Estimation and Model Selection Based Inference in Single and Multiple Threshold Models[1]

Jesus Gonzalo

Universidad Carlos III de Madrid

Department of Statistics and Econometrics

E-mail: jgonzalo@elrond.uc3m.es

and

Jean-Yves Pitarakis

University of Reading

Department of Economics

E-mail: J.Pitarakis@rdg.ac.uk

September 2000

ABSTRACT

This paper evaluates the properties of a joint and sequential estimation procedure for estimating the parameters of single and multiple threshold models. We initially proceed under the assumption that the number of regimes is known à priori but subsequently relax this assumption via the introduction of a model selection based procedure that allows the estimation of both the unknown parameters and their number to be performed jointly. Theoretical properties of the resulting estimators are derived and their finite sample properties investigated.

Keywords: Threshold Models, Multiple Regimes, Model Selection.

JEL Classification: C1, C2

[1]The first author gratefully acknowledges the financial support of the Spanish Secretary of Education (DGYCIT PB98-0026). Address for Correspondence: Jean-Yves Pitarakis, University of Reading, Department of Economics, PO Box 218 Whiteknights, RG6 6AA Reading, United-Kingdom. Tel:+44-118-9875123 Ext. 7470, Fax: +44-118-9750236.

1 Introduction

The recent applied and theoretical econometrics literature has witnessed a growing interest in the class of threshold models characterized by piecewise linear processes separated according to the magnitude of a threshold variable. When each linear regime follows an autoregressive process for instance we have the well known threshold autoregressive family of models, the statistical properties of which have been investigated in early work by Tong and Lim (1980), Tong (1983, 1990), and more recently reconsidered and extended in Hansen (1996, 1997, 1999a, 1999b, 2000), Caner and Hansen (2000), Gonzalez and Gonzalo (1998) among others. Given their rich dynamic structure and their ability to capture nonlinearities and asymmetries within an intuitive mathematical framework, this class of nonlinear models has also generated a growing interest among economists interested in capturing economically meaningful nonlinearities. Examples include the analyis of asymmetries in persistence in the US output growth (Beaudry and Koop (1993), Potter (1995)), nonlinearities in unemployment rates (Hansen (1997), Koop and Potter (1998)), threshold effects in cross-country growth regressions (Durlauf and Johnson (1997)) and in international relative prices (Obstfeld and Taylor (1997), O'Connell and Wei (1997)) among numerous others. Although economic theory is often silent about the specific type of nonlinearities, it frequently suggests models with switching behavior as in the case of the speculative storage model recently analyzed in Ng (1993) or situations where macroeconomic variables such as output or employment present different dynamics according to the stage of the business cycle (see Koop, Pesaran and Potter (1996), Altissimo and Violante (1999)). It is also important to point out that the threshold family of models is only one among a multitude of other possible specifications able to capture nonlinearities in economic variables. The choice is typically dictated by the particular stylized facts the model is designed to capture as well as the availability of statistical tools for conducting inferences. Alternative formulations include Hamilton's regime switching model (Hamilton (1989)), the standard change-point model, bilinear processes, among numerous others (see Carrasco (1997) for an encompassing testing strategy covering a wide range of nonlinear specifications). Although the multitude of potential specifications may suggest that the threshold family of models is only a narrow subset, recently Petruccelli (1992) has shown that the latter may also be viewed as an approximation to a more general class of nonlinear processes.

Despite their ability to capture interesting asymmetric features and jump phenomena observed in economic and financial time series the use of threshold models in the applied economics literature has been quite limited when compared with specifications such as Hamilton's regime switching model. Among the significant problems encountered when modelling data with threshold type of models are the prohibitive computational costs when estimating specifications with more than two regimes and on the theoretical side the difficulties in tabulating the limiting distributions of LR type statistics for detecting single or multiple threshold effects. For the latter case for instance, inferences are nonstandard due to the well known unidentified nuisance parameters problem together with the fact that the relevant limiting distributions tend to depend on model specific moments, thus ruling out any general tabulation. Tsay (1989) proposed a very interesting graphical approach for detecting the number and location of the thresholds and more recently, Hansen (1996) has developed a general methodology for the treatment of the at most two regime case which to our knowledge is the only technique that can handle very general threshold models including SETAR's of any order but its applicability to models with possibly more than two regimes is unclear.

In this paper our aim is to focus on some of the above mentioned computational and theoretical difficulties by first formally establishing the large sample properties of a sequential estimation approach that makes the estimation of multiple-threshold models computationally feasible. We subsequently concentrate on the possibility of using an alternative to testing for a data based determination of the unknown number of regimes. The plan of the paper is as follows. Section II focuses on the sequential estimation of the parameters of a multiple threshold model under the assumption that the number of regimes is fixed and known. Section III extends the results to the case of an unknown number of regimes by investigating the properties of a model selection based approach for the joint determination of the threshold parameters and their number. Section III concludes. All proofs are relegated to the appendix.

2 Joint and Sequential Estimation under a known number of thresholds

We consider the following multiple threshold model expressed in matrix form

$$\mathbf{y} = \sum_{j=1}^{m+1} \mathbf{X}_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y}$ denotes the $T \times 1$ vector of observations on the dependent variable, $\mathbf{X}_j \equiv \mathbf{X} * \mathbf{I}(\gamma_{j-1} < z \leq \gamma_j)$ with $\mathbf{X}$ denoting the $T \times K$ matrix of regressors, $\boldsymbol{\beta}_j$ the corresponding $K \times 1$ vector of coefficients and $\mathbf{I}(\gamma_{j-1} < z \leq \gamma_j)$ is the stacked $T \times 1$ vector of indicator variables with z referring to the threshold variable that triggers the regime changes and $\boldsymbol{\epsilon}$ is a random disturbance term with zero mean and variance σ_ϵ^2. The corresponding threshold parameters are denoted $(\gamma_1, \ldots, \gamma_m)$ with $\gamma_0 = -\infty$, $\gamma_{m+1} = \infty$ and $*$ is the Hadamard product operator. The threshold variable z could be a component of the regressor matrix or a variable that is external to the system. Given data collected in $\mathbf{y}$, $\mathbf{X}$ and z, and assuming that the number of regimes is known, our objective is to estimate the regression coefficients together with the threshold parameters. Specifically the unknown $(m+1)K + m$ dimensional parameter vector is given by $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{m+1}, \gamma_1, \ldots, \gamma_m)$. It is also worth noting that within the specification in (1) we have $\mathbf{X} = \sum_{j=1}^{m+1} \mathbf{X}_j$ and the regressors are such that $\mathbf{X}_i'\mathbf{X}_j = \mathbf{0}\ \forall i \neq j$. Before proceeding with the estimation of $\boldsymbol{\theta}$ we introduce a set of preliminary assumptions, ensuring the estimability of the unknown parameter vector. We define $\mathbf{X}_\zeta = \mathbf{X} * \mathbf{I}(\gamma_j^0 - \zeta < z \leq \gamma_j^0)$ and $\bar{\mathbf{X}}_\zeta = \mathbf{X} * \mathbf{I}(\gamma_j^0 < z \leq \gamma_j^0 + \zeta)$ for a small $\zeta - neighborhood$ of each of the m true threshold parameters and $\forall j$.

Assumption A1 *(i) The threshold parameters are such that* $\gamma_i \in \Gamma_m$, $i = 1, \ldots, m$, *where* $\Gamma_m = \{(\gamma_1, \ldots, \gamma_m) : -\infty < \underline{\gamma} < \gamma_1 < \ldots < \gamma_m < \bar{\gamma} < \infty\}$, *(ii) The minimum eigenvalues of* $\mathbf{X}_\zeta'\mathbf{X}_\zeta/T$ *and* $\bar{\mathbf{X}}_\zeta'\bar{\mathbf{X}}_\zeta/T$ *are bounded away from zero and (iii) the threshold variable* z *has a positive density in* $[\underline{\gamma}, \bar{\gamma}]$.

Part *(i)* of the above assumption restricts all threshold parameters to lie in the bounded subset $[\underline{\gamma}, \bar{\gamma}]$ of the threshold variable sample space, *(ii)* ensures that there are enough observations in each regime for estimability purposes and *(iii)* rules out the possibility that two distinct threshold values produce the same fit. In practice the analysis is conducted by imposing an ad-hoc lower bound for the number of observations in each regime. Following Hansen (1999a) for instance and letting T_j

$\forall j = 1, \ldots, m+1$ denote the number of observations in regime j, it is common practice to require $T_j/T \geq \lambda$, with λ typically set to 10%. Conditional on $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$ the model in (1) is linear in the $\boldsymbol{\beta}_j's$ and thus the application of the least squares principle leads to the concentrated sum of squared errors function

$$S_T(\gamma_1, \ldots, \gamma_m) = \mathbf{y}'\mathbf{y} - \sum_{j=1}^{m+1} \mathbf{y}'\mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{y} \tag{2}$$

from which the threshold parameters can be estimated through the following optimization program

$$(\hat{\gamma}_1, \ldots, \hat{\gamma}_m) = \arg \min_{(\gamma_1, \ldots, \gamma_m) \in \Gamma_m} S_T(\gamma_1, \ldots, \gamma_m). \tag{3}$$

The slope parameter estimates can then be computed as $\hat{\boldsymbol{\beta}}_j = \hat{\boldsymbol{\beta}}_j(\hat{\gamma}_1^0, \ldots, \hat{\gamma}_m^0)$. We next introduce a set of high level assumptions (uniform LLN type results) which will allow us to establish the limiting properties of both the joint and sequential threshold parameter estimators. We let $(\gamma_1^0, \ldots, \gamma_m^0)$ denote the true configuration of threshold parameters and $\mathbf{X}_j^0 = \mathbf{X} * \mathbf{I}(\gamma_{j-1}^0 < z \leq \gamma_j^0)$ $\forall j = 1, \ldots, m+1$ refers to the corresponding regressor matrix.

Assumption A2 *As* $T \to \infty$, *uniformly over* $\gamma_j \in \Re$

(i) $\dfrac{\mathbf{X}_j'\mathbf{X}_j^0}{T} \xrightarrow{p} \Big[\mathbf{G}(\gamma_j \wedge \gamma_j^0) - \mathbf{G}(\gamma_{j-1} \wedge \gamma_j^0)\Big] - \Big[\mathbf{G}(\gamma_j \wedge \gamma_{j-1}^0) - \mathbf{G}(\gamma_{j-1} \wedge \gamma_{j-1}^0)\Big],$

(ii) $\dfrac{\mathbf{X}_j'\boldsymbol{\epsilon}}{T} \xrightarrow{p} \mathbf{0},$

(iii) $\dfrac{\mathbf{X}_j'\boldsymbol{\epsilon}}{\sqrt{T}} = O_p(1),$

where $\mathbf{G}(\gamma_j^0)$ *are symmetric positive definite matrices* $\forall j$ *and the* $\mathbf{G}(\gamma_j)'s$ *are symmetric positive definite matrices, absolutely continuous and strictly increasing functions of* γ_j, $\forall j = 1, \ldots, m+1$.

