15

Threshold Effects in Multivariate Error Correction Models*

Jesùs Gonzalo and Jean-Yves Pitarakis

Abstract

We propose a testing procedure for assessing the presence of threshold effects in nonstationary vector autoregressive models with or without cointegration. Our approach involves first testing whether the long-run impact matrix characterizing the VECM type representation of the VAR switches according to the magnitude of some threshold variable and is valid regardless of whether the system is purely I(1), I(1) with cointegration or stationary. Once the potential presence of threshold effects is established we subsequently evaluate the cointegrating properties of the system in each regime through a model selection based approach whose asymptotic and finite sample properties are also established. This subsequently allows us to introduce a novel non-linear permanent and transitory decomposition of the vector process of interest.

*We wish to thank the Spanish Ministry of Education for supporting this research under grants No. SEC01-0890 and SEJ2004-0401ECON.

15.1 Introduction

A growing body of research in the recent time series literature has concentrated on incorporating nonlinear behavior into conventional linear reduced form specifications such as autoregressive and moving average models. The motivation for moving away from the traditional linear model with constant parameters has typically come from the observation that many economic and financial time series are often characterized by regime-specific behavior and asymmetric responses to shocks. For such series the linearity and parameter constancy restrictions are typically inappropriate and may lead to misleading inferences about their dynamics.

Within this context, and a univariate setting, a general class of models that has been particularly popular from both a theoretical and applied perspective is the family of threshold models, which are characterized by piecewise linear processes separated according to the magnitude of a threshold variable which triggers the changes in regime. When each linear regime follows an autoregressive process, we have the well-known threshold autoregressive class of models, the statistical properties of which have been investigated in the early work of Tong and Lim (1980), Tong (1983, 1990), Tsay (1989) and Chan (1990, 1993), and more recently reconsidered and extended in Hansen (1996, 1997, 1999a, 1999b, 2000), Caner and Hansen (2001), Gonzalez and Gonzalo (1997), Gonzalo and Montesinos (2000) and Gonzalo and Pitarakis (2002), among others. The two key aspects on which this theoretical research has focused is the development of a distributional theory for tests designed to detect the presence of threshold effects and the statistical properties of the resulting parameter estimators characterising such models.

Given their ability to capture a very rich set of dynamic behavior, including persistence and asymmetries, the use of this class of models has been advocated in numerous applications aiming to capture economically meaningful nonlinearities. Examples include the analysis of asymmetries in persistence in US output growth (Beaudry and Koop, 1993; Potter, 1995), asymmetries in the response of output prices to input price increases versus decreases (Borenstein, Cameron and Gilbert, 1997; Peltzman, 2000), nonlinearities in unemployment rates (Hansen, 1997; Koop and Potter, 1999), and threshold effects in cross-country growth regressions (Durlauf and Johnson, 1995) and in international relative prices (Michael, Nobay and Peel, 1997; Obstfeld and Taylor, 1997; O'Connell and Wei, 1997; Lo and Zivot, 2001), among numerous others.

Although the vast majority of the theoretical developments in the area of testing and estimation of univariate threshold models have been obtained under the assumption of stationarity and ergodicity, another important motivation for their popularity came from the observation that a better description of the dynamics of numerous economic variables can be achieved by interacting the pervasive nature of unit roots with that of threshold effects within the same specification. This was also motivated by the observation that there might be much weaker support for the unit root hypothesis when the alternative hypothesis under consideration allows for the presence of threshold type effects in the time series of interest. In Pippenger and Goering (1993), for example, the authors documented a substantial fall in the power of the Dickey–Fuller test when the stationary alternative was allowed to include threshold effects. This also motivated the work of Enders and Granger (1998), who proposed a simple test of the null hypothesis of a unit root against asymmetric adjustment instead of a linear stationary alternative.

One important property of threshold models that contributed to this line of research is their ability to capture persistent behavior while remaining globally stationary. This can be achieved, for example, by allowing a time series to follow a unit-root-type process such as a random walk within one regime while being stationary in another. Numerous economic and financial variables, such as unemployment rates or interest rates, must be stationary by the mere fact that they are bounded. However, at the same time, conventional unit roots tests are typically unable to reject the null hypothesis of a unit root in their autoregressive representation. This observation has prompted numerous researchers to explore the possibility that the dynamics of these series may be better described by threshold models that allow the nonstationary component to occur within a corridor regime. A well-known example highlighting this point is the behavior of real exchange rate series, which are typically found to be unit root processes, implying a lack of international arbitrage and violation of the PPP hypothesis. Once allowance is made for the presence of threshold effects, capturing aspects such as transaction costs, it has typically been found that this nonstationarity only occurs locally (e.g., between transaction cost bounds) and that the process is, in fact, globally stationary (see Bec, Ben-Salem and Carrasco, 2001, and references therein). Within a related context, Gonzalez and Gonzalo (1998) introduced a globally stationary process, referred to as a threshold unit root model, that combines the presence of a unit root with threshold effects, and found strong support in favor of such a specification when modeling interest rate series.

