

Estimation and Inference in Threshold Type Regime Switching Models

Jesús Gonzalo
Universidad Carlos III de Madrid
Department of Economics
Calle Madrid 126
28903 Getafe (Madrid) - Spain

Jean-Yves Pitarakis
University of Southampton
Economics Division
Southampton SO17 1BJ
United-Kingdom

January 2, 2012

Abstract

¹Financial support from the ESRC is gratefully acknowledged. Address for Correspondence: Jean-Yves Pitarakis, University of Southampton, School of Social Sciences, Economics Division, Southampton, SO17 1BJ, United-Kingdom. Email: j.pitarakis@soton.ac.uk

1 Introduction

The recognition that linear time series models may be too restrictive to capture economically interesting asymmetries and empirically observed nonlinear dynamics has over the past twenty years generated a vast research agenda on designing models which could capture such features while remaining parsimonious and analytically tractable. Models that are capable of capturing nonlinear dynamics have also been the subject of a much earlier and extensive research led by Statisticians as well as practitioners in fields as broad as Biology, Physics and Engineering with a very wide range of proposed specifications designed to capture, model and forecast field specific phenomena (e.g. Bilinear models, Random Coefficient Models, State Dependent Models etc.). The amount of research that has been devoted to describing the nonlinear dynamics of Sunspot Numbers and Canadian Lynx data is an obvious manifestation of this quest (see Tong (1990), Granger and Terasvirta (1995), Hansen (1999), Terasvirta, Tjostheim and Granger (2010), and references therein).

A particular behaviour of interest to economists has been that of regime change or regime switching whereby the parameters of a model are made to change depending on the occurrence of a particular event, episode or policy (e.g. recessions or expansions, periods of low/high stock market valuations, low/high interest rates etc) but are otherwise constant within regimes. Popular models that can be categorised within this group are the well known Markov switching models popularised by Hamilton's early work (see Hamilton (1989)) and which model parameter change via the use of an unobservable discrete time Markov process. This class of models in which parameter changes are triggered by an unobservable binary variable has been used extensively as an intuitive way of capturing policy shifts in Macroeconomic models as well as numerous other contexts such as forecasting economic growth and dating business cycles. In Leeper and Zha (2003), Farmer, Waggoner and Zha (2009), Davig and Leeper (2007), Benhabib (2010) for instance the authors use such models to introduce the concept of monetary policy switches and regime specific Taylor rules. Other particularly fruitful areas of application of such regime switching specifications has involved

the dating of Business Cycles, the modelling of time variation in expected returns among numerous others (see Hamilton (2011), Perez-Quiros and Timmermann (2000) etc.).

An alternative, parsimonious and dynamically very rich way of modelling regime switching behaviour in economic data is to take an explicit stand on what might be triggering such switches and adopt a piecewise linear setting in which regime switches are triggered by an *observed* variable crossing an unknown threshold. Such models have been proposed by Howell Tong in the mid 70s and have gone through an important revival following their adoption by Economists and Econometricians during the 80s and 90s following the methodological work of Bruce Hansen (see also Hansen (2011) and references therein for a historical overview), Ruey Tsay (Tsay (1989), Tsay (1991)), Koop, Pesaran and Potter (1996), Koop and Potter (1999) and others. When each regime is described by an autoregressive process and the threshold variable causing the regime change is also a lagged value of the variable being modelled we have the well known Self Exciting Threshold AutoRegressive class of models (SETAR) extensively studied in the early work of Tong and others (see Tong and Lim (1980), Tong (1983, 1990), Chan (1990, 1993)). In general however the threshold principle may apply to a wider range of linear univariate or multivariate models and need not be solely confined to autoregressive functional forms. Similarly the threshold variable triggering regime switches may or may not be one of the variables included in the linear part of the model. Despite their simplicity, such models have been shown to be able to capture a very diverse set of dynamics and asymmetries particularly relevant to economic data. Important examples include the modelling of phenomena such as costly arbitrage whereby arbitrage occurs solely after the spread in prices exceeds a threshold due for instance to transport costs (see Lo and Zivot (2001), Obstfeld and Taylor (1997), O'Connell and Wei (1997), Balke and Fomby (1997)). Other areas of application include the study of asymmetries in the Business Cycles explored in Beaudry and Koop (1993), Potter (1995), Koop and Potter (1999), Altissimo and Violante (2001), the modelling of asymmetries in gasoline and crude oil prices (Borenstein, Cameron and Gilbert (1997)) and other markets (Balke (2000), Gospodinov (2005), Griffin, Nardari and Stultz (2007) etc).

Threshold models are particularly simple to estimate and conduct inferences on and despite the lack of guidance offered by economic theory for a particular nonlinear functional form such piecewise linear structures can be viewed as approximations to a wider range of functional forms as discussed in Petrucci (1992) and Tong (1990, pp. 98-100). Two key econometric problems that need to be addressed when contemplating the use of such models for one's own data involve tests for detecting the presence of threshold effects and if supported by the data the subsequent estimation of the underlying model parameters.

The purpose of this paper is to offer a pedagogical overview of the most commonly used inference and estimation techniques developed in the recent literature on threshold models. In so doing, we also aim to highlight the key strengths, weaknesses and limitations of each procedure and perhaps more importantly discuss potential areas requiring further research and interesting extensions. The plan of the paper is as follows. Section 2 concentrates on tests for detecting the presence of threshold nonlinearities against linear specifications. Section 3 explores methods of estimating the model parameters and their properties. Section 4 discusses important extensions and interesting areas for future work. Section 5 concludes.

2 Detecting Threshold Effects

In what follows we will be interested in methods for assessing whether the dynamics of a univariate time series y_t and a p -dimensional regressor vector x_t may be plausibly described by a threshold specification given by

$$y_t = \begin{cases} x_t' \beta_1 + u_t & q_t \leq \gamma \\ x_t' \beta_2 + u_t & q_t > \gamma \end{cases} \quad (1)$$

with q_t denoting the threshold variable triggering the regime switches and u_t the random disturbance term. At this stage it is important to note that our parameterisation in (1) is general enough to also be viewed as encompassing threshold autoregressions by requiring x_t to contain lagged values of y_t . Similarly, the threshold variable q_t may be one of the components of x_t or some external variable. The threshold parameter γ is assumed unknown

throughout but following common practice we require $\gamma \in \Gamma$ with $\Gamma = [\underline{\gamma}, \bar{\gamma}]$ denoting a compact subset of the threshold variable sample space. Given our specification in (1) the first concern of an empirical investigation is to test the null hypothesis of linearity $H_0 : \beta_1 = \beta_2$ against $H_1 : \beta_1 \neq \beta_2$.