In what follows it will also be understood that $\mathbf{G}(\gamma_0^0 \wedge .) \equiv \mathbf{0}$, $\mathbf{G}(\gamma_0 \wedge .) \equiv \mathbf{0}$, $\mathbf{G}(\gamma_{m+1} \wedge \gamma_m^0) \equiv \mathbf{G}(\gamma_m^0)$, and $\mathbf{G}(\gamma_m \wedge \gamma_{m+1}^0) \equiv \mathbf{G}(\gamma_m)$. Within our notational conventions it is also implicit that $\mathbf{G}(\gamma_{m+1} \wedge \gamma_{m+1}^0) = \mathbf{G} \succ 0$ together with $\mathbf{G}(\gamma_{m+1}) \equiv \mathbf{G}(\gamma_{m+1}^0) \equiv \mathbf{G} \succ 0$. Thus an immediate consequence of assumption **A2**(i) is that $\mathbf{X}'\mathbf{X}/T \xrightarrow{p} \mathbf{G} \succ 0$. The above assumptions can be shown to hold under a wide range of specifications considered in applied work. If y is generated by

a SETAR process for instance then from Chan (1990), **A2** holds provided that the relevant characteristic polynomials have roots that lie inside the unit circle, the error process is iid with a bounded and continuous pdf and $E|\epsilon_t|^{4r} < \infty$ for some $r > 1$ (see also Hansen (1996, pp. 420-422) for a more general discussion on specifications under which **A2** holds). The limiting behaviour of the jointly estimated threshold parameters is summarized in the following proposition

Proposition 2.1 *As $T \to \infty$ and under A1 and A2(i)-(ii) we have $\hat{\gamma}_i \xrightarrow{p} \gamma_i^0$, $i = 1, \ldots, m$.*

The above joint estimators are straightforward to compute when the model is characterized by two regimes (m=1) since the optimization program requires a one-dimensional grid search only. When $m > 1$ however, the computational burden becomes substantial, requiring multi-parameter grid based simulations over all possible values of all threshold parameters taken together. The problem in hand is analogous to the computational problems arising when dealing with multiple-change point problems, recently investigated by Bai (1997), Bai and Perron (1998a, 1998b) and in the earlier work of Hawkins (1976) and Vostrikova (1981). In that literature it has been suggested that one may proceed sequentially by estimating the change-points one at a time since the change-point estimator obtained as an optimizer of a misspecified single parameter based objective function (derived from a fitted model with a single break while the true model contains more than one) maintains its consistency property for one of the true change-points. Given the similarities between threshold and change-point models, Hansen (1999) also conjectured that a similar feature should hold when fitting threshold models. To our knowledge however the recent literature does not provide any formal proof of the above result in the context of general threshold models such as the specification considered in (1) and even in the context of standard change-point models, the properties of the sequential estimation approach have only been established for simple mean shift models with no other included regressors (see Bai (1997) and Bai and Perron (1998)). Our next objective therefore is to formally establish the properties of threshold estimators obtained via a sequential estimation approach, requiring solely a single parameter grid search in each sequence. We concentrate on the limiting behaviour of a single threshold parameter estimate obtained from a fitted two-regime specification when the true model is given by (1). This will subsequentially allow us to formally establish the properties of a

sequential algorithm for estimating all the threshold parameters one at a time. The fitted model is given by

$$\mathbf{y} = \mathbf{Z_1}\delta_1 + \mathbf{Z_2}\delta_2 + \mathbf{u} \tag{4}$$

where $\mathbf{Z_1} = \mathbf{X} * I(z \leq r)$ and $\mathbf{Z_2} = \mathbf{X} * I(z > r)$ while the true model is specified as in (1). Note that $\mathbf{Z_1} + \mathbf{Z_2} = \mathbf{X}$ and $\mathbf{Z_1}'\mathbf{Z_2} = \mathbf{0}$. Applying the conditional least squares approach outlined previously to (4) leads to the following optimization program for the threshold parameter estimate

$$\hat{r} = \arg\min_{r \in \Gamma_1} S_T(r) \tag{5}$$

where

$$S_T(r) = \mathbf{y}'\mathbf{y} - \sum_{j=1}^{2} \mathbf{y}'\mathbf{Z}_j(\mathbf{Z}_j'\mathbf{Z}_j)^{-1}\mathbf{Z}_j'\mathbf{y} \tag{6}$$

and Γ_1 is the sample space of the threshold variable given by the "merged" version of Γ_m, i.e. $\Gamma_1 = [\underline{\gamma}, \bar{\gamma}]$. For greater technical convenience for what follows it is useful to define an alternative objective function $J_T(r) = S_T - S_T(r)$, with S_T denoting the sum of squared errors under the restriction $\boldsymbol{\beta}_1 = \ldots = \boldsymbol{\beta}_{m+1}$ in (1) and reformulate the optimization programme in (5) as

$$\hat{r} = \arg\max_{r \in \Gamma_1} J_T(r) \tag{7}$$

with

$$J_T(r) = (\hat{\boldsymbol{\delta}}_2 - \hat{\boldsymbol{\delta}}_1)'\mathbf{Z}_2'\mathbf{Z}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}_1'\mathbf{Z}_1(\hat{\boldsymbol{\delta}}_2 - \hat{\boldsymbol{\delta}}_1) \tag{8}$$

and where the $\hat{\boldsymbol{\delta}}'s$ in (8) denote the regime specific least squares estimators of the slopes for given r. Their dependence on the unknown threshold parameter is omitted for notational parsimony. The limiting behaviour of a properly normalized version of $J_T(r)$ is established in the following lemma

Lemma 2.1 *As $T \to \infty$ and under A1 and A2(i)-(ii) we have*

$$\sup_{r \in \Gamma} \left| \frac{J_T(r)}{T} - J_\infty(r) \right| \xrightarrow{p} 0$$

where $J_\infty(r)$ is nonstochastic continuous given by

$$\begin{aligned} J_\infty(r) = & \left[\sum_{\ell=1}^{m} \rho_\ell' \mathbf{G}(r \wedge \gamma_\ell^0)\mathbf{G}(r)^{-1} + \sum_{\ell=1}^{m} \rho_\ell' (\mathbf{G}(r \wedge \gamma_\ell^0) - \mathbf{G}(\gamma_\ell^0))(\mathbf{G} - \mathbf{G}(r))^{-1} \right] \\ & (\mathbf{G} - \mathbf{G}(r))\mathbf{G}^{-1}\mathbf{G}(r) \\ & \left[\mathbf{G}(r)^{-1} \sum_{\ell=1}^{m} \mathbf{G}(r \wedge \gamma_\ell^0)\rho_\ell + (\mathbf{G} - \mathbf{G}(r))^{-1} \sum_{\ell=1}^{m} (\mathbf{G}(r \wedge \gamma_\ell^0) - \mathbf{G}(\gamma_\ell^0))\rho_\ell \right] \end{aligned} \tag{9}$$

with $\boldsymbol{\rho}_\ell = (\boldsymbol{\beta}_\ell - \boldsymbol{\beta}_{\ell+1})$.

The above limit function $J_\infty(r)$ will have different expressions over the $m+1$ regimes. For $r = \gamma_k^0$ and $k = 1, \ldots, m$ for instance we have

$$
\begin{aligned}
J_\infty(r = \gamma_k^0) \quad = \quad & \left[\sum_{\ell=1}^{k} \rho_\ell' \mathbf{G}(\gamma_\ell^0) \mathbf{G}(\gamma_k^0)^{-1} + \sum_{\ell=k+1}^{m} \rho_\ell' (\mathbf{G} - \mathbf{G}(\gamma_\ell^0)(\mathbf{G} - \mathbf{G}(\gamma_k^0))^{-1} \right] \\
& (\mathbf{G} - \mathbf{G}(\gamma_k^0)) \mathbf{G}^{-1} \mathbf{G}(\gamma_k^0) \\
& \left[\mathbf{G}(\gamma_k^0)^{-1} \sum_{\ell=1}^{k} \mathbf{G}(\gamma_\ell^0) \rho_\ell + (\mathbf{G} - \mathbf{G}(\gamma_k^0))^{-1}) \sum_{\ell=k+1}^{m} (\mathbf{G} - \mathbf{G}(\gamma_\ell^0) \rho_\ell \right]. \qquad (10)
\end{aligned}
$$

Following the derivation of the uniform limit in (9), the most important subsequent step in the evaluation of the asymptotic properties of the extremum estimator defined in (7) involves establishing the existence of a unique maximum of $J_\infty(r)$. Since the uniform limit in (9) may have multiple local maxima we initially introduce an assumption ensuring that one of the true regimes dominates in the data and subsequently establish that $J_\infty(r)$ has a unique maximum.

Assumption A3 *There exists a single threshold parameter say* γ_s^0 *such that* $J_\infty(r = \gamma_s^0) > J_\infty(r = \gamma_k^0)$ $\forall k \neq s$ *and* $k = 1, \ldots, m$.

Given the expression of $J_\infty(r = \gamma_k^0)$ in (10) and assuming $m = 2$ with $s = 1$ (i.e. assuming that in a three regime model the first regime *dominates*) for instance the above assumption translates into the following requirement on the limiting objective function

$$
\begin{aligned}
J_\infty(\gamma_1^0) - J_\infty(\gamma_2^0) \quad = \quad & \rho_1' \mathbf{G}(\gamma_1^0) \mathbf{G}(\gamma_2^0)^{-1} (\mathbf{G}(\gamma_2^0) - \mathbf{G}(\gamma_1^0)) \rho_1 \\
- \quad & \rho_2' (\mathbf{G}(\gamma_2^0) - \mathbf{G}(\gamma_1^0))(\mathbf{G} - \mathbf{G}(\gamma_1^0))^{-1} (\mathbf{G} - \mathbf{G}(\gamma_2^0)) \rho_2 > 0. \qquad (11)
\end{aligned}
$$

Lemma 2.2 *Under A3 the limiting functional* $J_\infty(r)$ *in (9) is uniquely maximized at* $r = \gamma_s^0$.

The following two propositions focus on the consistency and rate of convergence of the threshold parameter estimator defined in (5) and (7).

Proposition 2.2 *As* $T \to \infty$ *and under A1, A2(i)-(ii) and A3 we have* $\hat{r} \xrightarrow{p} \gamma_s^0$.

Proposition 2.3 *As* $T \to \infty$ *and under A1, A2(i)-(ii) and A3 we have* $T|\hat{r} - \gamma_s^0| = O_p(1)$.

Propositions 2.1 and 2.3 establish that the single threshold parameter estimator obtain from a misspecified two regime model is T-consistent for the dominant true threshold from a model that has $m+1$ regimes. In practice since the *dominant* threshold parameter is unknown (it could correspond to any of the m true threshold parameters) we use subsamples $[\underline{\gamma}, \hat{r})$ and $(\hat{r}, \bar{\gamma}]$ to estimate the second threshold parameter. It is then straightforward to extend the arguments of the two propositions to show that this second threshold parameter will also be T-consistent for the *next dominant* threshold parameter. The above results can thus be used to implement a sequential algorithm for the estimation of all m parameters one at a time via a sequence of m one-dimensional optimization programmes as in (7) over appropriately defined search domains. Under $m = 2$ for instance and once the first stage estimator, say $\hat{r}^{(1)}$ has been obtained from (7)-(8), we proceed conditional on $\hat{r}^{(1)}$ and search for the second threshold by evaluating the second stage objective function over $r^{(2)} \in [\underline{\gamma}, \hat{r}^{(1)}) \cup (\hat{r}^{(1)}, \bar{\gamma}]$. Note that this latter objective function will have different expressions over the two sub-intervals since in practice it is not known whether the first stage estimate $\hat{r}^{(1)}$ is consistent for γ_1^0 or γ_2^0. Specifically, the second stage objective function can be formulated as

$$J_T(r^{(2)}) \quad = \quad J_T^L(r^{(2)}) I(r^{(2)} < \hat{r}^{(1)}) + J_T^R(r^{(2)}) I(r^{(2)} > \hat{r}^{(1)})$$

where $J_T^L(r^{(2)})$ and $J_T^R(r^{(2)})$ are analogous to (8) but derived from the following two canonical forms

$$Q^L y \quad = \quad \mathbf{Z_1^{(2)}} \delta_1 + \mathbf{Z_2^{(2)}} \delta_2 + \mathbf{u} \tag{12}$$

$$Q^R y \quad = \quad \mathbf{Z_1^{(2)}} \delta_1 + \mathbf{Z_2^{(2)}} \delta_2 + \mathbf{u}, \tag{13}$$

where $\mathbf{Z_1^{(2)}} = \mathbf{X} * \mathbf{I}(z \leq r^{(2)})$, $\mathbf{Z_2^{(2)}} = \mathbf{X} * \mathbf{I}(z > r^{(2)})$, and Q^L and Q^R are orthogonal matrices with orthogonal columns spanning the orthogonal complement of the column space of $\mathbf{X} * \mathbf{I}(z \leq \hat{r}^{(1)})$ and $\mathbf{X} * \mathbf{I}(z > \hat{r}^{(1)})$ respectively. Given the above notation, the consistency of the second stage threshold parameter estimator $\hat{r}^{(2)}$ can then be established in exactly the same manner as for $\hat{r}^{(1)}$. Although it is beyond our scope to concentrate on the limiting distributions of the threshold parameter estimators it is also important to mention that analogous to the change-point framework of Bai (1997), the first stage sequential estimator, denoted $\hat{r}^{(1)}$ above, will not have the same limiting distribution as its jointly estimated counterpart since the former has been estimated using a misspecified objective function contaminated by the wrongly omitted thresholds and as a result will be less efficient regardless of the sample size.

