediction errors.

In fact, evidence suggests that even a substantive model can be wrong. An effective non-likelihood approach stands some chance in exposing the inadequacy of a substantive model. Xia and Tong (2011) have given an example.

5 Change points over state and over time

Change points over state and over time correspond to nonlinearity and nonstationarity respectively. They are also called structural breaks, which have attracted much attention in the econometrics literature. See, e.g., Aue and Horváth (2013) for a recent survey.

Consider the following example. Let e_t denote a Gaussian white noise with zero mean and unit variance, and t_0 be an unknown positive integer. Consider Model A and Model B:

(A) $X_t = e_t$, if $X_{t-1} < 3$; $10 + e_t$ otherwise;

(B) $X_t = e_t$, if $t < t_0$; $10 + e_t$ otherwise.

Here, Model A, being a TAR model, has a change point at the state value 3 and is nonlinear, while Model B has a change point at the time value t_0 and is nonstationary. Now, given just one realization, it is almost impossible to distinguish between the two models. This means that at a deeper level, to distinguish between nonlinearity (NL) and nonstationarity (NS) we may need more than one realization.

What is in common between NL and NS is the existence of change points. Therefore, to detect a change point is of particular relevance. For NS, we may perform the detection of a change point over time by examining the data in their chronological order. For NL, we may first arrange the data in descending (or ascending) order and perform the detection of a change point over states by examining the order statistics. The order statistics idea was actually implemented in FORTRAN codes in Appendix A10 and Appendix A11 in Tong (1983, pp 291-292).[3] Tsay (1989) apparently re-discovered this idea and called it *arranged autoregression.*

The hypothesis-test-based detection of change points over time has a vast literature, one of the most recent being, e.g., Dehling *et al.* (2013). For on-line detection based on Akaike's information criterion (AIC), see Ozaki and Tong (1975) and Kitagawa and Akaike (1978). The possibility of adapting the above procedures to detect change points over states is worth exploring. In this connection, Tong (2013) has discussed the need and one way of revising the penalty term in the application of AIC and its cousins when dealing with setups involving nuisance parameters.

Now, to fit TAR models to a time series of sample size say n, the standard method is based on a grid-search over the order statistics. However, this can be quite time consuming because it requires $O(n)$ least squares operations for each trial TAR model. The situation is particularly serious for panel threshold models. Consider the following TAR model of order p.

$$X_t = \beta_1' \xi_{t-1} I(X_{t-d} \leq r) + \beta_2' \xi_{t-1} I(X_{t-d} > r) + \varepsilon_t,$$

where $\xi_{t-1} = (1, X_{t-1}, \ldots, X_{t-p})', I(\cdot)$ is the indicator function, β_1 and β_2 are (column) vectors of coefficients, and ε_t is a real-valued martingale difference with respect to some increasing sequence of σ-fields generated by $\{(X_j, \varepsilon_j) : j \leq t\}$. Here, $p \geq 0$ and $d > 0$ are integers temporarily assumed known, and r is an unknown real constant that is assumed to lie in the bounded subset of the reals $[\underline{r}, \bar{r}]$. For simplicity, let $q = \max\{p, d\}, \mathbf{y} = (X_{q+1}, \ldots, X_n)', \varepsilon = (\varepsilon_{q+1}, \ldots, \varepsilon_n)'$ and $\mathbf{I}(s) = (a_{ij})_{(n-q)\times(p+1)}$ with $a_{ij} = I(X_{q+i-d} \leq r)$. Suppose we have the sample $\{x_1, \ldots, x_n\}$. Then the TAR model can be re-written (now replacing X_j by the observations x_j, all j) as

$$\mathbf{y} = \mathbf{X}_1(r)\beta_1 + \mathbf{X}_2(r)\beta_2 + \varepsilon,$$

[3]Kung-Sik Chan contributed much to the construction of the FORTRAN codes.