Before proceeding with the various testing procedures it is useful to document alternative and occasionally more convenient formulations of the threshold model by introducing relevant indicator functions. Letting $I(q_t \leq \gamma)$ be such that $I(q_t \leq \gamma) = 1$ when $q_t \leq \gamma$ and $I(q_t \leq \gamma) = 0$ otherwise we define $x_{1t}(\gamma) = x_t * I(q_t \leq \gamma)$ and $x_{2t}(\gamma) = x_t * I(q_t > \gamma)$ so that (1) can also be written as

$$y_t = x_{1t}(\gamma)' \beta_1 + x_{2t}(\gamma)' \beta_2 + u_t \quad (2)$$

or in matrix notation as

$$y = X_1(\gamma) \beta_1 + X_2(\gamma) \beta_2 + u \quad (3)$$

with $X_i(\gamma)$ stacking the elements of $x_{it}(\gamma)$ for $i = 1, 2$ and which is such that $X_1(\gamma)' X_2(\gamma) = 0$. Our notation in (2)-(3) also makes it clear that for a known γ , say $\gamma = 0$, the above models are linear in their parameters and we are in fact in a basic textbook linear regression setting. This latter observation also highlights the importance of recognising the role played by the unknown threshold parameter when it comes to conducting inferences in threshold models. The price to pay for our desire to remain agnostic about the possible magnitude of γ and whether it exists at all is that we will need to develop tests that are suitable for any $\gamma \in \Gamma$. Naturally, we will also need to develop methods of obtaining a good estimator of γ once we are confident that the existence of such a quantity is supported by the data.

Within the general context of threshold models such as (1) the main difficulty for testing hypotheses such as $H_0 : \beta_1 = \beta_2$ arises from the fact that the threshold parameter γ is unidentified under this null hypothesis of linearity. This can be observed very clearly from our formulation in (3) since setting $\beta_1 = \beta_2$ leads to a linear model via $X_1(\gamma) + X_2(\gamma) \equiv X$ and in which γ plays no role. This problem is occasionally referred to as the Davies problem (see

Davies (1977, 1987) and Hansen (1996)) and is typically addressed by viewing the traditional Wald, LM or LR type test statistics as functionals of γ and subsequently focusing inferences on quantities such as the supremum or average of the test statistics across all possible values of γ .

Letting $X = X_1(\gamma) + X_2(\gamma)$ denote the p -dimensional regressor matrix in the linear model we can write its corresponding residual sum of squares as $S_T = y'y - y'X(X'X)^{-1}X'y$ while that corresponding to the threshold model is given by

$$S_T(\gamma) = y'y - \sum_{i=1}^2 y'X_i(\gamma)(X_i(\gamma)'X_i(\gamma))^{-1}X_i(\gamma)'y \quad (4)$$

for any $\gamma \in \Gamma$. This then allows us to write a Wald type test statistic for testing $H_0 : \beta_1 = \beta_2$ as

$$W_T(\gamma) = \frac{T(S_T - S_T(\gamma))}{S_T(\gamma)}. \quad (5)$$

Naturally we could also formulate alternative test statistics such as the likelihood ratio or LM in a similar manner e.g. $LR_T(\gamma) = T \ln S_T/S_T(\gamma)$ and $LM_T(\gamma) = T(S_T - S_T(\gamma))/S_T$. Due to the unidentified nuisance parameter problem inferences are typically based on quantities such as $\sup_{\gamma \in \Gamma} W_T(\gamma)$ or their variants (see Hansen (1996)).

For practical purposes the maximum Wald statistic is constructed as follows.

Step 1: Let \mathbf{qs} denote the $T \times 1$ dimensional sorted version of q_t . Since we operate under the assumption that $\gamma \in \Gamma$ a compact subset of $\{\mathbf{qs}[1], \dots, \mathbf{qs}[T]\}$ we trim a given fraction π from the top and bottom components of the $T \times 1$ vector \mathbf{qs} so as to obtain a new vector of threshold variable observations $\mathbf{qss} = \mathbf{qs}[T\pi : T(1 - \pi)]$. If $T = 1000$ for instance and $\pi = 10\%$ the new sorted and trimmed version of the threshold variable is given by $\mathbf{qss} = \mathbf{qs}[100 : 900]$. Let T_s denote the number of observations included in \mathbf{qss} .

Step 2: For each $i = 1, \dots, T_s$ construct the top and bottom regime regressor matrices given by $X_1[i] = x[1 : T] * I(q_t \leq \mathbf{qss}[i])$ and $X_2[i] = x[1 : T] * I(q_t > \mathbf{qss}[i])$. Note that

for each possible value of i , $X_1[i]$ and $X_2[i]$ are $T \times p$ regressor matrices with $*$ denoting the element by element multiplication operator and $x[1 : T]$ refers to the $T \times p$ original regressor matrix X .

Step 3: Using $X_1[i]$, $X_2[i]$ and X construct

$$S_T[i] = y'y - X_1[i]'(X_1[i]'X_1[i])^{-1}X_1[i] - X_2[i]'(X_2[i]'X_2[i])^{-1}X_2[i],$$

$S_T = y'y - y'X(X'X)^{-1}X'y$ and obtain a magnitude of the Wald statistics as defined above for each i , say $W_T[i]$ with $i = 1, \dots, T_s$.

Step 4: Use $\max_{1 \leq i \leq T_s} W_T[i]$ as the supremum Wald statistic and proceed similarly for $\max_{1 \leq i \leq T_s} LR_T[i]$ or $\max_{1 \leq i \leq T_s} LM_T[i]$ as required. Alternative test statistics may involve the use of averages such as $\sum_{i=1}^{T_s} W_T[i]/T_s$.

Upon completion of the loop, the decision regarding $H_0 : \beta_1 = \beta_2$ involves rejecting the null hypothesis for *large* values of the test statistics. Cutoffs and implied pvalues are obviously dictated by the limiting distribution of objects such as $\max_i W_T[i]$ which may or may not be tractable, an issue we concentrate on below.

The early research on tests of the null hypothesis of linearity focused on SETAR versions of (1) and among the first generation of tests we note the CUSUM type of tests developed in Petrucci and Davis (1986) and Tsay (1989). Chan (1990, 1991) subsequently extended this testing toolkit by obtaining the limiting distribution of a maximum LR type test statistic whose construction we described above. Chan (1990, 1991) established that under the null hypothesis $H_0 : \beta_1 = \beta_2$, suitable assumptions requiring stationarity, ergodicity and the iid'ness of the u_t 's, the limiting distribution of the supremum LR is such that $\sup_{\gamma} LR_T(\gamma) \Rightarrow \sup_{\gamma} \zeta(\gamma)' \Omega(\gamma) \zeta(\gamma) \equiv \sup_{\gamma} G_{\infty}(\gamma)$ with $\zeta(\gamma)$ denoting a zero mean Gaussian process and $\Omega(\gamma)$ its corresponding covariance kernel. Naturally the same result would hold for the Sup Wald or Sup LM statistics.