It is however possible to iterate the sequential procedure so as to make both sequentially estimated threshold parameters have the same asymptotic distribution as say $\hat{\gamma}_1$ and $\hat{\gamma}_2$ under $m = 2$. In the context of the above example this can be achieved simply by re-estimating the first stage threshold parameter $r^{(1)}$ taking $\hat{r}^{(2)}$ obtained in the second stage as given and subsequently re-estimating the second stage threshold parameter once more. This is the principle adopted in the analysis that follows.

2.1 Empirical Properties

Having established the consistency of the joint and sequential estimators, our next objective is to evaluate their relative behaviour in finite samples, viewing the joint estimation as the benchmark case. Our empirical results will also provide an overall picture of the finite sample behaviour and quality of estimators derived from threshold type specifications, features that to our knowledge have not been investigated in the recent time series literature and that are crucial for applied research. Given the computational burden that arises when dealing with models having more than three regimes we limit our analysis of the properties of the jointly estimated threshold parameters to models with at most two threshold parameters (three regimes).

Before proceeding with the empirical performance of the threshold parameter estimators however, it is important to highlight some difficulties that arise when designing a threshold type data generating process. The problem is related to the sensitivity of the variance of the estimators of the slopes (and implicitly that of the threshold parameter estimators) to the choice of the true threshold level. In a two regime (single threshold parameter) setup for instance one would expect to obtain more accurate estimates of both the threshold parameter and slopes if the true threshold parameter is set equal to the median or mean of the distribution of the threshold variable. In practice however it is often impossible to evaluate the moments of the threshold variable appearing in the DGP analytically making the interpretation of the resulting estimators (empirical bias, variance etc) extremely sensitive to the choice of the true threshold parameter. It is this latter aspect that we wish to initially illustrate by concentrating on a very simple DGP that lends itself to analytically tractable results. This will then allow us to achieve a fairer interpretation of our subsequent simulations based on richer dynamic structures. We initially consider the following two regime

model

$$y_t = \beta_1 I(y_{t-1} \leq \gamma_1) + \beta_2 I(y_{t-1} > \gamma_1) + \epsilon_t \tag{14}$$

where $\epsilon_t \equiv NID(0, \sigma_\epsilon^2)$ with σ_ϵ^2 set equal to 1 with no loss of generality. We also let γ_1^0 denote the true value of the threshold parameter and $\beta_1(\gamma_1)$, $\beta_2(\gamma_1)$ and $\sigma^2(\gamma_1)$ refer to the probability limits of $\hat{\beta}_1(\gamma_1)$, $\hat{\beta}_2(\gamma_1)$ and $\hat{\sigma}^2(\gamma_1)$ respectively. Letting $\Phi()$ denote the c.d.f. of a standard normal random variable and noting that $I(y_{t-1} \leq \gamma_1)$ is a Markov Chain, standard calculations using its transition matrix lead to $P(y_t \leq \gamma_1) = \Phi(\gamma_1 - \beta_2)/(1 - \Phi(\gamma_1 - \beta_1) + \Phi(\gamma_1 - \beta_2)) \equiv \pi(\gamma_1)$ from which it is straightforward to obtain

$$\beta_2(\gamma_1) - \beta_1(\gamma_1) = (\beta_2 - \beta_1) \frac{\pi(\gamma_1 \wedge \pi_1^0) - \pi(\gamma_1)\pi(\gamma_1^0)}{\pi(\gamma_1)(1 - \pi(\gamma_1))} \tag{15}$$

and

$$\begin{aligned} \sigma^2(\gamma_1) &= \sigma_\epsilon^2 + (\beta_2 - \beta_1)^2 \pi(\gamma_1^0)(1 - \pi(\gamma_1^0)) \\ &\quad -(\beta_2 - \beta_1)^2 \frac{[\pi(\gamma_1 \wedge \gamma_1^0) - \pi(\gamma_1)\pi(\gamma_1^0)]^2}{\pi(\gamma_1)(1 - \pi(\gamma_1)} \end{aligned} \tag{16}$$

where $\pi(\gamma_1 \wedge \gamma_1^0) = \pi(\gamma_1) I(\gamma_1 \leq \gamma_1^0) + \pi(\gamma_1^0) I(\gamma_1 > \gamma_1^0)$. From the expression of $\pi(\gamma_1)$ it is clear that under the above DGP we will have $\pi(\gamma_1) = 0.5$ when $\gamma_1 = 0.5(\beta_1 + \beta_2)$ also implying that $\gamma_1 = E(y_t)$. In other words choosing a true threshold parameter equal to the average of the parameters appearing in each regime ensures that it will also equal to the mean and median of the threshold variable, thus leaving an equal number of observations in both regimes. Our next objective is to evaluate the limiting behaviour of the variance of $\hat{\beta}_2(\gamma_1) - \hat{\beta}_1(\gamma_1)$. The latter should provide valuable information about the impact of the location of the true threshold parameter γ_1^0 on the estimators of the parameters. Standard calculations lead to

$$V_T(\hat{\beta}_2(\gamma\gamma_{1o} \; 13.00010g2o1.(o)1002()1.(o)100124.0001.0001(1)112(o).000111.()1412.0001.141\Gamma\hat{}2 \; JqT$$

to estimators with an extremely high variance, relative to the most favourable mean (or median) location. Under $\beta_1 = 1$ and $\beta_2 = 2$ for instance, the parabola is centered at $\gamma_1 = 1.5\ \forall \gamma_1^0$ with the corresponding variance equal to 4 while the variance corresponding to $\gamma_1^0 = 0$ for instance is close to 40, a ten-fold increase. In order to illustrate the usefulness of the above points we conducted a simulation experiment using the DGP in (14) and evaluated the empirical bias and variance of $\hat{\gamma}_1$ for different values of γ_1^0 together with the corresponding magnitudes for the slope estimates. Results are displayed in Table 1.

Table 1 about here

It is immediately clear that the threshold parameter estimate becomes highly inaccurate for values of γ_0 that fall outside the [1,2] range, with a typical greater than five-fold increase in its empirical standard deviation. The third and fourth columns of Table 1 display the empirical means and standard deviations of the resulting estimated slope parameters $\hat{\beta}_1(\hat{\gamma}_1)$ and $\hat{\beta}_2(\hat{\gamma}_1)$. It is interesting to note that the latter display a substantially smaller bias and a more stable variability when compared with that of the threshold parameter estimates. In summary the purpose of this preliminary exercise was to highlight the importance of experiment design when considering threshold type DGPs and that extreme caution should be taken when selecting the magnitude of γ_0. Ideally for results to give a sufficiently global picture it is an important imperative to scan across a wide range of possible true threshold parameter values since for models with richer dynamics, many of our analytical results would be unfeasible to obtain. We next concentrate on a similar specification with three regimes given by

$$(18) \qquad y_t = \beta_1 I(y_{t-1} \leq \gamma_1) + \beta_2 I(\gamma_1 < y_{t-1} \leq \gamma_2) + \beta_3 I(y_{t-1} > \gamma_2) + \epsilon_t.$$

Under the above true model and using standard but tedious algebra we have

$$(19) \qquad P(y_t \leq \gamma_1) = \frac{\Phi(\gamma_2 - \beta_3)\Phi(\gamma_1 - \beta_2) + \Phi(\gamma_1 - \beta_3)\Phi(\beta_2 - \gamma_2)}{\Delta(\gamma_1, \gamma_2)}$$

and

$$(20) \qquad P(\gamma_1 < y_t \leq \gamma_2) = \frac{\Phi(\gamma_2 - \beta_3)\Phi(\beta_1 - \gamma_1) - \Phi(\gamma_1 - \beta_3)\Phi(\beta_1 - \gamma_2)}{\Delta(\gamma_1, \gamma_2)}$$

where

$$\begin{aligned} \Delta(\gamma_1, \gamma_2) &= [\Phi(\gamma_1 - \beta_2) - \Phi(\gamma_1 - \beta_3)][\Phi(\beta_1 - \gamma_2) + \Phi(\gamma_2 - \beta_3)] \\ &+ [\Phi(\beta_2 - \gamma_2) + \Phi(\gamma_2 - \beta_3)][\Phi(\beta_1 - \gamma_1) + \Phi(\gamma_1 - \beta_3)] \end{aligned} \qquad (21)$$

and $P(y_t > \gamma_2) = 1 - P(y_t \leq \gamma_1) - P(\gamma_1 < y_t \leq \gamma_2)$. Our next objective therefore involves comparing the finite sample properties of the joint and sequential estimation approaches when applied to (18). We concentrate on DGPs given by (18) with $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$ and $\epsilon_t \equiv NID(0,1)$. The chosen threshold parameter structure encompasses a wide range of configurations leading to models with approximately equally divided regime proportions as well as models in which a single regime dominates. Specifically we consider $(\gamma_1^0, \gamma_2^0) = (1,2), (1.5,2.5), (1,3)$ and $(2,3)$ which using (19)-(21) imply regime proportions of approximately $(10\%, 20\%, 70\%)$, $(35\%, 30\%, 35\%)$, $(20\%, 60\%, 20\%)$ and $(70\%, 20\%, 10\%)$ respectively. All our experiments are performed using $T = 200$ across $N = 2000$ replications. The empirical means and corresponding standard deviations of the sequentially and jointly estimated threshold parameters together with the implied $\hat{\beta}'s$ are displayed in Table 2a.

Table 2a about here

As expected the precision of the estimates for both the joint and sequential approaches are highly sensitive to the location of the true threshold parameters with the most favourable scenario occuring when all three regimes have an approximately equal amount of observations. The increase in the variability of the threshold parameter estimators also translates into more imprecise estimated slopes with a quantitatively similar shift in magnitudes. When comparing both methods of estimation it is immediately apparent that the figures corresponding to the sequential and joint approaches are remarkably close, even for the moderately small sample size used in the experiment. Both the point estimates and their corresponding standard errors are virtually identical across all configurations of the true threshold parameters. Table 2b displays the results of a similar exercise using a SETAR(3;1,1,1) model given by $y_t = 0.2y_{t-1}I(y_{t-1} \leq -0.5) + 0.8y_{t-1}I(-0.5 < y_{t-1} \leq 0.5) - 0.5y_{t-1}I(y_{t-1} > 0.5) + \epsilon_t$. The choice of the true parameters is such that the regime proportions are approximately (40%, 35%, 25%).

Table 2b about here

For this scenario, results based on both T=200 and T=400 are presented. Focusing first on the relative behaviour of both estimation techniques it is again clear that they lead to estimates that remain very similar in terms of their finite sample variability and bias even in the context of models with richer dynamic structures. When evaluating the overall quality of the resulting estimators however and regardless of

the estimation technique it is important to note the drastic deterioration (in terms of loss of precision and finite sample bias) of both the threshold and slope estimates when moving from the simple threshold model with no conditional mean dynamics in each regime towards a more general SETAR process. In the latter case despite small finite sample biases the threshold parameters display a very high degree of variability which persists even as we move from T=200 to T=400.

