ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Threshold models in time series analysis—Some reflections

Howell Tong

London School of Economics & Political Science, United Kingdom

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C22

Keywords:
All-step-ahead prediction
Asymmetry
Bayesian decision
Business cycle
Catastrophe
Conditionally heteroscedastic autoregressive models with thresholds
GARCH model
Hidden Markov chain
Hysteresis
Jump resonance
Markov switching model
Mis-specified model
Mixture of distributions
Non-likelihood approach
Nonlinear unit root
Non-stationarity
Open-loop system
Panel threshold model
Positive-valued time series
Smooth threshold autoregressive models
Splines
Stochastic volatility
Structural breaks
Threshold autoregressive models
Threshold moving average models
Threshold principle
Threshold unit root
Volatility
Wrong model

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.

1. Introduction

This paper focuses on univariate time series, although many of the key ideas are also relevant to 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 (Tong, 1977) 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 organized by the Research Section on 19th March 1980 . The paper has distinguished itself by having no serious theorems but being perhaps rich in ideas, some of which

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

http://dx.doi.org/10.1016/j.jeconom.2015.03.039
0304-4076/

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.

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 irreversibility, amplitude–frequency dependency, phase transition, chaos, deeper insights and others. (2) How to do it? 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 following as many years. In fact, even with the publication of Tong (1983, 1990), the threshold approach had to wait till the late 1990s before its leaps in growth.

On looking back, evidence suggests that it has achieved, to a larger or lesser extent, 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. For example, 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. Chen et al. (2011) have given a similarly extensive review of the threshold approach in finance. Stenseth (2009) has summarized the importance of the threshold autoregressive model for understanding the structure of ecological dynamics. Unfortunately, cross fertilization does not seem to be as widespread as it should be.

Tong (2011) has given a fairly broad coverage of the historical background and motivation that led to the introduction of the threshold models in time series analysis as well as a fairly systematic account of the development of these models in the past thirty years or so. Unlike Tong (2011), my reflections here will be much narrower but, it is hoped, deeper by focusing 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 underpinnings

Let $\{X_t : t = 0, \pm 1, \pm 2, \ldots\}$ 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 $\mu(x)$ is a 'smooth' function. From a purely deterministic perspective, we can approximate the function $\mu(x)$ arbitrarily closely by a series of step functions on invoking the Weierstrass theorem, as described in Tong and Lim (1980). Petruccelli (1992) has later shown rigorously that a threshold autoregressive model can almost surely approximate a general class of time series processes. Operationally speaking, we can consider at least two different ways to approximate $\mu(x)$. 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 $\mu(x)$ with some purpose in mind, e.g. forecasting, control, filtering, etc. A natural setting to proceed is to apply Bayesian decision theory by starting with an approximation in the form of a Bayesian linear model, i.e. $\mu(x) = \theta$, and with Gaussian belief:

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

The 'prior' slope c is expected to be a good approximation only over a local range of x, beyond which it will need to be shifted to a new value, say $c + \delta$ for some δ. This entails the introduction of a decision space say D, which is a collection of decisions; the decision of shifting c to $c + \delta$ is denoted by δ. Clearly we need some measure of closeness of the approximation. One convenient way to measure the closeness of the approximating linear model to the 'true' model is by the loss function that is conjugate to the Gaussian distribution, namely

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

where h and k are positive real constants. In many practical situations, we would not expect to have to invoke a large δ for a 'smooth' function $\mu(x)$. Put another way, a drastic decision increases the uncertainly of belief. This consideration leads to the definition of V as a function of δ by

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

where $\alpha, \beta > 0$. To evaluate the impact of the decision δ, we need to evaluate the expected loss $E_V(\delta)$ of making the decision δ, which is defined 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 after the decision δ has been employed. Here, $F_V(\theta|\delta)$ is $N(c+\delta, V(\delta))$. With $V(\delta)$ denoted as V and $\mu(x)$ as μ to simplify the notation,

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

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 to be made to c for those x for which $0 < \mu(x) - c < \{(1+\gamma^2)^{\frac{1}{2}} - 1\}\gamma^{-1}$, where $\gamma = \beta(k+\alpha)^{-\frac{1}{2}}$.

The above simple discussion provides the Bayesian decision theoretic underpinnings of the threshold approach to nonlinear time series analysis. It makes explicit the interplay between the principle of parsimony on the one hand and the purpose of modelling on the other. It is curious that the Bayesian underpinnings have gone totally unnoticed in the time series literature as well as the econometrics literature.