where $\mathbf{X}_1(r) \equiv \mathbf{X} * \mathbf{I}(r)$ and $\mathbf{X}_2(r) \equiv \mathbf{X} * \{\mathbf{I}(\infty) - \mathbf{I}(r)\}$ with $\mathbf{X}$ being an $(n - q) \times (p+1)$ matrix, whose i-th row is $(1, x_{q+i-1}, \ldots, x_{q+i-p})$. Here, '$*$' denotes the Hadamard product of matrices. Denote the parameter vector $(\beta_1', \beta_2', r)'$ by θ and its true value by θ_0. The TAR model is linear in β_i's and least squares estimation gives the squared error function

$$S_n(r) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_1(r)(\mathbf{X}_1(r)'\mathbf{X}_1(r))^{-1}\mathbf{X}_1(r)'\mathbf{y} - \mathbf{y}'\mathbf{X}_2(r)(\mathbf{X}_2(r)'\mathbf{X}_2(r))^{-1}\mathbf{X}_2(r)'\mathbf{y},$$

from which we obtain the least squares estimate of r by

$$\hat{r} = \arg \min_{r \in [\underline{r}, \overline{r}]} S_n(r).$$

Let

$$J_n(r) = S_n - S_n(r),$$

where $S_n = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Assuming that $\{\varepsilon_t\}$ and $\{X_t\}$ are both strictly stationary and ergodic, $EX_t^2 < \infty$ and X_t has a continuous density that is positive on the real line, Li and Tong (2013) have proved that

$$\sup_{r \in [\underline{r},\ \overline{r}]} \left| \frac{J_n(r)}{n} - J(r) \right| \xrightarrow{p} 0,$$

and $J(r)$ is strictly monotonically increasing in $[\underline{r},\ r_0]$ and strictly monotonically decreasing in $[r_0,\ \overline{r}]$ provided that $\beta_{10} \neq \beta_{20}$. The unimodality of $J(r)$ can be exploited to furnish a two-stage grid-search algorithm that reduces the computational cost of fitting models with change points to a level much below $O(n)$. Essentially, a crude search is followed by a fine search in the vicinity of a crude minimum. However, the unimodality property is unlikely to hold where there are two thresholds or more, for which case further research is necessary. One possible way is to divide the order statistics into segments similar to the way that Ozaki and Tong (1975) modelled nonstationary time series.

A member of the TAR model family is the open-loop TAR model (labeled TARSO in Tong and Lim (1980), in which the input time series is an exogenous time series. Much of the above discussion applies to a TARSO as well. Structural breaks in multiple time series seem not to have attracted much attention. Since a close-loop TAR model is a bivariate time series, it may be an obvious first candidate for attention in this respect.

6 Threshold unit root and catastrophe

Consider the simple TAR model

$$X_t = \{\alpha X_{t-1} + \gamma\} I(X_{t-1} < r) + \{\beta X_{t-1} + \delta\} I(X_{t-1} \geq r) + \varepsilon_t,$$

where $I(\cdot)$ denotes the indicator function, and $\{\varepsilon_t\}$ is a sequence of iid random variables, with zero mean and variance σ^2. This model is stationary and ergodic if $\alpha < 1, \beta < 1$, and $\alpha\beta < 1$. The behaviour outside this region is more delicate. See Chan *et al.* (1985). Pham *et al.* (1991) studied the case $\gamma = \delta =$ a known constant. Without loss of generality, let $r = 0$. Let (α_0, β_0) be the true value of (α, β), and $(\hat{\alpha}_n, \hat{\beta}_n)$ be the least squares estimates of α, β. Pham *et al.* (1991) proved that $(\hat{\alpha}_n, \hat{\beta}_n) \to (\alpha_0, \beta_0)$ almost surely if and only if one of the following conditions holds:

$$\alpha_0 \leq 1, \beta_0 \leq 1, \gamma = 0;$$

$$\alpha_0 < 1, \beta_0 \leq 1, \gamma < 0;$$

$$\alpha_0 \leq 1, \beta_0 < 1, \gamma > 0.$$

Moreover, for the case (α_0, β_0) lying on the boundary $\alpha_0\beta_0 = 1, \hat{\alpha}_n$ is strongly consistent.

As far as I know, Pham *et al.* (1991) was the first paper dealing with the estimation of a non-stationary TAR model. They also coined the name *nonlinear unit root.* Perhaps, it can also be called a *threshold unit root*, reflecting the fact that the above conditions include cases of unit root in one of the regimes. When both regimes have a unit root, the threshold unit root reduces to the conventional unit root. Threshold unit roots opened the possibility of exploring regime specific random walks. Ling (2009) attacked some of the unsolved problems mentioned in Pham *et al.*(1991). Karlsen and Tjøstheim (2001) considered the unit root problem in a nonlinear setting by way of null recurrent time series.