3 Estimation under an unknown number of thresholds: A Sequential Model Selection Approach

In the preceding section our analysis was conducted under the assumption that the number of regimes of the threshold models is known. In practice however economic theory rarely offers an intuitive rationale for an à priori imposition of a specific number of regimes in the data. Numerous empirical applications aiming to describe the dynamics of macroeconomic variables have taken the ad-hoc view that two regimes may be appropriate for describing alternative dynamics for expansions and recessions. Others (e.g. Pesaran and Potter (1997), Koop and Potter (1999)) have argued that perhaps three regimes, encompassing bad times, good times *and* normal times should be modelled. Given this uncertainty it is then natural to inquire about data-based methods for the determination of the number of regimes. The literature on threshold models does not seem to offer any formal methodology for detecting the number of regimes in threshold type specifications, beyond the case involving testing single threshold versus linear models. In Chan (1990) for instance, the author obtained the limiting distribution of an LR type test statistic in the context of a general two regime SETAR model, but with the exception of a few special cases the limiting distribution does not lend itself to conventional tabulations due to its dependence on a large number of unknown parameters (e.g. moments of the regressors). More recently Hansen (1996), developed a bootstrap based procedure that allows the construction of asymptotically valid p-values for a large number of test statistics for the null of linearity versus two regimes. To our knowledge, Hansen's (1996) asymptotic p-value based approach is the only technique that allows the treatment of general threshold type models such as SETAR's of any order and its implementation is not restricted to models with simple dynamics. Although its validity is established for the treatment of the at most two regimes case it is not clear whether Hansen's (1996) can be

legitimately extended to a framework that allows the sequential determination of the number of regimes when the latter could be greater than two (see Hansen (1999)). Given the numerous unresolved difficulties arising in this context our objective here is to propose an alternative to sequential testing.

We propose to view the problem of specifying the number of regimes from a model selection perspective in which our main task is to select the optimal model among a portfolio of nested specifications and where the selection is made via the optimization of a penalized objective function. The objective function is such that one of its component is a monotonic function of the model dimension (e.g. the residual variance) and its other component penalizes the increase or decrease of the first component caused by the increase in the model dimension. Within our threshold framework the purpose of the penalty term is to penalize over-segmentation as m is allowed to increase. Formally, letting $S_T(\gamma_1, \ldots, \gamma_m)$ denote the concentrated sum of squared errors defined in (2), then in the spirit of the traditional model selection literature we introduce the following criterion

$$IC_T(\gamma_1, \ldots, \gamma_m) = \log\ S_T(\gamma_1, \ldots, \gamma_m) + \frac{\lambda_T}{T}[K(m+1)] \tag{22}$$

where λ_T is a deterministic function of the sample size (or a constant independent of T) that is in turn multiplied by the number of free parameters. Clearly an increase in m will lead to a reduction in $S_T(\gamma_1, \ldots, \gamma_m)$, a reduction that will be penalized due to the resulting increase in the number of estimated parameters. It is also important to observe that the minimization of the above objective function for given m will lead to the same estimates of the threshold parameters as in (3) since the penalty term does not depend on the magnitude of the threshold parameters. In a related study, Liu, Wu and Zidek (1997) also considered a criterion similar to (22) for the estimation of the number of threshold parameters. They used simulation based evidence to introduce a penalty term playing the role of λ_T in (22). Their analysis however is based on a direct joint estimation of the concentrated sum of squared errors function $S_T(\gamma_1, \ldots, \gamma_m)$ and differs from ours in its implementation and probabilitic framework.

Noting that under the linear specification the objective function, say $IC_T(0) = \log S_T + \frac{\lambda_T}{T}K$, does not depend on the threshold parameters we can introduce a

modified criterion defined as

$$Q_T(m) \;=\; IC_T(0) - \min_{\gamma_1,\ldots,\gamma_m} IC_T(\gamma_1,\ldots,\gamma_m)$$

or more specifically as

$$Q_T(m) \;=\; \max_{\gamma_1,\ldots,\gamma_m} \log\left[\frac{\hat{\sigma}^2}{\hat{\sigma}^2(\gamma_1,\ldots,\gamma_m)}\right] - \frac{\lambda_T}{T} K\ m. \tag{23}$$

The model selection based estimator of the number of unknown threshold parameters can then be formally defined as

$$\hat{m} \;=\; \arg\max_{0\leq m\leq M} Q_T(m) \tag{24}$$

for some upperbound M. Note that the threshold parameter estimates are implicitely obtained as a by-product of the above regime determination procedure. It is also useful to observe that T times the first component in the right hand side of (23) is the likelihood ratio statistic for testing linearity against $m+1$ regimes. Thus if we let $F_T(\gamma)$ denote any of the conventional LR, Score or Wald type test statistics we can also consider alternative versions of the objective function in (23) by introducing

$$\bar{Q}_T(m) \;=\; \max_{\gamma_1,\ldots,\gamma_m} F_T(\gamma_1,\ldots,\gamma_m) - \lambda_T K m, \tag{25}$$

as a more general version of $Q_T(m)$ in (23). This also suggests that the approach can accomodate the presence of heteroscedasticity via the use of heteroscedasticity robust versions of $F_T(.)$ in (25). We next concentrate on the theoretical and empirical properties of the model selection based estimates obtained as a solution to (24).

3.1 m=0 versus m=1 case

When our objective is to select between a linear and a two-regime specification we have $\hat{m} = \arg\max_{0\leq m\leq 1} Q_T(m)$. Recalling that $Q_T(0) = 0$ by construction the model selection procedure involves accepting the linear specification (m=0) if $Q_T(1) < Q_T(0)$ or equivalently if

$$IC_T(0) \;\leq\; \min_{\gamma_1\in\Gamma_1} IC_T(\gamma_1) \tag{26}$$

and decide for the threshold model when

$$IC_T(0) \;>\; IC_T(\gamma_1) \tag{27}$$

for some $\gamma_1 \in \Gamma_1$. Using the expressions of $IC_T(0)$ and $IC_T(\gamma_1)$ given above it is useful to note that the selection rule in (26) can be reformulated as

$$\max_{\gamma_1 \in \Gamma_1} \; T \; \log \left[\frac{\hat{\sigma}^2}{\hat{\sigma}^2(\gamma_1)} \right] \quad \leq \quad \lambda_T \; K \tag{28}$$

or equivalently as

$$\max_{\gamma_1 \in \Gamma_1} \; \frac{T(\hat{\sigma}^2 - \hat{\sigma}^2(\gamma_1))}{\hat{\sigma}^2(\gamma_1)} \quad \leq \quad \lambda_T K \tag{29}$$

since $T[\exp\left(\frac{\lambda_T K}{T}\right) - 1] \simeq \lambda_T K$. At this stage it is again interesting to note that the quantities appearing on the left hand side of (28) and (29) are conventional likelihood ratio and Wald type test statistics for the hypothesis of linearity versus a two regime threshold model. Their limiting distributions typically depend on unknown and model specific moments and cannot be tabulated. An important advantage of the model selection approach is that it does not rely on the critical values of the test statistics for deciding between the linear and threshold specifications. Instead the decision rule is based on the deterministic penalty term, solely function of the sample size multiplied by the number of free parameters. Equivalently when seen from a conventional testing perspective the above decision rule can be interpreted as using a test statistic in which the significance level is allowed to converge to zero as the sample size increases. Such a strategy has often been advocated when one performs a sequence of nested tests so as to avoid a build up of Type I errors or more generally to make the testing strategy lead to consistent estimates. We next show that the above model selection procedure leads to an estimator of m_0 that is weakly consistent. The result is summarized in the following proposition

Proposition 3.1 *Letting m_0 denote the true number of threshold parameters with $m_0 \in \{0, 1\}$, $\hat{m}$ defined as in (24) with λ_T such that (i) $\lambda_T \to \infty$ and (ii) $\frac{\lambda_T}{T} \to 0$ then under A1-A2(i)-(iii) we have $P(\hat{m} = m_0) \to 1$ as $T \to \infty$.*

The above proposition establishes that with probability tending to one and assuming that $m_0 \in \{0, 1\}$, the model selection procedure leads to an estimated number of threshold parameters that coincides with the true number provided that the penalty term satisfies conditions (i) and (ii). A possible candidate for the choice of the penalty term is $\lambda_T = \log T$ corresponding to a Schwarz type criterion but clearly the set of possible choices is extremely wide making it difficult to argue for an optimal penalty

choice. To our knowledge theoretical guidelines about specific choices of λ_T remain an open question in most frameworks that advocate the use of model selection criteria. Under some special classes of SETAR models however it is possible to gain further insight on the large and small sample behaviour of the model selection approach using the fact that the limiting distribution of the maximum LR or Wald type test statistics can be expressed in terms of the supremum of normalized squared Brownian Bridges. The availability of limiting results that may lead to proper tabulations of critical values is determined by the choice of the delay parameter in relation to the choice of the autoregressive lag length in each regime as well as the presence or absence of a constant term in the fitted models (see Tong (1990, pp. 239-246)). Here we consider the following two-regime model

$$y_t = \beta_1 \, I(y_{t-1} \leq \gamma_1) + \beta_2 \, I(y_{t-1} > \gamma_1) + \epsilon_t. \tag{30}$$

Following Chan (1990) the limiting distribution of the maximum likelihood ratio type statistics for testing $H_0 : \beta_1 = \beta_2 = \beta$ versus $H_1 : \beta_1 \neq \beta_2$ is given by

$$\max_{\gamma_1 \in \Gamma} F_T(\gamma_1) \rightarrow \max_{\delta \in \Delta} \frac{[B(\delta) - \delta B(1)]^2}{\delta(1-\delta)} \tag{31}$$

where $B(.)$ is a standard Brownian Motion and $\delta = P(y_t \leq \gamma_1)$ with $\Delta = [\delta_1, \delta_2]$ where $\delta_1 = P(y_t \leq \underline{\gamma})$ and $\delta_2 = P(y_t \leq \bar{\gamma})$. The above result can allow us to evaluate analytically the behaviour of the model selection approach outlined above. Assuming that the true model is characterized by $m_0 = 0$ (i.e. it is linear) and letting $P_\infty^{1|0}$ denote the limiting probability of falsely pointing to $m = 1$ we have

$$P_\infty^{1|0} = P\left[\max_{\delta \in \Delta} \frac{[B(\delta) - \delta B(1)]^2}{\delta(1-\delta)} > \kappa_\infty\right] \tag{32}$$

where $\kappa_\infty = \lim_{T\to\infty} T(e^{\frac{\lambda_T}{T}} - 1) \approx \lim_{T\to\infty} \lambda_T$. The above expression can also be reformulated as the absorption probability of an Ornstein-Uhlenbeck process, as

$$P[\max_{0 \leq t \leq \rho} |U(t)| > \sqrt{\kappa_\infty}], \tag{33}$$

where $\rho = \frac{1}{2}\log[\delta_2(1-\delta_1)/\delta_1(1-\delta_2)]$ and $U(t)$ is a zero mean Ornstein-Uhlenbeck process with covariance kernel $E[U(t)U(s)] = e^{|t-s|}$. Next using a result due to Dirkse (1975)[2], (33) can be approximated as

$$P[\max_{0 \leq t \leq \rho} |U(t)| > \sqrt{\kappa_\infty}] \approx \sqrt{\frac{2}{\pi}} e^{-\frac{\kappa_\infty}{2}} \sqrt{\kappa_\infty} \left[\rho - \frac{\rho}{\kappa_\infty} + \frac{2}{\kappa_\infty}\right]. \tag{34}$$

[2]See also Miller and Siegmund (1982) for a minor correction brought to Dirkse's (1974) original formulation.

From the above expression we can infer that the probability of over-segmentation under $\lambda_T = log\ T$ for instance (within this linear versus single threshold framework) will be converging to zero at the rate $O_p\left(\sqrt{\frac{\log T}{T}}\right)$. Note that this is in contrast with the standard multiple regression framework under which the BIC's probability of "overfitting" vanishes faster, at the rate $1/\sqrt{T}$. This suggests that in our threshold class of models, BIC based inferences will not necessarily lead to overly parsimonious choices (i.e. undersegmentation) in finite samples. It is also clear from (34) that if we were to choose an Akaike type penalty (i.e. $\lambda_T = 2$) we would be facing an extremely high probability of overfitting. Under $\kappa_\infty = 2$ for instance (corresponding to the Akaike penalty), $\beta = 0$ and 10% trimming in both sides of Γ (implying $\underline{\gamma} = -1.28$ and $\bar{\gamma} = 1.28$ under normality) the approximation in (34) leads to a spurious probability of pointing to the threshold model of approximately 87%. Although the above arguments were based on a simple model similar results also hold for a wider class of SETAR type specifications for which the limiting null distribution of the maximum LR statistic is again given by (31) albeit with a different structure of the set Δ. Indeed it is interesting to also consider the case where each regime is specified as an AR(1) process with no constant. Specifically we consider $y_t = \beta_1 y_{t-1} I(y_{t-1} \leq \gamma_1) + \beta_2 y_{t-1} I(y_{t-1} > \gamma_1) + \epsilon_t$ with the null given by $H_0 : \beta_1 = \beta_2 = \beta$. Under this scenario the limiting distribution of the LR type statistics is again given by (31) but with δ defined as $\delta(\gamma_1) = E(y_t^2 I(y_t \leq \gamma_1))/E(y_t^2)$. Since y_t is unconditionally normal, standard calculations lead to $\delta_1 = 0.3249$ and $\delta_2 = 0.6751$ under 10% trimming. Thus under $\kappa_\infty = 2$ the above approximation leads to a spurious probability of selecting the threshold specification of approximately 56.7%.