These results were obtained within a SETAR setting with the covariance kernel of $\zeta(\gamma)$ depending on model specific population moments in a complicated manner (e.g. unknown

population quantities such as $E[x_t^2 I(q_t \leq \gamma)]$ etc.). This latter aspect is important to emphasise since it highlights the unavailability of universal tabulations for $\sup_{\gamma} G_{\infty}(\gamma)$. Differently put the limiting distribution given by $G_{\infty}(\gamma)$ depends on model specific nuisance parameters and can therefore not be tabulated for practical inference purposes. There are however some very restrictive instances under which $G_{\infty}(\gamma)$ may simplify into a random variable with a familiar distribution that is free of any nuisance parameters. This can happen for instance if the threshold variable is taken as external, say independent of x_t and u_t . In this instance $G_{\infty}(\gamma)$ can be shown to be equivalent to a normalised squared Brownian Bridge Process identical to the limiting distribution of the Wald, LR or LM statistic for testing the null of linearity against a single structural break tabulated in Andrews (1993). More specifically, the limiting distribution is given by $[W(\lambda) - \lambda W(1)]^2 / \lambda(1 - \lambda)$ with $W(\lambda)$ denoting a standard Brownian Motion associated with u_t . Tong (1990, pp. 240-244) documents some additional special cases in which the limiting random variable takes the simple Brownian Bridge type formulation. See also Wong and Li (1997) for an application of the same test to a SETAR model with conditional heteroskedasticity. Note also that inferences would be considerably simplified if we were to proceed for a given value of γ , say $\gamma = 0$. This scenario could arise if one were interested in testing for the presence of threshold effects at a specific *location* such as q_t crossing the zero line. In this instance it can be shown that since $\zeta(\gamma = 0)$ is a multivariate normally distributed random variable with covariance $\Omega(\gamma = 0)$ the resulting Wald statistic evaluated at $\gamma = 0$, say $W_T(0)$, will have a χ^2 limit.

The lack of universal tabulations for test statistics such as $\max_i W_{\top}[i]$ perhaps explains the limited take up of threshold based specifications by Economists prior to the 90s. In an important paper, Hansen (1996) proposed a broadly applicable simulation based method for obtaining asymptotic pvalues associated with $\max_i W_{\top}[i]$ and related test statistics. Hansen's method is general enough to apply to both SETAR or any other threshold model setting, and bypasses the constraint of having to deal with unknown nuisance parameters in the limiting distribution. Hansen's simulation based method proposes to replace the

population moments of the limiting random variable with their sample counterparts and simulates the score under the null using NID(0,1) draws. This simulation based method is justified by the multiplier CLT (see Van der Wart and Wellner (1996)) and can in a way be viewed as an *external bootstrap*. It should not be confused however with the idea of obtaining critical values from a bootstrap distribution.

A useful exposition of Hansen’s simulation based approach which we repeat below can be found in Hansen (1999). For practical purposes Hansen’s (1996) method involves writing the sample counterpart of $G_\infty(\gamma)$, say $G_T(\gamma)$ obtained by replacing the population moments with their sample counterparts (the scores are simulated using NID(0,1) random variables). One can then obtain a large sample of draws, say $N=10000$, from $\max_{1 \leq i \leq T_s} G_T[i]$ so as to construct an approximation to the limiting distribution given by $\sup_\gamma G_\infty(\gamma)$. The computed test statistic $\max_{1 \leq i \leq T_s} W_T[i]$ can then be compared with either the quantiles of the simulated distribution (e.g. 9750th sorted value) or alternatively pvalues can be computed.

It is important to note that this approach is applicable to general threshold specifications and is not restricted to the SETAR family. Gauss, Matlab and R codes applicable to a general threshold specification as in (1) can be found as a companion code to Hansen (1997). The general format of the procedure involves the arguments y , x and q (i.e. the data) together with the desired level of trimming π and the number of replications N . The output then consists of $\max_{1 \leq i \leq T_s} W_T[i]$ together with its pvalue, say

$$\text{TEST}(y, x, q, \pi, N) \rightarrow \left(\max_{1 \leq i \leq T_s} W_T[i], \text{ pval} \right). \quad (6)$$

The above approach allows one to test the null hypothesis $H_0 : \beta_1 = \beta_2$ under quite general conditions and is commonly used in applied work.

An alternative and equally general model selection based approach that does not require any simulations has been proposed more recently by Gonzalo and Pitarakis (2002). Here, the problem of detecting the presence of threshold effects is viewed as a model selection problem among two competing models given by the linear specification $y_t = x_t' \beta + u_t$, say \mathcal{M}_0 , and \mathcal{M}_1 its threshold counterpart (2). The decision rule is based on an information

theoretic criterion of the type

$$IC_T(\gamma) = \ln S_T(\gamma) + 2 p \frac{c_T}{T}. \quad (7)$$

Here $2p$ refers to the number of estimated parameters in the threshold model (i.e. p slopes in each regime) and c_T is a deterministic penalty term. Naturally, under the linear model \mathcal{M}_0 we can write the criterion as

$$IC_T = \ln S_T + p \frac{c_T}{T}. \quad (8)$$

Intuitively, as we move from the linear to the less parsimonious threshold specification, the residual sum of squares declines and this decline is balanced by a greater penalty term (i.e. $2 p c_T$ versus $p c_T$). The optimal model is then selected as the model that leads to the smallest value of the IC criterion. More formally, we choose the linear specification if

$$IC_T < \min_{\gamma \in \Gamma} IC_T(\gamma) \quad (9)$$

and opt for the threshold model otherwise. It is interesting to note that this decision rule is very much similar to using a maximum LR type test statistic since $IC_T - \min_{\gamma} IC_T(\gamma) = \max_{\gamma} [IC_T - IC_T(\gamma)] = \max_{\gamma} [\ln(S_T/S_T(\gamma)) - p c_T/T]$. Equivalently, the model selection based approach points to the threshold model when $\max_{\gamma} LR_T(\gamma) > p c_T$. Thus, rather than basing inferences on the quantiles of the limiting distribution of $\max_{\gamma} LR_T(\gamma)$ we instead reach our decision by comparing the magnitude of $\max_{\gamma} LR_T(\gamma)$ with the deterministic quantity $p c_T$. This also makes it clear that the practical implementation of this model selection approach follows trivially once Steps 3 and 4 above have been completed. More specifically noting that the model selection based approach points to the threshold specification when

$$\max_{\gamma} \frac{T(S_T - S_T(\gamma))}{S_T(\gamma)} > T \left(e^{\frac{p c_T}{T}} - 1 \right) \quad (10)$$

it is easy to see that the decision rule can be based on comparing $\max_{1 \leq i \leq T_s} W_{\top}[i]$ with the deterministic term $T(e^{\frac{p c_T}{T}} - 1)$.