As I have repeatedly stressed, e.g. Tong and Lim (1980) and Tong (1990, 2011), the regimes of a TAR model can be defined in a very flexible way, so that the modeller can delineate them to best suit his/her purpose. For example, prompted by catastrophe theory, Tong and Lim (1980, p.283) used the following: $\{X_{t-1} - X_{t-2} \geq 0, X_{t-2} \leq r_1\} \cup \{X_{t-1} - X_{t-2} < 0, X_{t-2} \leq r_2\}$ as one regime and $\{X_{t-1} - X_{t-2} \geq 0, X_{t-2} > r_1\} \cup \{X_{t-1} - X_{t-2} < 0, X_{t-2} > r_2\}$ as another. Note that, on letting $r_1 \to \infty$ and $r_2 \to -\infty$, this model includes, as a special case, the momentum-TAR model used by Enders and Granger (1998) to characterize asymmetric growth rates as in monthly or quarterly unemployment rates; they called $X_{t-1} - X_{t-2}$ momentum.

7 Concluding discussion

Not all the ideas and aspirations in Tong and Lim (1980) have been fully explored in the literature. Let me mention a few that might be relevant to econometrics.

(1) It is conceivable that in a nonlinear economic system, structural breaks may be associated with what is called jumped resonance in nonlinear oscillations. In a nonlinear system driven by a periodic force, the output may exhibit sudden jumps as the frequency (or amplitude) of the driving force gradually increases/decreases. Simulated examples of jumped resonance are shown in Tong and Lim (1980, 252-254). Note that the jumps can occur at different frequencies depending on whether the driving force is increasing or decreasing. This may have very interesting economics interpretations. I am not aware of any systematic exploration of the phenomenon in the statistical time series literature or in econometrics.

(2) TAR models for multivariate time series are still in their infancy. While open-loop and close-loop threshold autoregressive systems (Tong and Lim, 1980) are early attempts, some fresh ideas and powerful optimization algorithms are still badly needed. Recently, panel time series analysis has been attracting much attention in the econometrics literature, (e.g., Baltagi, 2008). However, it seems that the threshold idea is only beginning to filter into this exciting area. See, e.g., Mitchell *et al.* (2012). In this regard, it is perhaps pertinent to draw attention to experiences in ecology. Stenseth *et al.* (1999) studied common dynamic structure of Canadian lynx populations within three climatic regions. Chan *et al.* (2004), found that the lynx data over Canada share similar dynamics in the decrease phase but they appear to be different in the increase phase. Yao *et al.* (2000) studied common threshold structure in panels of short ecological time series. Essentially, they showed that the mink-muskrat interactions shared some

common structure across 81 trapping stations in Canada. For further ecological experiences, see Stenseth (2009).

(3) Although conditional heteroscedasticity was recognized by TAR models not later than Tong and Lim (1980), a systematic development is available only recently in Chan *et al.* (2013), due to early pre-occupation with the conditional mean function. On hindsight, this represents a belated opportunity to model volatility from the threshold perspective. Be that as it may, exciting possibilities are still available. One of the many advantages of the threshold approach to volatility, under the acronym T-CHARM, is that only very mild conditions need to be imposed on the model parameters, unlike the situation with GARCH models. In its simplest form, a T-CHARM is given by

$$X_t = \sigma(X_{t-1})\eta_t,$$

where $\{\eta_t\}$ is a sequence of iid random variables (not necessarily Gaussian) with mean zero and unit variance, and the function $\sigma(x)$ is piecewise constant. Thus, X_t is a dynamic and interpretable mixing of distributions of possibly different variances. Recall that mixtures of distributions are a well tried way of modelling heteroscedasticity. The simplest form of $\sigma(x)$ is

$$\sigma(x) = \sigma_1 I(x \leq r) + \sigma_2 I(x > r), \quad \sigma_1 > 0, \sigma_2 > 0,$$

where I denotes an indicator function and r is a real constant, the threshold. Let ρ_k denote $corr(\sigma^2(X_t), \sigma^2(X_{t-k}))$. Chan *et al.* (2013) showed that the above T-CHARM is stationary and

$$\rho_k = \{Pr(\sigma_2\eta_t > r) - Pr(\sigma_1\eta_t > r)\}^k.$$

Note that on the parameters σ_1 and σ_2, only positivity is needed and no further condition is necessary to estimate ρ_k. Contrast this with the well-known GARCH(1,1) model

$$\sigma^2(X_t) = \sigma_t^2 = \alpha_0 + (\alpha\eta_t^2 + \beta)\sigma_{t-1}^2,$$

where α_0, α, β are all positive. For stationarity, we need to impose $\alpha + \beta < 1$, under which $\rho_{gk} = (\alpha + \beta)^k$. (The suffix g is added to signify the GARCH model.) Note that while ρ_{gk} is always positive and monotonically decreasing,

ρ_k is more flexible in that it can also alternate between positive and negative values. It is well known that to estimate ρ_{gk}, a fourth moment condition is needed: $2\alpha^2 + (\alpha + \beta)^2 < 1$, which is quite restrictive and can cause problems in applications. Chan *et al.* (2013) gave an example with real data.