REMARK (**Spurious Thresholds**) Since under an AR(1) process we have $E(y_t^2) = \sigma_\epsilon^2/(1-\beta^2)$ it becomes immediately clear that for $\beta = 1$ (i.e. when the true DGP is given by a random walk) the conventional tests as well as the model selection based inferences will point to the threshold alternative with probability one since the set Δ in (31) will be shrinking to zero.

Our next objective is to evaluate the finite sample performance of the alternative criteria across a wider range of DGPs. We initially concentrate on linear models (i.e. $m_0 = 0$) and evaluate the performance (correct decision frequencies) of the various criteria when used for distinguishing between linearity and single threshold

type nonlinearity. For the choice of our specifications we follow our above discussion and consider an AR(1) process $y_t = \rho y_{t-1} + \epsilon_t$. The corresponding fitted threshold model is then given by $y_t = \rho_1^{(1)} y_{t-1} I(y_{t-1} \leq \gamma_1) + \rho_1^{(2)} y_{t-1} I(y_{t-1} > \gamma_1) + \epsilon_t$. Table 3a presents the correct decision frequencies (i.e. choosing $m = 0$ over $m = 1$) across three sample size ($T = 200, 400$ and $T = 600$) and where BIC, AIC, HQ, BIC2 and BIC3 refer to the model selection criteria with penalty terms $\lambda_T = \log T$, $\lambda_T = 2$, $\lambda_T = 2 \log\log T$, $\lambda_T = 2 \log T$ and $\lambda_T = 3 \log T$ respectively. The main motivation for the inclusion of the less familiar penalty terms labeled as BIC2 and BIC3 is to provide a sufficiently general description of the sensitivity of the model selection based decision frequencies to the magnitude of λ_T.

Table 3a about here

The results based on the AIC criterion confirm our previous theoretical analysis, with the criterion shown to point spuriously to the threshold model more than 50% of the times. This empirical frequency further deteriorates as the autoregressive parameter ρ approaches the unit root region. Similarly the HQ criterion, despite its ability to point to the true model asymptotically, is also performing poorly in moderately large samples by wrongly selecting the threshold model close to 30% of the times. As expected from proposition 3.2 the criterion improves its ability to point to the true model as the sample size grows but this latter improvement occurs very slowly reflecting the weakness of the HQ penalty. Among all model selection criteria the best performance is displayed by the BIC and its variants, denoted BIC2 and BIC3. Under $|\rho| < 1$ for instance and for reasonably large sample sizes the BIC is able to point to the linear model more than 93% of the times with a deterioration occuring only under the random walk model. Also, contrary to the linear regression framework the BIC does not appear to lead to spurious parsimonious choices. Both the BIC2 and BIC3 are pointing to the correct model with a probability close to 1. At this stage however the BIC2 and BIC3 based frequencies must be interpreted with caution since a close to 100% correct decision frequency might be due to a spurious choice of the most parsimonious structure due to the strength of the penalty terms characterizing BIC2 and BIC3.

We next consider a threhold DGP (i.e. $m_0 = 1$) of the form $y_t = \rho y_{t-1} I(y_{t-1} \leq 0) - \rho y_{t-1} I(y_{t-1} > 0) + \epsilon_t$ with $\rho = \{-0.40, -0.25, -0.15, -0.10, -0.05\}$. Note that as the magnitude of $|\rho|$ decreases, the existence of a two regime process will become more

and more difficult to detect. The empirical correct decision frequencies corresponding to this experiment are presented in Table 3b.

Table 3b about here

Taking into consideration the previously analyzed behaviour of the criteria under $m_0 = 0$, Table 3b suggests that the BIC and to a lesser extent the BIC2 display the best overall performance, with an excellent ability to point to the true model even for moderately small sample sizes. As expected, the ability of all criteria to point to the correct threshold model decreases with $|\rho|$ but even under $|\rho| = 0.15$ and T=600 the BIC is still able to select the true specification more than 93% of the times, compared with 60% for the BIC2.

3.2 General Case

Here we consider the case where there may be more than one threshold parameters (i.e. more than two regimes) in the set of possible models. Taking advantage of our general result on the consistency of the threshold parameter estimators in underspecified models we propose a sequential model selection based strategy for the estimation of the unknown number of threshold parameters, regardless of their number. Specifically the idea involves first proceeding as in the above section, deciding between a linear model (m=0) and a two regime threshold specification (m=1). If $Q_T(0) > Q_T(1)$ the procedure stops and we decide that the data support the linear model. If $Q_T(0) < Q_T(1)$ we obtain the estimate of the first threshold parameter, say $\hat{r}^{(1)}$ and conditional on this first stage threshold parameter estimator we proceed with a second stage $m = 0$ versus $m = 1$ decision process conducted on both subsamples in order to detect the eventual presence of a second threshold. The procedure continues until the model selection procedure leads to the choice $m = 0$ on all subsamples. More formally, letting $Q_T^{(i,j)}(1)$ denote the magnitude of (23) or (25) obtained in step i and subsample j, the stopping rule involves concluding for the presence of $m + 1$ regimes (or m threshold parameters) when $Q_T^{(m+1,j)}(1) < 0$ for all $j = 1, \ldots, m + 1$. The following proposition summarizes the asymptotic properties of the sequential threshold parameter estimator.

Proposition 3.2 *Letting $\hat{m}_{seq}$ denote the number of threshold parameters estimated via the sequential procedure with (i) $\lambda_T \to \infty$ and (ii) $\frac{\lambda_T}{T} \to 0$ then under A1-A3 we*

have $P(\hat{m}_{seq} = m_0) \to 1$ *as* $T \to \infty$.

At this stage it is important to reiterate that the sequential model selection based approach described above is very much similar to conducting a sequence of F tests for the determination of the number of threshold parameters but with the decision rule based on the penalty term rather than the critical values of the relevant limiting distributions. In the present multiple threshold framework however these distributions depend on all the parameters of the DGP and cannot be tabulated for practical purposes. In order to evaluate the finite sample behaviour of the sequential model selection based approach we conducted two sets of experiments using models with $m_0 = 1$ (two regimes) and $m_0 = 2$ (three regimes) respectively. We concentrate solely on the properties of the BIC and its two variants since our previous analysis demonstrated the unreliability of alternative criteria such as the AIC or HQ. Results corresponding to the two-regime specification are presented in Table 4a. Note first that the convergence of $\hat{m}$ to its true value $m_0 = 1$ is clearly visible across the increasing sample sizes, with the BIC detecting the true number of threshold parameters more than 90% of the times under T=600 and close to 95% of the times under T=800.

Tables 4a and 4b about here

It is also important to note that the procedure does not display any tendency to under-segment in the sense that the wrong decisions are mostly clustered at $\hat{m} = m_0 + 1$. An overall similar picture also arises from the results corresponding to a true model with three regimes (see Table 4b). The BIC and its variants do not display any tendency to under-segment and the wrong decisions are again clustered at $m_0 + 1$. Overall the BIC displays desirable large sample properties and a reasonably good finite sample behaviour. Obviously for the latter one should interpret any experimental result with caution since finite sample simulation based performance can be highly DGP specific. Under our DGP in Table 4a for instance, our choice of true parameter values is such that each regime has an approximately equal number of observations (50%). If we were to modify the magnitude of the slope and/or threshold parameters in such a way that one regime strongly dominates then it is natural to expect a deterioration in performance of the model selection criteria in small samples.

4 Conclusion

In this paper our objective was to provide a model selection based framework for estimating and conducting inferences in the context of multiple threshold models. We formally proved that estimating the threshold parameters one at a time leads to consistent estimates and subsequently investigated the asymptotic and finite sample properties of a model selection based approach for the determination of the number of regimes.

APPENDIX

As a matter of notation we refer to the fact that a symmetric matrix $\mathbf{A}$ is positive (semi) definite by writing $\mathbf{A} \succ (\succeq) 0$. More specifically matrix $\mathbf{A}$ is said to be larger than another symmetric matrix $\mathbf{B}$ if $\mathbf{A} - \mathbf{B} \succeq 0$. Equivalently, $\mathbf{A} \succeq \mathbf{B} \Leftrightarrow \mathbf{A} - \mathbf{B} \succeq 0 \Leftrightarrow x'\mathbf{A}x \geq x'\mathbf{B}x$ together with $\mathbf{A} \succ \mathbf{B} \Leftrightarrow \mathbf{A} - \mathbf{B} \succ 0 \Leftrightarrow x'\mathbf{A}x > x'\mathbf{B}x$.

PROOF OF PROPOSITION 2.1: We reparameterize the true specification as $y = \mathbf{W}\boldsymbol{\rho} + \mathbf{X}\boldsymbol{\beta}_{m+1} + \boldsymbol{\epsilon}$, with $\mathbf{W} = [\mathbf{X}_{\gamma_1^0}, \mathbf{X}_{\gamma_2^0}, \ldots, \mathbf{X}_{\gamma_m^0}]$, $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots, \boldsymbol{\rho}_m)'$, $\mathbf{X}_{\gamma_i^0} = \mathbf{X} * \mathbf{I}(z \leq \gamma_i^0)$ and $\boldsymbol{\rho}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}_{i+1}$, $\forall i = 1, \ldots, m$. The fitted model is given by (1) and we further define $\mathbf{M} = I - \sum_{i=1}^{m+1} \mathbf{P}_i$ with $\mathbf{P}_i = \mathbf{X}_i(\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i$. With the above notation and using straightforward algebra, the concentrated sum of squared errors function can be reformulated as $S_T(\gamma_1, \ldots, \gamma_m) = \boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\mathbf{W}\boldsymbol{\rho} + \boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon} + 2\boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\boldsymbol{\epsilon}$. Using assumptions **A2**(i)-(ii) we have $\frac{\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}}{T} = \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{T} + o_p(1)$ and $\frac{\boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\boldsymbol{\epsilon}}{T} = o_p(1)$ uniformly over $\gamma_i \in \boldsymbol{\Gamma}_m$, leading to

$$\frac{S_T(\gamma_1, \ldots, \gamma_m)}{T} - \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{T} = \frac{\boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\mathbf{W}\boldsymbol{\rho}}{T} + o_p(1). \tag{A.1}$$

Letting $R_T(\gamma_1, \ldots, \gamma_m) = S_T(\gamma_1, \ldots, \gamma_m) - S_T(\gamma_1^0, \ldots, \gamma_m^0)$ and using

$$\frac{S_T(\gamma_1^0, \ldots, \gamma_m^0)}{T} = \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{T} + o_p(1), \tag{A.2}$$

we can write

$$\frac{R_T(\gamma_1, \ldots, \gamma_m)}{T} = \frac{\boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\mathbf{W}\boldsymbol{\rho}}{T} + o_p(1). \tag{A.3}$$