Although all of this research lay within a univariate setup, the recent time series literature has also witnessed a growing interest in the inclusion of threshold effects in multivariate settings such as vector error correction models. A key factor that triggered this research was the observation that threshold effects may also have an intuitive appeal when it comes to modeling the adjustment process toward a long-run equilibrium characterizing two or more variables.

From the early work of Engle and Granger (1987), for instance, it is well known that two or more variables that behave like unit root processes individually may, in fact, be linked via a long-run equilibrium relationship that makes particular linear combinations of these variables stationary or, as it is commonly known, cointegrated. When this happens, the variables in question admit an error correction model representation that allows for the joint modeling of both their long-run and short-run dynamics. In its linear form, such an error correction specification restricts the adjustment process to remain the same across time, thereby ruling out the possibility of lumpy and discontinuous adjustment. An important paper, which relaxed this linearity assumption by introducing the possibility of threshold effects in the adjustment process toward the long-run equilibrium, thereby capturing phenomena such as changing speeds of adjustment, was Balke and Fomby (1997), where the authors introduced the concept of threshold cointegration (see also Tsay, 1998).

The inclusion of such nonlinearities in error correction models has been found to have a very strong intuitive and economic appeal, allowing, for example, for the possibility that the adjustment process toward the long-run equilibrium behaves differently depending on how far off the system is from the long-run equilibrium itself (i.e., depending on the magnitude of the equilibrium error). This also allows for the possibility that the adjustment process shuts down over certain periods. Consider, for instance, the prices of the same asset in two different geographical regions. Although both prices will be equal in the long-run equilibrium, due to the presence of transaction costs, arbitrage only kicks in when the difference in price (i.e., the equilibrium error) is sufficiently large.

The concept of threshold cointegration, as introduced in Balke and Fomby (1997), has attracted considerable attention from practitioners interested in uncovering nonlinear adjustment patterns in relative prices and other variables (see Wohar and Balke, 1998; Baum, Barkoulas and Caglayan, 2001; Enders and Falk, 1998; Lo and Zivot, 2001; O'Connell and Wei, 1997). From a methodological point of view, Balke and Fomby (1997) proposed to assess such occurences within a simple setup which consisted of adapting the approach developed in Hansen (1996) to an Engle–Granger type test performed on the cointegrating residuals. Their setup also implicitly assumed the existence of a known and single cointegrating vector linking the variables of interest. In a related study, Enders and Siklos (2001) extended Balke and Fomby's methodology by adapting the work of Enders and Granger (1998) to a cointegrating framework.

Despite the substantial interest generated by the introduction of the concept of threshold cointegration in Balke and Fomby (1997), a full statistical treatment within a formal multivariate error correction type of specification has only been available since the recent work of Hansen and Seo (2002) (see also Tsay (1998), who introduced an arranged regression approach for testing for the presence of threshold effects in VARs). Although also dealing with a multivariable cointegration setup, the methodology proposed in Balke and Fomby (1998) or Enders and Siklos (2001) focused on the direct treatment of the cointegrating residuals akin to the familiar Engle–Granger test for cointegration. In Hansen and Seo (2002), however, the authors developed a maximum likelihood based estimation and testing theory, starting directly from a vector error correction model representation of a cointegrated system with potential threshold effects in its adjustment process. More specifically, Hansen and Seo (2002) considered a VECM, assumed to contain a single cointegrating vector, in which the threshold effects are driven by the error correction term. Their analysis also implicitly assumes that the researcher knows in advance the cointegration properties of the system (i.e., the system is known to be cointegrated with a single cointegrating vector) and interest solely lies in detecting the presence of threshold effects in the adjustment process toward the equilibrium. This simplifying assumption avoids the need to test for cointegration in the presence of a potentially nonlinear adjustment process. In more recent research, Seo (2004) concentrated on this latter issue by developing a new distributional theory for directly testing the null of no cointegration against the alternative of threshold cointegration. In Seo's (2004) framework it is again the case that cointegration, if present, is solely characterized by a single cointegrating vector and, as in Hansen and Seo (2002), the threshold variable of interest is taken to be the error correction term itself.