(4) Another area that has not attracted as much attention as it deserves is the threshold moving average model or more generally threshold autoregressive/moving average model. For some recent attempts, see, e.g. Li *et al.* (2012) and the references therein. It should be recalled that the famous Exponential Weighted Moving Average filter is rooted in a moving average model and that Evgeny Evgenievich Slutsky created the moving average model in his famous paper in 1927 as an approach to business cycle theory. Will a threshold moving average model throw deeper insights on business cycles?

Acknowledgements

I thank Professor Kung-Sik Chan for comments and suggestions. The research was partially supported by the Distinguished Visiting Professorship at the University of Hong Kong and a Senior Visiting Fellowship at the University of Bologna, Italy.

References

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.*, **34**, 1-16.

Baltagi, B.H. (2008).*Econometric analysis of panel data.* New York: John Wiley and Sons.

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

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

Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. of Time Series Anal.*, **7**, 179-190.

Chan, K.S. and Tong, H. (2002). A note on the equivalence of two approaches for specifying a Markov process. *Bernoulli*, **8**, 117-122.

Chan, K.S., Tong, H. and Stenseth, N.C. (2004). Testing for common structures in a panel of threshold models. *Biometrics*, **60**, 225-232.

Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985). A multiple threshold AR(1) model. *J. Appl. Prob.*, **22**, 267-279.

Chan, K.S., Li, D., Ling, S. and Tong, H. (2013). On conditionally heteroscedastic AR models with thresholds. *Stat. Sinica.* (To appear).

Davies, P.L. (2008) Approximating data (with discussion). *J. of Korean Stat. Soc.*, **37**, 191-240.

Dehling, H., Rooch, A. and Taqqu, M.S. (2013) Non-parametric change-point tests for long-range dependent data. *Scan. J. Stat.*, **40**, 153-173.

Ekner, L.E. and Nejstgaard, E. (2013). Parmaeter identification in the logistic STAR model. Discussion paper No. 13-07, Department of Economics, University of Copenhagen.

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. and Econ. Stat.*, **16**, 304-311.

Hamilton, J.D. (1989). A new approach to economic analysis of nonstationary time series and the business cycle. *J. Econometrics*, **57**, 357-384.

Hansen, B.E. (2011). Threshold autoregression in economics. *Stat. and Its Interface*, **4**, 123–128.

Karlsen, H. and Tjøstheim, D. (2001) Nonparametric estimation in null recurrent time series models. *Ann. Stat.*, **29**, 372-416.

Kitagawa, G. and Akaike, H. (1978). A procedure for the modeling of non-stationary time series. *Ann. Inst. Statist. Math.*, **30**, 351-363.

Lawrance, A.J. and Kottegoda, N.T. (1977). Stochastic modelling of riverflow time series (with discussion). *J. Roy. Statist. Soc.*, A, **140**, 1-47.

Lawrance, A.J. and Lewis, (1980). The exponential autoregressive-moving average EARMA(p,q) process. *J Roy. Statist. Soc.*, B, **42**, 150-161.

Li, D. and Tong, H. (2013). A two-stage grid-search algorithm for fitting models with change points. (under preparation.)

Li, D., Ling, S. and Tong, H. (2012). On moving-average models with feedback. *Bernoulli*, **18**, 735-745.

Ling, S. (2009). Some remarks on Professor Tong's two papers. In *Exploration of a nonlinear world*, ed. Kung-Sik Chan, Singapore:World Scientific, 289-295.

Mitchell, J., Kapetanios, G. and Shin, Y. (2012). A nonlinear panel data model of cross-sectional dependence. *Working Paper No. 12/01*, Department of Economics, Univ. of Leicester, U.K.

Ozaki, T. and Tong, H. (1975). On the fitting of non-stationary autoregressive models in time series analysis. *Proc. 8th Hawaii Int. Conf. on System Sc.*, California: Western Periodicals, 225-226.