Since $\boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\mathbf{W}\boldsymbol{\rho}$ can be rewritten as $(\mathbf{W}\boldsymbol{\rho})'\mathbf{M}(\mathbf{W}\boldsymbol{\rho})$ with $\mathbf{M}$ idempotent, we have $\boldsymbol{\rho}'\mathbf{W}'\mathbf{M}\mathbf{W}\boldsymbol{\rho} \succeq 0, \forall\ \gamma_i$ for $i = 1, \ldots, m$. Our next objective is to establish that the non-stochastic and continuous uniform probability limit of (A.3) say $\boldsymbol{\rho}' R_\infty(\gamma_1, \ldots, \gamma_m)\boldsymbol{\rho}$, reaches its unique minimum value of zero if and only if $(\gamma_1, \ldots, \gamma_m) = (\gamma_1^0, \ldots, \gamma_m^0)$. Letting $R_\infty^{ii}(\gamma_1, \ldots, \gamma_m)$ denote the diagonal components of $R_\infty(\gamma_1, \ldots, \gamma_m)$ and using assumptions **A2**(i)-(ii) we obtain

$$\begin{aligned} R_\infty^{ii}(\gamma_1, \ldots, \gamma_m) &= \mathbf{G}(\gamma_i^0) - \sum_{j=1}^{m+1} [\mathbf{G}(\gamma_i^0 \wedge \gamma_j) - \mathbf{G}(\gamma_i^0 \wedge \gamma_{j-1})][\mathbf{G}(\gamma_j) - \mathbf{G}(\gamma_{j-1})] \\ &\quad [\mathbf{G}(\gamma_i^0 \wedge \gamma_j) - \mathbf{G}(\gamma_i^0 \wedge \gamma_{j-1})] \end{aligned} \tag{A.4}$$

for $i = 1, \ldots, m$ and

$$\begin{aligned} R_\infty^{ij}(\gamma_1, \ldots, \gamma_m) &= \mathbf{G}(\gamma_i^0) - \sum_{j=1}^{m+1} [\mathbf{G}(\gamma_i^0 \wedge \gamma_j) - \mathbf{G}(\gamma_i^0 \wedge \gamma_{j-1})][\mathbf{G}(\gamma_j) - \mathbf{G}(\gamma_{j-1})] \\ &\quad [\mathbf{G}(\gamma_i^0 \wedge \gamma_j) - \mathbf{G}(\gamma_i^0 \wedge \gamma_{j-1})] \end{aligned} \tag{A.5}$$

for $j \neq i$, $j = 1, \ldots, m$ and where $\mathbf{G}(\gamma_i^0, \gamma_0) \equiv \mathbf{0}$, $\mathbf{G}(\gamma_i^0, \gamma_{m+1}) \equiv \mathbf{G}(\gamma_i^0)$ and $\mathbf{G}(\gamma_{m+1}) \equiv \mathbf{G}$. As a direct consequence of (A.4) and (A.5) we have $R_\infty^{ii}(\gamma_1, \ldots, \gamma_m) \succeq R_\infty^{ij}(\gamma_1, \ldots, \gamma_m)$ $\forall j \neq i$ and

$$\begin{aligned} R_\infty^{ii}(\gamma_1, \gamma_2, \ldots, \gamma_{i-1}, \gamma_i^0, \gamma_{i+1}, \ldots, \gamma_m) &= 0 \\ R_\infty^{ii}(\gamma_1^0, \gamma_2^0, \ldots, \gamma_{i-1}^0, \gamma_i, \gamma_{i+1}^0, \ldots, \gamma_m^0) &\succ 0 \end{aligned} \tag{A.6}$$

implying that $R_\infty(\gamma_1, \ldots, \gamma_m) = 0$ *if and only if* $\gamma_i = \gamma_i^0$ $\forall i = 1, \ldots, m$. Since $(\hat{\gamma}_1, \ldots, \hat{\gamma}_m) = \arg\min[R_T(\gamma_1, \ldots, \gamma_m)/T]$ and (A.3) converges uniformly in probability to the nonstochastic continuous functional $\boldsymbol{\rho} R_\infty(\gamma_1, \ldots, \gamma_m)\boldsymbol{\rho}$ that is uniquely minimized at $(\gamma_1^0, \ldots, \gamma_m^0)$ it follows from Theorem 2.2 of Newey and McFadden (1998) that $(\hat{\gamma}_1, \ldots, \hat{\gamma}_m) \xrightarrow{p} (\gamma_1^0, \ldots, \gamma_m^0)$.

PROOF OF LEMMA 2.1: We use the same parameterization as in the proof of Proposition 2.1. Using assumptions **A2**(i)-(ii) and the fact that $\mathbf{Z}_2'\mathbf{X} = \mathbf{Z}_2'\mathbf{Z}_2$ and $\mathbf{Z}_1'\mathbf{X} = \mathbf{Z}_1'\mathbf{Z}_1$ we can express $\hat{\boldsymbol{\delta}}_2 - \hat{\boldsymbol{\delta}}_1$ in (8) as

$$\hat{\boldsymbol{\delta}}_2 - \hat{\boldsymbol{\delta}}_1 = \left[\frac{\mathbf{Z}_2'\mathbf{Z}_2}{T}\right]^{-1}\left[\frac{\mathbf{Z}_2'\mathbf{W}}{T}\right]\boldsymbol{\rho} - \left[\frac{\mathbf{Z}_1'\mathbf{Z}_1}{T}\right]^{-1}\left[\frac{\mathbf{Z}_1'\mathbf{W}}{T}\right]\boldsymbol{\rho} + o_p(1). \tag{A.7}$$

and the result in (9) follows by noting from assumption **A2**(i) that

$$\sup_r \left| \frac{\mathbf{Z}_1'\mathbf{X}_{\gamma_i^0}}{T} - \mathbf{G}(\gamma_i^0 \wedge r) \right| \xrightarrow{p} 0, \tag{A.8}$$

and

$$\sup_r \left| \frac{\mathbf{Z}_2'\mathbf{X}_{\gamma_i^0}}{T} - (\mathbf{G}(\gamma_i^0) - \mathbf{G}(\gamma_i^0 \wedge r)) \right| \xrightarrow{p} 0. \tag{A.9}$$

LEMMA A.1: For $r \in (\gamma_1^0, \gamma_2^0)$ and letting

$$\begin{aligned} K_1 &= \mathbf{G}(\gamma_1^0)\mathbf{G}(r)^{-1}(\mathbf{G}(r) - \mathbf{G}(\gamma_1^0)) \\ K_2 &= \mathbf{G}(\gamma_2^0)\mathbf{G}(r)^{-1}(\mathbf{G}(r) - \mathbf{G}(\gamma_1^0)(\mathbf{G} - \mathbf{G}(\gamma_1^0))^{-1}(\mathbf{G} - \mathbf{G}(\gamma_2^0)) \\ M_1 &= \mathbf{G}(\gamma_1^0)\mathbf{G}(\gamma_2^0)^{-1}(\mathbf{G}(\gamma_2^0) - \mathbf{G}(\gamma_1^0)) \\ M_2 &= (\mathbf{G}(\gamma_2^0) - \mathbf{G}(\gamma_1^0))(\mathbf{G} - \mathbf{G}(\gamma_1^0))^{-1}(\mathbf{G} - \mathbf{G}(\gamma_2^0)), \end{aligned}$$

we have

(i) $M_1 \succ K_1$ and $M_2 \succ K_2$

(ii) $\forall x \neq 0, 0 < \frac{x'K_1x}{x'M_1x} < 1$

(iii) $\forall x \neq 0, z \neq 0$ and $x \neq z$, $\left[\frac{x'M_1x}{x'K_1x}\right]\left[\frac{z'K_2z}{z'M_2z}\right] \leq 1$

PROOF OF LEMMA A.1: (i) From assumption **A2**(i), $\mathbf{G}(r)$ is a continuous strictly increasing function of r. Since $r \in (\gamma_1^0, \gamma_2^0)$ it follows that $\mathbf{G}(\gamma_2^0) \succ \mathbf{G}(r)$ directly implying that $M_1 \succ K_1$. The result $M_2 \succ K_2$ follows using the same argument. (ii) Since $K_1 \prec M_1$ we have $K_1M_1^{-1} \prec I$ and therefore $\lambda^{max}(M_1^{-1}K_1) < 1$ which together with $\lambda^{min}(M_1^{-1}K_1) > 0$ implies the desired result. (iii) First note that $K_2M_2^{-1} = (\mathbf{G}(\gamma_2^0) - \mathbf{G}(\gamma_1^0))M_1^{-1}K_1(\mathbf{G}(\gamma_2^0) - \mathbf{G}(\gamma_1^0))^{-1}$, implying that $K_2M_2^{-1}$ and $M_1^{-1}K_1$ have the same characteristic roots. Next we have

$$\frac{z'K_2z}{z'M_2z} \leq \lambda^{max}(M_2^{-1}K_2)$$

and

$$\frac{x'M_1x}{x'K_1x} \leq \lambda^{max}(K_1^{-1}M_1) = \lambda^{max}((M_1^{-1}K_1)^{-1}) = [\lambda^{max}(M_1^{-1}K_1)]^{-1}$$

which implies the desired result.

PROOF OF LEMMA 2.2: With no loss of generality we provide the proof assuming $m = 2$ and setting $s = 1$ in the context of the requirements of assumption **A3**. The proof is in three parts. Since $J_\infty(r)$ takes different expressions over the three regions given by $[\underline{\gamma}, \gamma_1^0)$, (γ_1^0, γ_2^0) and $(\gamma_2^0, \bar{\gamma}]$, the result will follow by showing that the maximum of $J_\infty(r)$ cannot occur in any of the three regions in the sense that $J_\infty(\gamma_1^0) > J_\infty(r)$, $J_\infty(\gamma_2^0) > J_\infty(r)$, and the requirement that $J_\infty(\gamma_1^0) > J_\infty(\gamma_2^0)$. We start by treating the case $r \in (\gamma_1^0, \gamma_2^0)$. Using the expression of $J_\infty(r)$ in (9) and setting $m = 2$ we have

$$\begin{aligned}
J_\infty(\gamma_1^0) - J_\infty(r) &= \boldsymbol{\rho}_1'\mathbf{G}(\gamma_1^0)\mathbf{G}(r)^{-1}(\mathbf{G}(r) - \mathbf{G}(\gamma_1^0))\boldsymbol{\rho}_1 \\
&- \boldsymbol{\rho}_2'(\mathbf{G} - \mathbf{G}(\gamma_2^0))(\mathbf{G} - \mathbf{G}(r))^{-1}(\mathbf{G}(r) - \mathbf{G}(\gamma_1^0)) \\
&\quad (\mathbf{G} - \mathbf{G}(\gamma_1^0))^{-1}(\mathbf{G} - \mathbf{G}(\gamma_2^0))\boldsymbol{\rho}_2 \\
&> \boldsymbol{\rho}_1'K_1\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2'K_2\boldsymbol{\rho}_2 \\
&= \frac{\boldsymbol{\rho}_1'K_1\boldsymbol{\rho}_1}{\boldsymbol{\rho}_1M_1\boldsymbol{\rho}_1}\left[\boldsymbol{\rho}_1'M_1\boldsymbol{\rho}_1 - \frac{\boldsymbol{\rho}_1'M_1\boldsymbol{\rho}_1}{\boldsymbol{\rho}_1K_1\boldsymbol{\rho}_1}\frac{\boldsymbol{\rho}_2'K_2\boldsymbol{\rho}_2}{\boldsymbol{\rho}_2M_2\boldsymbol{\rho}_2}\boldsymbol{\rho}_2M_2\boldsymbol{\rho}_2\right] \\
&> \frac{\boldsymbol{\rho}_1'K_1\boldsymbol{\rho}_1}{\boldsymbol{\rho}_1M_1\boldsymbol{\rho}_1}\left[J_\infty(\gamma_1^0) - J_\infty(\gamma_2^0)\right] > 0. \qquad \text{(A.10)}
\end{aligned}$$

The first inequality follows by observing that $\mathbf{G}(\gamma_2^0)\mathbf{G}(r)^{-1} \prec (\mathbf{G} - \mathbf{G}(\gamma_2^0))(\mathbf{G} - \mathbf{G}(r))^{-1}$ since $r < \gamma_2^0$ and the last inequality follows from Lemma A.1 (iii) and the

fact that $J_\infty(\gamma_1^0) - J_\infty(\gamma_2^0) = \boldsymbol{\rho}_1 M_1 \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2 M_1 \boldsymbol{\rho}_2$ obtained in (11) and thus implying that the maximum of $J_\infty(r)$ cannot occur in (γ_1^0, γ_2^0). We next concentrate on the case $r < \gamma_1^0$. Using (9) and standard algebra we can write

$$J_\infty(\gamma_1^0) - J_\infty(r) \quad = \quad \mathbf{w}'(\mathbf{G} - \mathbf{G}(\gamma_1^0))^{-1}(\mathbf{G}(\gamma_1^0) - \mathbf{G}(r))(\mathbf{G} - \mathbf{G}(r))^{-1}\mathbf{w} \tag{A.11}$$

with $\mathbf{w} = [(\mathbf{G} - \mathbf{G}(\gamma_1^0))\boldsymbol{\rho}_1 + (\mathbf{G} - \mathbf{G}(\gamma_2^0))\boldsymbol{\rho}_2]$. Next note that $\mathbf{w} = 0$ implies $J_\infty(\gamma_1^0) < J_\infty(\gamma_2^0)$ which is ruled out by assumption, thus $\mathbf{w} \neq 0$ and therefore the above quadratic form is strictly positive, implying that the maximum of $J_\infty(r)$ cannot occur for $r < \gamma_1^0$. The treatment of the case $r > \gamma_2^0$ is identical.