Gonzalo and Pitarakis (2002) further established that this model selection based approach leads to the *correct* choice of models (i.e. $\lim_{T \rightarrow \infty} P(\mathcal{M}_1 | \mathcal{M}_0) = \lim_{T \rightarrow \infty} P(\mathcal{M}_0 | \mathcal{M}_1) = 0$) provided that the chosen penalty term is such that $c_T \rightarrow \infty$ and $c_T/T \rightarrow 0$. Through extensive simulations Gonzalo and Pitarakis (2002) further argued that a choice of $c_T = \ln T$ leads to excellent finite sample results.

In Table 1 below we present a small simulation experiment in which we contrast the size properties of the test based approach with the ability of the model selection approach to point to the linear specification when the latter is true (i.e. correct decision frequencies). Our Data Generating Process is given by $y_t = 1 + 0.5x_{t-1} + u_t$ with x_t generated from an AR(1) process given by $x_t = 0.5x_{t-1} + v_t$. The random disturbances $w_t = (u_t, v_t)$ are modelled as an $NID(0, \Omega_2)$ random variable with $\Omega = \{(1.0.5), (0.5, 1)\}$. The empirical size estimates presented in Table 1 are obtained as the number of times across the N replications that the empirical p-value exceeds 1%, 2.5% and 5% respectively. The empirical p-values associated with the computed Wald type $\max W_T[i]$ test statistic are obtained using Bruce Hansen's publicly available *thrtest* routine. The correct decision frequencies associated with the model selection procedure correspond to the number of times across the N replications that $\max_{\gamma} T(S_T - S_T(\gamma))/S_T(\gamma) < T(e^{p \ln T/T} - 1)$.

Table 1. Size Properties of $\max_i W_T[i]$ and Model Selection Based Correct Decision Frequencies under a Linear DGP

	0.010	0.025	0.050	<i>MSEL</i>
$T = 100$	0.009	0.019	0.041	0.862
$T = 200$	0.013	0.029	0.055	0.902
$T = 400$	0.011	0.023	0.052	0.964

The above figures suggest that the test based on $\sup_{\gamma} W_T(\gamma)$ has good size properties even under small sample sizes. We also note that the ability of the model selection procedure to point to the true model converges to 1 as we increase the sample size. This is expected

from the underlying theory since the choice of a BIC type of penalty $c_T = \ln T$ satisfies the two conditions ensuring vanishing probabilities of over and under fitting.

In summary, we have reviewed two popular approaches for conducting inferences about the presence or absence of threshold effects within multiple regression models that may or may not include lagged variables. Important operating assumptions include stationarity and ergodicity, absence of serial correlation in the error sequence u_t , absence of endogeneity, and a series of finiteness of moments assumptions ensuring that laws of large numbers and CLTs can be applied. Typically, existing results are valid under a martingale difference assumption on u_t (see for instance Hansen (1999)) so that some forms of heterogeneity (e.g. conditional heteroskedasticity) would not be invalidating inferences. In fact all of the test statistics considered in Hansen (1996) are heteroskedasticity robust versions of Wald, LR and LM. It is important to note however that regime dependent heteroskedasticity is typically ruled out. A unified theory that may allow inferences in a setting with threshold effects in both the conditional mean and variance (with possibly different threshold parameters) is not readily available although numerous authors have explored the impact of allowing for GARCH type effects in threshold models (see Wong and Li (1997), Gospodinov (2005, 2008)). It will also be interesting to assess the possibility of handling serial correlation in models such as (1). Finally, some recent research has also explored the possibility of including persistent variables (e.g. near unit root processes) in threshold models. This literature was triggered by the work of Caner and Hansen (2001) who extended tests for threshold effects to models with unit root processes but much more remains to be done in this area (see Pitarakis (2008), Gonzalo and Pitarakis (2011, 2012)).

3 Estimation of Threshold Models and Further Tests

The natural objective of an empirical investigation following the rejection of the null hypothesis of linearity is the estimation of the unknown true threshold parameter, say γ_0 , together with the unknown slope coefficients β_{10} and β_{20} .

3.1 Threshold and Slope Parameter Estimation

The true model is now understood to be given by $y_t = x_{1t}(\gamma_0)' \beta_{10} + x_{2t}(\gamma_0)' \beta_{20} + u_t$ and our initial goal is the construction of a suitable estimator for γ_0 . A natural choice is given by the least squares principle which we write as

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} S_T(\gamma) \quad (11)$$

with $S_T(\gamma)$ denoting the concentrated sum of squared errors function. In words, the least squares estimator of γ is the value of γ that minimises $S_T(\gamma)$. It is also important to note that this argmin estimator is numerically equivalent to the value of γ that maximises the *homoskedastic* Wald statistic for testing $H_0 : \beta_1 = \beta_2$ i.e $\hat{\gamma} = \arg \max_{\gamma} W_T(\gamma)$ with $W_T(\gamma) = T(S_T - S_T(\gamma))/S_T(\gamma)$. From a practical viewpoint therefore $\hat{\gamma}$ is a natural byproduct of the test procedure described earlier (see Appendix A for a simple Gauss code for estimating $\hat{\gamma}$). We have

Step 1: Record the index $i = 1, \dots, T_s$ that maximises $W_T[i]$, say \hat{i}

Step 2: $\hat{\gamma}$ is obtained as `qss[\hat{i}]`.

The asymptotic properties of $\hat{\gamma}$ that have been explored in the literature have concentrated on its super consistency properties together with its limiting distribution. Early work on these properties was completed in Chan (1993) in the context of SETAR type threshold models (see also Koul and Qian (2002)). Chan (1993) established the important result of the T-consistency of $\hat{\gamma}$ in the sense that $T(\hat{\gamma} - \gamma_0) = O_p(1)$. This result was also obtained by Gonzalo and Pitarakis (2002) who concentrated on general threshold models with multiple regimes instead. Proving the consistency of the argmin estimator $\hat{\gamma}$ is typically done following a standard two step approach. In a first instance it is important to show that the objective function, say $S_T(\gamma)/T$ satisfies

$$\sup_{\gamma \in \Gamma} \left| \frac{S_T(\gamma)}{T} - S_{\infty}(\gamma) \right| \xrightarrow{p} 0 \quad (12)$$

with $S_\infty(\gamma)$ denoting a nonstochastic limit with a unique minimum. The consistency of $\hat{\gamma}$ then follows by showing that $S_\infty(\gamma)$ is uniquely minimised at $\gamma = \gamma_0$ i.e. $S_\infty(\gamma) > S_\infty(\gamma_0)$ for $\gamma < \gamma_0$ and $S_\infty(\gamma) > S_\infty(\gamma_0)$ for $\gamma > \gamma_0$.

In Chan (1993) the author also obtained the limiting distribution of $T(\hat{\gamma} - \gamma_0)$ with the latter shown to be a function of a compound Poisson process. This limit did not lend itself to any practical inferences however since dependent on a large number of nuisance parameters besides being particularly difficult to simulate due to the presence of continuous time jump processes.