Pham, D.T., Chan, K.S. and Tong, H. (1991). Strong consistency of the least squares estimator for a non-ergodic threshold autoregressive model. *Stat. Sinica*, **1**, 361-369.

Rosenblatt, M. (1971). Markov Processes: Structure and Asymptotic Behavior. Berlin: Springer-Verlag.

Slutsky, E. (1927). The summation of random causes as a source of cyclical processes. *Voprosy Koniunktury* , **3**, 34-64. {English translation in *Econometrica*, **5**, 105-146.}

Smith, J.Q., Harrison, P.J. and Zeeman, E.C. (1981). The analysis of some discontinuous decision processes. *European J. Oper. Res.*, **7**, 30-43.

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., Chan, K.S., Tong, H., Boontra, R., Boutin, S., Krebs, C.J., Forchhammer, M.C., and Hurrell, J.W. (1999). Common dynamic structure of Canadian lynx populations within three climatic regions. *Science*, **285**, 1071-1073.

Teräsvirta, T., Tjøstheim, D. and Granger, C.W. (2010). *Modelling nonlinear economic time series.* Oxford: Oxford Univ. Press. Press.

Tong, H. (1977). Discussion on the Paper by Drs. Lawrance and Kottegoda. (1977). *J. Roy. Statist. Soc.*, series A, **140**, 34-35.

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

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

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

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

Tong, H. (2011). Threshold models in time series analysis – 30 years on. *Stat. and its Interface*, **4**, 107–118.

Tong, H. (2013). Review of *Modelling nonlinear ecoomic time series* by T. Teräsvirta, D. Tjøstheim and C.W.J. Granger, *J. Amer. Statist. Assoc.* (to appear).

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

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

van Dijk, D., Teräsvirta, T. and Franses, P.H. (2002). Smooth tranistion autoregressive models-a survey of recent developments. *Econometric Rev.*, **21**, 245-292.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **48**, 817-838.

Wong, C.S. and Li, W.K. (2002). On a mixture autoregressive model. *J. Roy. Statist. Soc.*, B, **62**, 95-115.

Xia, Y. and Tong, H. (2011). Feature matching in time series modeling (with discussion). *Stat. Sc.*, **26**, 21-46.

Yao, Q., Tong, H., Finkenstädt, B. and Stenseth, N.C. (2000). Common structure in panels of short ecological time-series. *Proc. Roy. Soc. London* series B, **267**, 2459-2467.

Department of Statistics, London School of Economics and Political Science.

E-mail: howell.tong@gmail.com

Current Departmental Research Reports

Serial No.	Date	Research Report Title	Author(s)
484 (Revised)	Jun-12	On conditionally heteroscedastic AR Models with thresholds	Kung-Sik Chan, Dong Li, Shiqing Ling and Howell Tong
496	Jul-12	A geometric process maintenance model and optimal policy	Yeh Lam
497	Oct-12	Test for homogeneity in gamma mixture models using likelihood ratio	T.S.T. Wong and W.K. Li
498	Oct-12	Buffered threshold autoregressive time series models	Guodong Li, Bo Guan, Wai Keung Li and Philip L.H. Yu
499	Nov-12	On a rescaled fractionally integrated GARCH model	Muyi Li, Wai Keung Li and Guodong Li
500	Nov-12	On the sphericity test with large-dimensional observations	Qinwen Wang and Jianfeng Yao
501	Nov-12	On generalized expectation based estimation of a population spectral distribution from high-dimensional data	Weiming Li and Jianfeng Yao
502	Nov-12	A note on multivariate CUSUM charts	Hualong Yang and Jianfeng Yao
503	Apr-13	A geometric process credibility model	Yeh Lam
504	Jun-13	On mixture double autoregressive time series models	Zhao Liu and Guodong Li
505	Jun-13	A simple formula for mixing estimators with different convergence rates	Stephen M.S. Lee and Mehdi Soleymani
506	Nov-13	On singular value distribution of large-dimensional autocovariance matrices	Zeng Li, Guangming Pan and Jianfeng Yao
507	Nov-13	Threshold models in time series analysis-some reflections	Howell Tong

❖❖❖ *The complete listing can be found at* ❖❖❖
http://www.saasweb.hku.hk/research/staff_research_report.php
Requests for off prints may be sent to saas@hku.hk by e-mail