PROOF OF PROPOSITION 2.2: The result follows from Lemmas 2.1, 2.2 and using Theorem 2.2 of Newey and McFadden (1998).

PROOF OF PROPOSITION 2.3: We proceed using the same simplifications as in the proof of Lemma 2.2, setting $m = 2$ and $s = 1$ with the true model given by $\mathbf{y} = \mathbf{X}_1^0\boldsymbol{\beta}_1 + \mathbf{X}_2^0\boldsymbol{\beta}_2 + \mathbf{X}_3^0\boldsymbol{\beta}_3 + \boldsymbol{\epsilon}$. To establish the T-consistency of $\hat{r}$ it suffices to show that $S_T(r) - S_T(\gamma_1^0) > 0$ for $T|r - \gamma_1^0|$ sufficiently large (see Chan (1993)). From Proposition 2.2 we operate in a $\epsilon-$neighborhood of γ_1^0 and treat the case $r < \gamma_1^0$. Formally we establish that for every $\nu > 0$, there exists an $0 < M < \infty$ such that for all T large we have

$$P\left[\min_{\frac{M}{T} < (\gamma_1^0 - r)} S_T(r) - S_T(\gamma_1^0) \leq 0\right] \quad < \quad \nu. \tag{A.12}$$

We initially write $S_T(r) - S_T(\gamma_1^0) = (S_T(r) - S_T(r, \gamma_1^0)) - (S_T(\gamma_1^0) - S_T(r, \gamma_1^0))$ where $S_T(r, \gamma_1^0)$ denotes the concentrated sum of squared errors function from the following auxiliary specification

$$\mathbf{y} \quad = \quad \mathbf{Z}_1\boldsymbol{\rho}_1 + \mathbf{X}_{r\gamma_1^0}\boldsymbol{\rho}_2 + \bar{\mathbf{X}}_{\gamma_1^0}\boldsymbol{\rho}_3 + \mathbf{u} \tag{A.13}$$

with $\mathbf{X}_{r\gamma_1^0} = \mathbf{X} * I(r < z \leq \gamma_1^0)$ and $\bar{\mathbf{X}}_{\gamma_1^0} = \mathbf{X} * I(z > \gamma_1^0)$. We can therefore write

$$\begin{aligned}\frac{S_T(r) - S_T(\gamma_1^0)}{T(\gamma_1^0 - r)} \quad = \quad & (\hat{\boldsymbol{\rho}}_3 - \hat{\boldsymbol{\rho}}_2)' \left[\frac{\mathbf{X}'_{r\gamma_1^0}\mathbf{X}_{r\gamma_1^0}(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\bar{\mathbf{X}}'_{\gamma_1^0}\bar{\mathbf{X}}_{\gamma_1^0}}{T(\gamma_1^0 - r)}\right](\hat{\boldsymbol{\rho}}_3 - \hat{\boldsymbol{\rho}}_2) - \\ & (\hat{\boldsymbol{\rho}}_2 - \hat{\boldsymbol{\rho}}_1)\left[\frac{\mathbf{Z}_1'\mathbf{Z}_1(\mathbf{X}_1^{0\prime}\mathbf{X}_1^0)^{-1}\mathbf{X}'_{r\gamma_1^0}\mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)}\right](\hat{\boldsymbol{\rho}}_2 - \hat{\boldsymbol{\rho}}_1),\end{aligned}$$

and using $\mathbf{X}'_{r\gamma_1^0}\mathbf{X}_{r\gamma_1^0}(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\bar{\mathbf{X}}'_{\gamma_1^0}\bar{\mathbf{X}}_{\gamma_1^0} = \mathbf{X}_{r\gamma_1^0}(\mathbf{I} - \mathbf{P}_{Z_2})\mathbf{X}_{r\gamma_1^0}$ together with

$$\|(\mathbf{Z}_1'\mathbf{Z}_1)(\mathbf{X}_1^{0\prime}\mathbf{X}_1^0)^{-1}\| < 1$$

since $r < \gamma_1^0$, we have

$$\frac{S_T(r) - S_T(\gamma_1^0)}{T(\gamma_1^0 - r)} \geq (\hat{\rho}_3 - \hat{\rho}_2)' \left[\frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right] (\hat{\rho}_3 - \hat{\rho}_2) - (\hat{\rho}_3 - \hat{\rho}_2)' \left[\frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{P}_{Z_2} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right] (\hat{\rho}_3 - \hat{\rho}_2) -$$

$$(\hat{\rho}_2 - \hat{\rho}_1) \left[\frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right] (\hat{\rho}_2 - \hat{\rho}_1). \tag{A.14}$$

Next it is straightforward to show that

$$\hat{\boldsymbol{\rho}}_2 - \hat{\boldsymbol{\rho}}_1 \xrightarrow{p} 0 \tag{A.15}$$

and

$$\hat{\boldsymbol{\rho}}_3 - \hat{\boldsymbol{\rho}}_2 \xrightarrow{p} (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1) + (\mathbf{G} - \mathbf{G}(\gamma_1^0))^{-1} (\mathbf{G} - \mathbf{G}(\gamma_2^0))(\boldsymbol{\beta}_3 - \boldsymbol{\beta}_2). \tag{A.16}$$

Thus from (A.15) and since under our assumptions $\left\| \frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right\| = O_p(1)$ the third term in the right hand side of (A.14) can be made arbitrarily small. Similarly since we are operating with r in a small neighborhood of γ_1^0 it follows that the second term on the right hand side of (A.14) can also be made arbitrarily small. This follows from

$$\left\| \frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{P}_{Z_2} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right\| \leq \left\| \frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{Z}_2}{T(\gamma_1^0 - r)} \left(\frac{\mathbf{Z}_2' \mathbf{Z}_2}{T} \right)^{-1} \right\| \left\| \frac{\mathbf{Z}_2' \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right\| (\gamma_1^0 - r).$$

Finally we have

$$(\hat{\boldsymbol{\rho}}_3 - \hat{\boldsymbol{\rho}}_2)' \left[\frac{\mathbf{X}'_{r\gamma_1^0} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right] (\hat{\boldsymbol{\rho}}_3 - \hat{\boldsymbol{\rho}}_2) \geq \lambda^{min} \left[\frac{\mathbf{X}_{r\gamma_1^0} \mathbf{X}_{r\gamma_1^0}}{T(\gamma_1^0 - r)} \right] \|\boldsymbol{\rho}_3 - \boldsymbol{\rho}_2\|^2. \tag{A.17}$$

Since by assumption the minimum eigenvalue of the moment matrix taken in the neighborhood of γ_1^0 is *strictly* positive and given the result in (A.16) it follows that $S_T(r) - S_T(\gamma_1^0) > 0$ on the relevant set, thus establishing the required result.

PROOF OF PROPOSITION 3.1: We first consider the case $m_0 = 0$ and prove that $P(\hat{m} = 1) \to 0$ as $T \to \infty$, which by (26) is equivalent to $P[IC_T(0) > IC_T(\gamma_1)] \to 0$ for some $\gamma_1 \in \Gamma_1$, thus implying that the procedure does not oversegment asymptotically. Using (23) we write

$$\begin{aligned} P[IC_T(0) > IC_T(\gamma_1)] &\leq P[IC_T(0) > \min_{\gamma_1 \in \Gamma_1} IC_T(\gamma_1)] \\ &= P\left[\max_{\gamma_1 \in \Gamma_1} T \log \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2(\gamma_1)} \right) > \lambda_T K \right]. \end{aligned} \tag{A.18}$$

Assuming with no loss of generality that $|\hat{\sigma}^2 - \hat{\sigma}^2(\gamma_1)|/\hat{\sigma}^2$ is small and using $\log(1+x) \simeq x$ together with $\hat{\sigma}^2(\gamma_1) \xrightarrow{p} \sigma_\epsilon^2$ and $\hat{\sigma}^2 \xrightarrow{p} \sigma_\epsilon^2$ implied by the LLN, we can reformulate (A.12) as

$$P[IC_T(0) > IC_T(\gamma_1)] \quad \leq \quad P[\max_{\gamma_1 \in \Gamma_1} F_T(\gamma_1) > \lambda_T K] \tag{A.19}$$

with $F_T(\gamma_1) = T(\hat{\sigma}^2 - \hat{\sigma}^2(\gamma_1))/\hat{\sigma}^2$. Next, under $m_0 = i0ea$nd when the fitted m9141Tfell

Proof of Proposition 3.2 We first show that the event $\{\hat{m} > m_0\}$ cannot occur as $T \to \infty$. Let $Q_T^{(i,j)}(1)$ denote the value of (23) evaluated in step i for subsample j. For the sequential model selection procedure to stop at m_0 (assuming all previous decisions to be correct since $\hat{m} > m_0$) it is required that $Q_T^{(m_0+1,j)}(1) < 0 \ \forall j = 1, 2, \ldots, m_0 + 1$. Thus the occurence of the event $\{\hat{m} > m_0\}$ implies the existence of at least one $j \in \{1, 2, \ldots, m_0 + 1\}$ for which $Q_T^{(m_0+1,j)}(1) > 0$. We can therefore write

$$P[\hat{m} > m_0] \quad \leq \quad \sum_{j=1}^{m_0+1} P[Q_T^{(m_0+1,j)}(1) > 0]$$

and

$$P[\hat{m} > m_0] \quad \leq \quad \sum_{j=1}^{m_0+1} P[\max_{r^{(m_0+1)} \in (\hat{r}^{(j-1)}, \hat{r}^{(j)})} F_T^{(j)}(r^{(m_0+1)}) > \lambda_T K] \to 0$$

provided that $\lambda_T \to \infty$ and where it is understood that $\hat{r}^{(0)} \equiv \underline{\gamma}$ and $\hat{r}^{(m_0+1)} \equiv \bar{\gamma}$. The case $\{\hat{m} < m_0\}$ follows in exactly the same manner as in Proposition 3.1 since in any subsample that has at least one threshold we have $\hat{\sigma}^2 - \hat{\sigma}^2(\hat{r}) \xrightarrow{p} C > 0$.