As a way out of these difficulties and for the purpose of developing a toolkit that can be used by practitioners, Hansen (2000) adopted an alternative parameterisation of the threshold model that was then shown to lead to a convenient nuisance parameter free limiting distribution for $\hat{\gamma}$. The price to pay for this more favourable limiting theory was a rate of convergence for $\hat{\gamma}$ that was slightly lower than T . The main idea behind Hansen's approach was to reparameterise the threshold model in (1) in such a way that the threshold effect vanishes with T in the sense that $\delta_T = \beta_2 - \beta_1 \rightarrow 0$ as $T \rightarrow \infty$. Assuming Gaussian errors and using this vanishing threshold framework Hansen (2000) was able to obtain a convenient distribution theory for $\hat{\gamma}$ that is usable for conducting inferences and confidence interval construction. In particular, Hansen (2000) derived the limiting distribution of a Likelihood Ratio test for testing the null hypothesis $H_0 : \gamma = \gamma_0$ and showed it to be free of nuisance parameters provided that $\delta_T \rightarrow 0$ at a suitable rate. As mentioned earlier, the price to pay for this asymptotically vanishing threshold parameterisation is the slightly slower convergence rate of $\hat{\gamma}$. More specifically $T^{1-2\alpha}(\hat{\gamma} - \gamma_0) = O_p(1)$ for $0 < \alpha < \frac{1}{2}$ which can be contrasted with the T -consistency documented under non vanishing threshold effects. Note that here α is directly linked to the rate of decay of $\delta_T = \beta_2 - \beta_1 = c/T^\alpha$ so that the faster the threshold is allowed to vanish the slower the ensuing convergence of $\hat{\gamma}$.

Hansen (2000) subsequently showed that a Likelihood Ratio type test for testing the null hypothesis $H_0 : \gamma = \gamma_0$ takes a convenient and well known limiting expression that is

free of nuisance parameters provided that u_t is assumed to be homoskedastic in the sense that $E[u_t^2|q_t] = \sigma_u^2$. More specifically, Hansen (2000) established that

$$LR_T(\gamma_0) \xrightarrow{d} \zeta \quad (13)$$

with $P(\zeta \leq x) = (1 - e^{-x/2})^2$. The practical implementation of the test is now trivial and can be performed in two simple steps. Suppose for instance that one wishes to test $H_0 : \gamma = 0$. This can be achieved as follows

Step 1: Construct $LR_T = T(S_T(\gamma = 0) - S_T(\hat{\gamma}))/S_T(\hat{\gamma})$ with $\hat{\gamma} = \arg \min_{\gamma \in \Gamma} S_T(\gamma)$.

Step 2: The pvalue corresponding to the test statistic is $p = 1 - (1 - e^{-LR_T/2})^2$.

Following the work of Hansen (2000) numerous authors explored the possibility of developing inferences about γ (e.g. confidence intervals) without the need to operate within a vanishing threshold framework with gaussian errors and/or assuming error variances that cannot shift across regimes. In Gonzalo and Wolf (2005) the authors developed a flexible subsampling approach in the context of SETAR models while more recently Li and Ling (2011) revisited the early work of Chan (1993) and explored the possibility of using simulation methods to make the compound Poisson type of limit usable for inferences. The above discussions have highlighted the important complications that are caused by the presence of the discontinuity induced by the threshold variable. This prompted Seo and Linton (2007) to propose an alternative approach for estimating the parameters of a threshold model that relies on replacing the indicator functions that appear in (2) with a smoothed function à la smoothed maximum score of Horowitz (1992).

Finally, following the availability of an estimator for γ , the remaining slope parameter estimators can be constructed in a straightforward manner as

$$\hat{\beta}_i(\hat{\gamma}) = (X_i(\hat{\gamma})'X_i(\hat{\gamma}))^{-1}X_i(\hat{\gamma})'y \quad (14)$$

for $i = 1, 2$. An important result that follows from the consistency of $\hat{\gamma}$ and that makes inferences about the slopes simple to implement is the fact that $\hat{\beta}_i(\hat{\gamma})$ and $\hat{\beta}_i(\gamma_0)$ are asymptotically equivalent. More formally, we have $\sqrt{T}(\hat{\beta}_i(\hat{\gamma}) - \hat{\beta}_i(\gamma_0)) \xrightarrow{p} 0$ so that inferences

about the slopes can proceed as if γ were known. Under conditional homoskedasticity for instance t-ratios can be constructed in the usual manner via the use of covariances given by $\hat{\sigma}_u^2(\hat{\gamma})(X_i(\hat{\gamma})'X_i(\hat{\gamma}))^{-1}$ with $\hat{\sigma}_u^2(\hat{\gamma}) = S_T(\hat{\gamma})/T$.

3.2 Finite Sample Properties

At this stage it is also useful to gain some insights on the behaviour of estimators such as $\hat{\gamma}$ and $\hat{\beta}_i(\hat{\gamma})$ in finite samples commonly encountered in Economics. The bias and variability of $\hat{\gamma}$ is of particular importance since the asymptotics of $\hat{\beta}_i(\hat{\gamma})$ rely on the fact that we may proceed as if γ_0 were known. As noted in Hansen (2000) it is unlikely that we will ever encounter a scenario whereby $\hat{\gamma} = \gamma_0$ and taking this uncertainty into account in subsequent confidence intervals about the β_i 's becomes particularly important.

In order to evaluate the finite sample behaviour of the threshold and slope parameter estimators we consider a simple specification given by

$$y_t = \begin{cases} \beta_{10} + \beta_{11}x_{t-1} + u_t & q_{t-1} \leq \gamma_0 \\ \beta_{20} + \beta_{21}x_{t-1} + u_t & q_{t-1} > \gamma_0 \end{cases} \quad (15)$$

with $x_t = \phi_x x_{t-1} + v_t$ and $q_t = \phi_q q_{t-1} + e_t$. Letting $w_t = (u_t, v_t, e_t)$ we take $w_t \equiv NID(0, \Omega)$ and set $\Omega = \{(1, 0.5, -0.3), (0.3, 1.0.4), (-0.5, 0.4, 1)\}$ so as to allow for some dependence across the random shocks while satisfying the assumptions of the underlying distributional theory. Regarding the choice of parameters we use $\{\phi_q, \phi_x\} = \{0.5, 0.5\}$ throughout and set the threshold parameter $\gamma_0 = 0.25$.

Our initial goal is to assess the finite sample bias and variability of $\hat{\gamma} = \arg \min S_T(\gamma)$. For this purpose we distinguish between two scenarios of strong and weak threshold effects. Results for this experiment are presented in Table 2 below which display averages and standard deviations across N=1000 replications.

Table 2. Finite Sample Properties of $\hat{\gamma}$ and $\hat{\beta}_i(\hat{\gamma})$

	$E(\hat{\gamma})$	$\sigma(\hat{\gamma})$	$E(\hat{\beta}_{10})$	$\sigma(\hat{\beta}_{10})$	$E(\hat{\beta}_{20})$	$\sigma(\hat{\beta}_{20})$	$E(\hat{\beta}_{11})$	$\sigma(\hat{\beta}_{11})$	$E(\hat{\beta}_{21})$	$\sigma(\hat{\beta}_{21})$
<i>Case 1 (strong) : $\beta_{10} = 1, \beta_{20} = 2, \beta_{11} = 0.5, \beta_{12} = 1, \gamma_0 = 0.25$</i>										
$T = 100$	0.227	0.183	0.991	0.142	2.012	0.199	0.515	0.138	1.009	0.163
$T = 200$	0.243	0.080	0.996	0.099	2.004	0.128	0.507	0.087	1.014	0.104
$T = 400$	0.246	0.034	0.999	0.069	2.000	0.087	0.502	0.059	1.004	0.073
<i>Case 2 (weak) : $\beta_{10} = 1, \beta_{20} = 1, \beta_{11} = 0.5, \beta_{12} = 1, \gamma_0 = 0.25$</i>										
$T = 100$	0.156	0.621	1.016	0.239	0.962	0.276	0.494	0.201	1.052	0.212
$T = 200$	0.219	0.396	0.994	0.126	0.981	0.156	0.489	0.109	1.041	0.131
$T = 400$	0.248	0.215	1.000	0.074	0.987	0.098	0.495	0.064	1.021	0.082

The above figures suggest that both the threshold and slope parameter estimators have good small sample properties as judged by their bias and variability. We note that $\hat{\gamma}$ has negligible finite sample bias even under small sample sizes such as $T=200$. However an interesting distinguishing feature of $\hat{\gamma}$ is its substantial variability relative to that characterising the slope parameter estimators. Under the weak threshold scenario for instance and the moderately large sample size of $T=400$ we note that $\sigma(\hat{\gamma}) \approx E(\hat{\gamma})$ whereas the standard deviations of the $\hat{\beta}_i(\hat{\gamma})$'s are substantially smaller. It will be interesting in future work to explore alternative estimators that may have lower variability.

The above Data Generating Process can also be used to assess the properties of the LR based test for testing hypotheses about γ . Using the same parameterisation as in Table 2 we next consider the finite sample size properties of the Likelihood Ratio test for testing $H_0 : \gamma = 0.25$. Results for this experiment are presented in Table 3 below which contrasts nominal and empirical sizes. Empirical sizes have been estimated as the number of times (across N replications) that the estimated pvalue is smaller than 1%, 2.5% and 5% respectively. The scenario under consideration corresponds to Case 2 under a weak threshold parameterisation.

Table 3. Size Properties of the LR test for $H_0 : \gamma = 0.25$

	0.010	0.025	0.050
$T = 100$	0.010	0.025	0.065
$T = 200$	0.017	0.030	0.065
$T = 400$	0.015	0.032	0.054
$T = 800$	0.010	0.024	0.055

Table 3 above suggests an excellent match of theoretical and empirical sizes across a wide range of small to moderately large sample sizes. Note also that this happens under a rather weak threshold effect forcing solely the slope parameters to switch once q_{t-1} cross the value 0.25. It is also important to recall that the above inferences based on a nuisance parameter free limiting distribution are valid solely under a homoskedasticity restriction forcing $E[u_t^2|q_t]$ to be constant.

4 Going Beyond the Standard Assumptions & Suggestions for Further Work

The various methods for detecting the presence of threshold effects and subsequently estimating the model parameters that we reviewed above crucially depend on the stationarity and ergodicity of the series being modelled. It is indeed interesting to note that despite the enormous growth of the unit root literature the vast majority of the research agenda on exploring nonlinearities in economic data has operated under the assumption of stationarity highlighting the fact that nonstationarity and nonlinearities have been mainly treated in isolation. In fact one could also argue that they have often been viewed as mutually exclusive phenomena with an important strand of the literature arguing that neglected nonlinearities might be causing strong persistence.

One area through which threshold specifications entered the world of unit roots is through the concept of cointegration, a statistical counterpart to the notion of a long run equilibrium linking two or more variables. This naturally avoided the technical problems

one may face when interacting nonlinearities with nonstationarities since cointegrated relationships are by definition stationary processes and their residuals can be interpreted as mean-reverting equilibrium errors whose dynamics may describe the adjustment process to the long run equilibrium. Consider for instance two I(1) variables y_t and x_t and assume that they are cointegrated in the sense that the equilibrium error z_t is such that $|\rho| < 1$ in

$$\begin{aligned} y_t &= \beta x_t + z_t \\ z_t &= \rho z_{t-1} + u_t. \end{aligned} \tag{16}$$

Researchers such as Balke and Fomby (1997) proposed to use threshold type specifications for error correction terms for capturing the idea that adjustments to long run equilibria may be characterised by discontinuities or that there may be periods during which the speed of adjustment to equilibrium (summarised by ρ) may be slower or faster depending on how far we are from the equilibrium or alternatively depending on some external variable summarising the state of the economy. More formally the equilibrium error or error correction term can be formulated as

$$\Delta \hat{z}_t = \begin{cases} \rho_1 \hat{z}_{t-1} + v_t & q_{t-1} \leq \gamma \\ \rho_2 \hat{z}_{t-1} + v_t & q_{t-1} > \gamma \end{cases} \tag{17}$$

with $\hat{z}_t = y_t - \hat{\beta}x_t$ typically taken as the threshold variable q_t . Naturally one could also incorporate more complicated dynamics to the right hand side of (17) in a manner similar to an Augmented Dickey Fuller regression. The natural hypothesis to test in this context is again that of linear adjustment versus threshold adjustment via $H_0 : \rho_1 = \rho_2$. This simple example highlights a series of important issues that triggered a rich literature on testing for the presence of nonlinear dynamics in error correction models. First, the above framework assumes that y_t and x_t are *known* to be cointegrated so that z_t is stationary under both the null and alternative hypotheses being tested. In principle therefore the theory developed in Hansen (1996) should hold and standard tests discussed earlier should be usable (see also Enders and Siklos (2001)). Another difficulty with the specification of a SETAR type of model for \hat{z}_t is that its stationarity properties are still not very well understood beyond some

simple cases (see Chan and Tong (1985) and Caner and Hansen (2001, pp. 1567-1568))¹

One complication with alternative tests such as $H_0 : \rho_1 = \rho_2 = 0$ is that under this null the threshold variable (when $q_t \equiv \hat{z}_t$) is no longer stationary. It is our understanding that some of these issues are still in need of a rigorous methodological research agenda. Note for instance that fitting a threshold model to \hat{z}_t in (17) involves using a generated variable via $y_t - \hat{\beta}x_t$ unless one is willing to assume that the cointegrating vector is known.