Table 1: Empirical Mean and Standard Deviation of Estimators

DGP: $y_t = \beta_1 I(y_{t-1} \leq \gamma_1^0) + \beta_2 I(y_{t-1} > \gamma_1^0) + \epsilon_t$

$\beta_1 = 1,\ \beta_2 = 2,\ T = 250\ \ and\ N = 5000$

γ_1^0	$\hat{\gamma}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.0	1.612	1.744	2.020
	(0.920)	(0.305)	(0.196)
0.5	0.788	1.180	2.018
	(0.526)	(0.347)	(0.107)
1.0	0.996	0.993	2.006
	(0.133)	(0.156)	(0.086)
1.5	1.491	0.995	2.004
	(0.096)	(0.105)	(0.105)
2.0	1.985	0.994	2.007
	(0.124)	(0.084)	(0.158)
2.5	2.154	0.984	1.805
	(0.564)	(0.112)	(0.364)
3.0	1.317	0.988	1.216
	(0.897)	(0.198)	(0.302)

Table 2a: Empirical Mean and Standard Deviation of Estimators

DGP: $y_t = \beta_1 I(y_{t-1} \leq \gamma_1^0) + \beta_2 I(\gamma_1^0 < y_{t-1} \leq \gamma_2^0) + \beta_3 I(y_{t-1} > \gamma_2^0) + \epsilon_t$

$\beta_1 = 1,\ \beta_2 = 2,\ \beta_3 = 3,\ T = 200,\ N = 2000.$

	Sequential Estimation				
(γ_1^0, γ_2^0)	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
(1,2)	1.244	2.230	1.183	2.211	3.002
	(0.364)	(0.539)	(0.379)	(0.445)	(0.136)
(1.5,2.5)	1.480	2.495	0.995	2.003	3.005
	(0.127)	(0.132)	(0.125)	(0.163)	(0.128)
(1,3)	0.996	2.973	0.998	2.001	3.002
	(0.138)	(0.146)	(0.174)	(0.104)	(0.175)
(2,3)	1.682	2.655	1.003	1.718	2.763
	(0.581)	(0.381)	(0.144)	(0.479)	(0.396)
	Joint Estimation				
(1,2)	1.247	2.213	1.186	2.106	3.003
	(0.360)	(0.512)	(0.372)	(0.434)	(0.143)
(1.5,2.5)	1.479	2.493	0.992	1.997	3.011
	(0.126)	(0.124)	(0.126)	(0.152)	(0.123)
(1,3)	0.993	2.972	0.986	1.997	3.006
	(0.144)	(0.169)	(0.175)	(0.116)	(0.181)
(2,3)	1.686	2.711	1.013	1.715	2.699
	(0.579)	(0.391)	(0.132)	(0.472)	(0.400)

Table 2b: Empirical Mean and Standard Deviation of Estimators

DGP: $y_t = \beta_1 y_{t-1} I(y_{t-1} \leq \gamma_1^0) + \beta_2 y_{t-1} I(\gamma_1^0 < y_{t-1} \leq \gamma_2^0) + \beta_3 y_{t-1} I(y_{t-1} > \gamma_2^0) + \epsilon_t$

$\beta_1 = 0.2,\ \beta_2 = 0.8,\ \beta_3 = -0.5,\ N = 2000.$

	Sequential Estimation				
$(\gamma_1^0 = -0.5, \gamma_2^0 = 0.5)$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
T=200	-0.582	0.483	0.179	1.028	-0.512
	(0.519)	(0.344)	(0.096)	(1.209)	(0.125)
T=400	-0.536	0.507	0.187	0.974	-0.509
	(0.418)	(0.207)	(0.066)	(0.683)	(0.086)
	Joint Estimation				
T=200	-0.511	0.415	0.180	1.250	-0.508
	(0.452)	(0.340)	(0.090)	(2.037)	(0.126)
T=400	-0.502	0.492	0.188	0.985	-0.501
	(0.409)	(0.205)	(0.071)	(0.675)	(0.082)

Table 3a: Correct Decision Frequencies: Linear Model

DGP: $y_t = \rho y_{t-1} + \epsilon_t$

	T=200				
ρ	BIC	AIC	HQ	BIC2	BIC3
0.5	88.4	44.9	70.9	99.3	99.8
0.7	87.8	45.1	69.7	98.8	99.8
0.9	85.6	41.2	66.3	98.7	99.9
1.0	50.0	9.7	24.9	89.8	98.7
	T=400				
0.5	92.4	44.9	72.5	99.2	100.0
0.7	91.1	44.7	71.9	99.5	100.0
0.9	90.9	42.0	70.7	99.4	100.0
1.0	56.9	9.7	27.4	93.2	99.7
	T=600				
0.5	93.5	45.7	74.1	99.7	100.0
0.7	92.1	45.8	74.4	99.7	100.0
0.9	91.9	42.6	73.5	99.6	100.0
1.0	60.1	9.8	29.2	94.2	99.7

Table 3b: Correct Decision Frequencies: Threshold Model

DGP: $y_t = \rho y_{t-1} I(y_{t-1} \leq 0) - \rho y_{t-1} I(y_{t-1} > 0) + \epsilon_t$

	T=200				
ρ	BIC	AIC	HQ	BIC2	BIC3
-0.40	100.0	100.0	100.0	99.3	95.1
-0.25	94.2	99.7	98.9	72.4	43.4
-0.15	63.3	91.3	80.4	25.2	7.4
-0.10	38.7	78.2	58.9	9.9	1.8
-0.05	19.0	61.5	38.2	2.7	0.2
	T=400				
-0.40	100.0	100.0	100.0	100.0	100.0
-0.25	99.9	100.0	100.0	97.0	84.2
-0.15	84.5	98.0	94.6	46.6	18.2
-0.10	53.2	88.6	75.2	16.5	3.5
-0.05	21.6	66.6	42.4	2.5	0.2
	T=600				
-0.40	100.0	100.0	100.0	100.0	100.0
-0.25	100.0	100.0	100.0	99.6	97.1
-0.15	93.5	99.6	98.8	60.5	36.3
-0.10	66.5	95.2	86.3	25.4	6.0
-0.05	22.7	72.1	46.9	2.8	0.2

Table 4a: Correct Decision Frequencies: Threshold Model ($m_0 = 1$)

Sequential Model Selection

$$y_t = \begin{cases} -3 + 0.5y_{t-1} - 0.9y_{t-2} + \epsilon_t & y_{t-2} \leq 1.5 \\ 2 + 0.3y_{t-1} + 0.2y_{t-2} + \epsilon_t & y_{t-2} > 1.5 \end{cases}$$

	T=400			
	$\hat{m} = 0$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$
BIC	0.8	80.5	18.7	0.0
BIC2	1.3	91.1	7.6	0.0
BIC3	1.3	91.5	7.1	0.0
	T=600			
	$\hat{m} = 0$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$
BIC	0.1	90.0	10.0	0.0
BIC2	0.2	96.3	3.5	0.0
BIC3	0.4	96.4	3.3	0.0
	T=800			
	$\hat{m} = 0$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$
BIC	0.1	94.4	5.4	0.0
BIC2	0.1	98.4	1.5	0.0
BIC3	0.2	98.5	1.3	0.0

Table 4b: Correct Decision Frequencies: Threshold Model ($m_0 = 2$)

Sequential Model Selection

$$y_t = \begin{cases} 2.7 + 0.8y_{t-1} - 0.2y_{t-2} + \epsilon_t & y_{t-2} \leq 5 \\ 6 + 1.9y_{t-1} - 1.2y_{t-2} + \epsilon_t & 5 < y_{t-2} \leq 12 \\ 1 + 0.7y_{t-1} - 0.3y_{t-2} + \epsilon_t & y_{t-2} > 12 \end{cases}$$

	T=400			
	$\hat{m} \leq 1$	$\hat{m} = 2$	$\hat{m} = 3$	$\hat{m} \geq 4$
BIC	0.0	79.7	20.3	0.0
BIC2	0.0	98.1	1.9	0.0
BIC3	0.0	99.1	0.9	0.0
	T=600			
	$\hat{m} \leq 1$	$\hat{m} = 2$	$\hat{m} = 3$	$\hat{m} \geq 4$
BIC	0.0	85.4	14.6	0.0
BIC2	0.0	99.0	1.0	0.0
BIC3	0.0	99.3	0.7	0.0
	T=800			
	$\hat{m} \leq 1$	$\hat{m} = 2$	$\hat{m} = 3$	$\hat{m} \geq 4$
BIC	0.0	88.1	11.9	0.0
BIC2	0.0	99.0	1.0	0.0
BIC3	0.0	99.8	0.2	0.0

REFERENCES

ALTISSIMO, F. AND G. L. VIOLANTE (1999) "The nonlinear dynamics of output and unemployment in the US," *University College London, Department of Economics, Unpublished Manuscript.*

BAI, J. (1997) "Estimating multiple breaks one at a time," *Econometric Theory,* 13, 315-352.

BAI, J. AND P. PERRON (1998a) "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica,* 66, pp. 47-78.

BAI, J. AND P. PERRON (1998b) "Computation and analysis of multiple structural change models," Unpublished Manuscript.

BEAUDRY, P. AND G. KOOP (1993) "Do recessions permanently change output?" *Journal of Monetary Economics,* 31, 149-164.

CANER, M. AND B. E. HANSEN (2000) "Threshold autoregression with a unit root", *Econometrica,* Forthcoming.

CARRASCO, M. (1997) "Misspecified Structural Change, Threshold and Markov Switching Models," *Ohio State University, Unpublished Manuscript.*

CHAN K. S. (1989) "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *Annals of Statistics,* 21, 520-533.

CHAN K. S. (1990) "Testing for Threshold Autoregression," *Annals of Statistics,* 18, 1886-1894.

CHAN, K. S. AND H. TONG (1990) "On likelihood ratio tests for threshold autoregression," *Journal of the Royal Statistical Society B,* 52, 469-476.

CHAN, K. S.(1991) "Percentage points of Likelihood Ratio for Threshold Autoregression," *Journal of the Royal Statistical Society B,* 53, 691-696.

DIRKSE J. P. (1975) "An absorption probability for the Ornstein-Uhlenbeck process," *Journal of Applied Probability,* 12, 595-599.

DURLAUF, S. N., AND P. A. JOHNSON (1995) "Multiple Regimes and Cross-Country Growth Behavior," *Journal of Applied Econometrics,* 10, 365-384.

GONZALEZ, M. and J. GONZALO (1998) "Threshold Unit Root Models," *Working Paper, Department of Statistics and Econometrics, Universidad Carlos III de Madrid.*

HAMILTON, J.D. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica,* 57, 357-384.

HANSEN, B. E. (1996) "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica,* 64, 413-430.

HANSEN, B. E. (1997) "Inference in TAR Models," *Studies in Nonlinear Dynamics and Econometrics,* 2, 1-14.

HANSEN, B. E. (1999a) "Testing for linearity," *Journal of Economic Surveys,* 13, 551-576.

HANSEN, B. E. (1999b) "Threshold effect in non-dynamic panels: Estimation, Testing, and Inference," *Journal of Econometrics,* Vol. 93, pp. 345-368.

HANSEN, B. E. (2000) "Sample Splitting and Threshold Estimation," *Econometrica,* 68, 575-603.

HAWKINS, D. M. (1976) "Point estimation of the parameters of piecewise regression models," *Applied Statistics,* 25, 51-57.

KOOP, G. AND S. M. POTTER (1999) "Dynamic asymmetries in U.S. Unemployment," *Journal of Business and Economic Statistics,* 17, 298-312.

LIU, J., S. WU AND J. W. ZIDEK (1997) "On Segmented Multivariate Regression," *Statistica Sinica,* 7, 497-525.

MOENADDIN, R. AND H. TONG (1988) "A comparison of likelihood ratio test and CUSUM test for threshold autoregression," *The Statistician,* 37, 213-225.

NEWEY, W. K., AND D. L. MCFADDEN (1994) "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics, Vol. IV,* ed. by R. F. Engle and D. L. McFadden. New-York: Elsevier, pp. 2113-2245.

OBSTFELD, M. AND A. TAYLOR (1997) "Nonlinear Aspects of Goods Market Arbitrage and Adjustment," *Journal of Japanese and International Economics,* 11, 441-79.

O'CONNELL, P. G. J. AND S. WEI (1997) "The bigger they are the harder they fall: How price differences across U.S. cities are arbitraged," *NBER Working Paper, No. W6089.*

PETRUCCELLI, J. D. (1992) "On the approximation of time series by threshold autoregressive models," *Sankhya, Series B, 54.*

POTTER, S. M. (1995) "A nonlinear approach to US GNP," *Journal of Applied Econometrics,* 2, 109-125.

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

TONG, H. (1983) *Threshold Models in Non-Linear Time Series Analysis: Lecture Notes in Statistics,* 21, Berlin, Springer-Verlag.

TONG, H. AND K. S. LIM (1980) "Threshold Autoregression, Limit Cycles and Cyclical Data," *Journal of The Royal Statistical Society, Series B,* 4, 245-292.

TSAY, R. S. (1989) "Testing and Modelling Threshold Autoregressive Processes," *Journal of the American Statistical Association,* Vol. 84, 231-240.

VOSTRIKOVA, L. J. (1981) "Detecting disorder in multidimensional random processes," *Soviet. Math. Dokl.*, 24, 55-59.