Perhaps a more intuitive and rigorous framework for handling all of the above issues is to operate within a multivariate vector error correction setting à la Johansen. Early research in this area has been developed in Hansen and Seo (2002) who proposed a test of the null hypothesis of linear versus threshold adjustment in the context of a VECM. Assuming a VECM with a single cointegrating relationship and a known cointegrating vector Hansen and Seo (2001) showed that the limiting theory developed in Hansen (1996) continues to apply in this setting. However, and as recognised by the authors the validity of the distributional theory under an estimated cointegrating vector is unclear. These two points are directly relevant to our earlier claim about testing $H_0 : \rho_1 = \rho_2$ in (17). If we are willing to operate under a known β then the theory of Hansen (1996) applies and inferences can be implemented using a $\sup_{\gamma} W_T(\gamma)$ or similar test statistic.

In Seo (2006) the author concentrates on the null hypothesis of no linear cointegration which would correspond to testing the joint null hypothesis $H_0 : \rho_1 = \rho_2 = 0$ within our

¹Caner and Hansen (2001) was in fact one of the first papers that sought to combine the presence of unit root type of nonstationarities and threshold type nonlinear dynamics. Their main contribution was the development of a new asymptotic theory for detecting the presence of threshold effects in a series which was restricted to be a unit root process under the null of linearity (e.g. testing $H_0 : \beta_1 = \beta_2$ in $\Delta y_t = \beta_1 y_{t-1} I(q_{t-1} \leq \gamma) + \beta_2 y_{t-1} I(q_{t-1} > \gamma) + u_t$ with $q_t \equiv \Delta y_{t-k}$ for some $k \geq 1$ when under the null of linearity we have $\Delta y_t = u_t$ so that y_t is a pure unit root process). Pitarakis (2008) has shown that when the fitted threshold model contains solely deterministic regressors such as a constant and deterministic trends together with the unit root regressor y_{t-1} the limiting distribution of $\max_i W_T[i]$ takes a familiar form given by a normalised quadratic form in Brownian Bridges and readily tabulated in Hansen (1997). Caner and Hansen (2001) also explore further tests such as $H_0 : \beta_1 = \beta_2 = 0$ which are directly relevant for testing $H_0 : \rho_1 = \rho_2 = 0$ in the above ECM.

earlier ECM specification. Seo's work clearly highlights the impact that a nonstationary threshold variable has since under this null hypothesis the error correction term used as the threshold variable is also I(1) and Hansen's (1996) distributional framework is no longer valid. It is also worth emphasising that Seo's distributional results operate under the assumption of a known cointegrating vector. In a more recent paper Seo (2011) explores in greater depth the issue of an unknown cointegrating vector and derives a series of large sample results about $\hat{\beta}$ and $\hat{\gamma}$ via a smoothed indicator function approach along the same lines as Seo and Linton (2007).

Overall there is much that remains to be done. We can note for instance that all of the above research operated under the assumption that threshold effects were relevant solely in the adjustment process to the long run equilibrium with the latter systematically assumed to be given by a single linear cointegrating regression. An economically interesting feature that could greatly enhance the scope of the VECMs is the possibility of allowing the cointegrating vectors to also be characterised by threshold effects. This would be particularly interesting for the statistical modelling of *switching equilibria*. Preliminary work in this context can be found in Gonzalo and Pitarakis (2006a, 2006b).

5 Conclusions

The purpose of this chapter was to provide a comprehensive methodological overview of the econometrics of threshold models as used by Economists in applied work. We started our review with the most commonly used methods for detecting threshold effects and subsequently moved towards the techniques for estimating the unknown model parameters. Finally we also briefly surveyed how the originally developed stationary threshold specifications have evolved to also include unit root variables for the purpose of capturing economically interesting phenomena such as asymmetric adjustment to equilibrium. Despite the enormous methodological developments over the past ten to twenty years this line of research is still at its infancy. Important new developments should include the full development of an es-

timation and testing methodology for Threshold VARs similar to Johansen's linear VAR analysis together with a full representation theory that could allow for switches in both the cointegrating vectors and their associated adjustment process. As dicussed in Gonzalo and Pitarakis (2006a, 2006b) such developments are further complicated by the fact that it is difficult to associate a formal definition of threshold cointegration with the rank properties of VAR based long run impact matrices as it is the case in linearly cointegrated VARs.

APPENDIX

The code below estimates the threshold parameter $\hat{\gamma} = \arg \min_{\gamma} S_T(\gamma)$ using the specification in (15). It takes as inputs the variables $y \equiv y_t$, $x \equiv x_{t-1}$ and $q_t \equiv q_{t-1}$ and outputs $\hat{\gamma}$. The user also needs to input the desired percentage of data trimming used in the determination of Γ (e.g. trimper=0.10).

```
proc gamhatLS(y,x,q,trimper);
local t,qs,top,bot,qss,sigsq1,r,xmat1,xmat2,thetahat,zmat,res1,idx;
t=rows(y); /* sample size */
qs=sortc(q[1:t-1],1); /* threshold variable */
top=t*trimper;
bot=t*(1-trimper);
qss=qs[top+1:bot]; /* Sorted and Trimmed Threshold Variable */
sigsq1=zeros(rows(qss),1); /* Initialisation: Defining some vector of length (bot-top) */
r=1; /* Looping over all possible values of qss */
do while r<=rows(qss);
xmat1=ones(t-1,1).*(q[1:t-1].<=qss[r])~x[1:t-1].*(q[1:t-1].<=qss[r]);
xmat2=ones(t-1,1).*(q[1:t-1].>qss[r])~x[1:t-1].*(q[1:t-1].>qss[r]);
zmat=xmat1~xmat2;
thetahat=invpd(zmat'zmat)*zmat'y[2:t];
res1=y[2:t]-zmat*thetahat;
sigsq1[r]=res1'res1; /* Residual Sum of Squares for each possible value of qss */
r=r+1;
endo;
idx=minindc(sigsq1); /* Fetch the index where the smallest value of sigsq1 is located */
retp(qss[idx]); /* Returns the threshold parameter estimator */
endp;
```

REFERENCES

- Altissimo, F. and G. L. Violante (1999), 'The nonlinear dynamics of output and unemployment in the US', *Journal of Applied Econometrics*, 16, 461-486.
- Andrews, D. K. W. (1993), 'Tests for Parameter Instability and Structural Change with Unknown Change Point', *Econometrica*, 61, 821-856.
- Balke, N. and T. Fomby (1997), 'Threshold Cointegration', *International Economic Review*, 38, 627-645.
- Balke, N. (2000), 'Credit and Economics Activity: Credit Regimes and Nonlinear Propagation of Shocks', *Review of Economics and Statistics*, 82, 344-349.
- Benhabib, J. (2010), 'Regime Switching, Monetary Policy and Multiple Equilibria', Unpublished Manuscript, Department of Economics, New York University.
- Beaudry, P. and G. Koop (1993) 'Do recessions permanently change output?', *Journal of Monetary Economics*, 31, 149-164.
- Borenstein, S., Cameron, A. C. and R. Gilbert (1997), 'Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?', *Quarterly Journal of Economics*, 112, 305-39.
- Caner, M. and B. E. Hansen (2001), 'Threshold autoregression with a unit root', *Econometrica*, 69, 1555-1596.
- Chan, K. S. (1990), 'Testing for Threshold Autoregression', *Annals of Statistics*, 18, 1886-1894.
- Chan, K. S. (1993), 'Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model', *Annals of Statistics*, 21, 520-533.
- Chan, K. S. and Tong, H. (1985), 'On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations', *Advances in Applied Probability*, 17, 666-678.
- Davies, R. B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, 64, 247-254.

- Davies, R. B. (1987), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, 74, 33-43.
- Davig, T. and E. M. Leeper (2007), 'Generalizing the Taylor Principle', *American Economic Review*, 97, 607-635.
- Farmer, R. E. A, Waggoner, D. F. and T. Zha (2009), 'Indeterminacy in a forward-looking regime switching model', *International Journal of Economic Theory*, 5, 69-84.
- Enders, W. and P. L. Siklos (2001), 'Cointegration and threshold adjustment' *Journal of Business and Economic Statistics*, 19, 166-176.
- Gonzalo, J. and J. Pitarakis (2002), 'Estimation and Model Selection Based Inference in Single and Multiple Threshold Models', *Journal of Econometrics*, 110, 319-352.
- Gonzalo, J. and J. Pitarakis (2011), 'Regime Specific Predictability in Predictive Regressions', *Journal of Business and Economic Statistics*, In Press.
- Gonzalo, J. and J. Pitarakis (2012), 'Detecting Episodic Predictability Induced by a Persistent Variable', Unpublished Manuscript, Economics Division, University of Southampton.
- Gonzalo, J. and J. Pitarakis (2006a), 'Threshold Effects in Cointegrating Relationships', *Oxford Bulletin of Economics and Statistics*, 68, 813-833.
- Gonzalo, J. and J. Pitarakis (2006b), Threshold Effects in Multivariate Error Correction Models, in T. C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics: Econometric Theory*, Ch. 18 Volume 1, Palgrave MacMillan.
- Gonzalo, J., and M. Wolf (2005), 'Subsampling inference in threshold autoregressive models', *Journal of Econometrics*, 127, 201-224.
- Gospodinov, N. (2005), 'Testing for Threshold Nonlinearity in Short-Term Interest Rates', *Journal of Financial Econometrics*, 3, 344 -371.
- Gospodinov, N. (2008), 'Asymptotic and bootstrap tests for linearity in a TAR-GARCH(1,1) model with a unit root', *Journal of Econometrics*, 146, 146-161.

- Griffin, J. M., F. Nardari and R. M. Stultz (2007), 'Do Investors Trade More When Stocks Have Performed Well? Evidence from 46 Countries', *Review of Financial Studies*, 20, 905-951.
- Granger, C.W.J. and T. Terasvirta (1993) *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Hamilton, J. D. (1989), 'A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle', *Econometrica*, 57, 357-384.
- Hamilton, J. D. (2011), 'Calling Recessions in Real Time', *International Journal of Forecasting*, 27, 1006-1026.
- 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. (1999), 'Testing for linearity', *Journal of Economic Surveys*, 13, 551-576.
- Hansen, B. E. (2000), 'Sample Splitting and Threshold Estimation', *Econometrica*, 68, 575-603.
- Hansen, B. E. (2011), 'Threshold Autoregressions in Economics', *Statistics and Its Interface*, 4, 123-127.
- Horowitz, J. L. (1992), 'A Smoothed Maximum Score Estimator for the Binary Response Model', *Econometrica*, 60, 505-31.
- Koop, G., H. M. Pesaran and S. M. Potter (1996), 'Impulse response analysis in nonlinear multivariate models', *Journal of Econometrics*, 74, 119-147.
- Koop, G., and S. M. Potter (1999), 'Dynamic asymmetries in U.S. Unemployment', *Journal of Business and Economic Statistics*, 17, 298-312.

- Koul, L.H. and Qian, L.F. (2002), 'Asymptotics of maximum likelihood estimator in a two-phase linear regression model', *Journal of Statistical Planning and Inference*, 108, 99-119.
- Leeper, E. M. and T. Zha (2003), 'Modest Policy Interventions', *Journal of Monetary Economics*, 50, 1673-1700.
- Li, D. and S. Ling (2011), 'On the least squares estimation of multiple-regime threshold autoregressive models', *Journal of Econometrics*, Forthcoming.
- Lo, M. C. and E. Zivot (2001), 'Threshold cointegration and nonlinear adjustment to the law of one price', *Macroeconomic Dynamics*, 5, 533-576.
- 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 arbitrated,' *NBER Working Paper, No. W6089*.
- Perez-Quiros, G. and A. Timmermann (2000), 'Firm Size and Cyclical Variations in Stock Returns', *Journal of Finance*, 55, 1229-1262.
- Petrucelli, J. D. (1992), 'On the approximation of time series by threshold autoregressive models', *Sankhya, Series B*, 54, 54-61.
- Petrucelli, J.D. and Davies N. (1986), 'A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series', *Biometrika*, 73, 687-694.
- Pitarakis, J. (2008), 'Threshold autoregression with a unit root revised', *Econometrica*, 76, 1207-1217.
- Potter, S. M. (1995), 'A nonlinear approach to US GNP', *Journal of Applied Econometrics*, 2, 109-125.
- Seo, M. H. (2006), 'Bootstrap testing for the null of no cointegration in a threshold vector error correction model', *Journal of Econometrics*, 134, 129-150.

- Seo, M. H. and O. Linton (2007), ‘A Smoothed Least Squares Estimator For Threshold Regression Models’, *Journal of Econometrics*, 141, 704-735.
- Terasvirta, T., Tjostheim, D. and C. W. J. Granger (2010), *Modelling Nonlinear Economic Time Series*, Oxford University Press, New-York, USA.
- 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.
- Tong, H. (1983), *Threshold Models in Non-Linear Time Series Analysis: Lecture Notes in Statistics*, 21, Berlin, Springer-Verlag.
- Tong, H. (1990) *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press: Oxford.
- Tsay, R. S. (1989), ‘Testing and Modeling Threshold Autoregressive Processes’, *Journal of the American Statistical Association*, 84, 231-240.
- Tsay, R. S. (1991), ‘Detecting and modeling nonlinearity in univariate time series analysis’, *Statistica Sinica*, 1, 431-451.
- Tsay, R. S. (1998), ‘Testing and Modeling Multivariate Threshold Models’, *Journal of the American Statistical Association*, 93, 1188-1202.
- Van der Vaart, A.W., and J. A. Wellner, (2009), *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Wong, C. S. and Li, W. K. (1997), ‘Testing of threshold autoregression with conditional heteroscedasticity’, *Biometrika*, 84, 407-418.