In the present chapter our goal is to contribute further to the analysis of threshold effects in possibly cointegrated multivariate systems of the vector error correction type. Our initial goal is to evaluate the properties of a Wald-type test for testing the null of linearity against threshold nonlinearity in the long-run impact matrix of a VECM. Our analysis does not presume any specific cointegration properties of the system and is valid regardless of whether the system is cointegrated or not. One additional difference from previous work is our view about the threshold variable that induces the presence of threshold effects. Instead of taking the error correction term to be the variable whose magnitude triggers threshold effects, we consider a general external threshold variable, which could be any economic or financial variable that is stationary and ergodic, such as the growth rate in the economy. Having established the existence of threshold effects in the VECM representation of our system, we subsequently evaluate the properties of least squares based estimators of the threshold parameter, focusing on both its large and small sample properties, followed by an analysis of the formal cointegration properties of the system when applicable. This then allows us to formally obtain a nonlinear permanent and transitory decomposition of the vector process of interest following the same methodology as in Gonzalo and Granger (1995).

The rest of this chapter is as follows. Section 15.2 develops the theory for testing for the presence of threshold effects in a Vector Error Correction type of model. Section 15.3 focuses on the theoretical properties of estimators of the threshold parameters. Section 15.4 proposes a methodology for assessing the cointegration properties of the system, Section 15.5 introduces a nonlinear permanent and transitory decomposition based on a VECM with threshold effects and Section 15.6 concludes. All proofs are relegated to the appendix.

15.2 Testing for threshold effects in a multivariate framework

15.2.1 The model and test statistic

We let the p-dimensional time series $\{Y_t\}$ be generated by the following vector error correction type specification, which allows for the presence of threshold effects in its long run impact matrix:

$$\Delta Y_t = \mu + \mathbf{\Pi}_1 Y_{t-1} I(q_{t-d} \leq \gamma) + \Pi_2 Y_{t-1} I(q_{t-d} > \gamma) + \sum_{j=1}^{k} \Gamma_j \Delta Y_{t-j} + u_t \qquad (15.1)$$

where Π_1, Π_2 and Γ_j are $p \times p$ constant parameter matrices, q_{t-d} is a scalar threshold variable, $I(.)$ is the indicator function, γ the threshold parameter, k and d the known lag length and delay parameters and u_t is the p-dimensional random disturbance vector.

The model in (15.1) is a multivariate generalization of an autoregressive model with threshold effects whose dynamics are characterized by a piecewise linear vector autoregression. The regime switches are governed by the magnitude of the threshold variable q_t crossing an unknown threshold value γ. The specification in (15.1) is similar to that considered in Seo (2004), except that no assumptions are made about the rank structure of either $\mathbf{\Pi}_1$ or $\mathbf{\Pi}_2$, and the threshold variable is not necessarily given by an error correction term such as $q_t = \beta' Y_t$, with β denoting the single cointegrating vector.

The initial question of interest in the context of the specification in (15.1) is whether the long-run impact matrix is truly characterized by threshold effects driven by the threshold variable q_t. In the absence of such effects we have a standard linear VECM with $\mathbf{\Pi}_1 = \mathbf{\Pi}_2$, and this restriction can be tested via a conventional Wald-type test statistic against the alternative $H_1 : \mathbf{\Pi}_1 \neq \mathbf{\Pi}_2$.

At this stage it is important to note that the sole purpose of testing the above null hypothesis is to uncover the presence or absence of threshold effects in the long run impact matrix. More importantly, we wish to conduct this set of inferences regardless of the stationarity properties of Y_t, in the sense that our null hypothesis may hold under a purely stationary set up or a unit root set up with or without cointegration. If the null hypothesis is not rejected we can then carry on with the process of exploring the stochastic properties of the data following, for example, Johansen's methodology (see Johansen, 1998 and references therein). Before proceeding further, and to motivate our working model, we consider two simple examples illustrating particular cases of our specification in (15.1).

EXAMPLE 1: Here we present a bivariate system of cointegrated I(1) variables with threshold effects in their adjustment process. Specifically, with $Y_t = (y_{1t}, y_{2t})'$ we write $y_{1t} = \beta y_{2t} + z_t$, where $\Delta y_{2t} = \epsilon_{2t}$ and $\Delta z_t = \rho_1 z_{t-1} I(q_{t-1} \leq \gamma) + \rho_2 z_{t-1} I(q_{t-1} > \gamma) + \epsilon_{1t}$ with $\rho_i < 0$ for $i = 1, 2$, and for simplicity we take q_t to be an iid random variable. In this example both y_{1t} and y_{2t} are I(1) and cointegrated with cointegrating vector $(1, -\beta)$, since z_t is a covariance stationary process following a threshold autoregressive scheme. It is now straightforward to reformulate the above model as in (15.1) by writing

$$\begin{aligned} \begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} &= \begin{pmatrix} \rho_1 \\ 0 \end{pmatrix} (1 - \beta) \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} \leq \gamma) \\ &\quad + \begin{pmatrix} \rho_2 \\ 0 \end{pmatrix} (1 - \beta) \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} \geq \gamma) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \end{aligned} \tag{15.2}$$

with $u_{1t} = \epsilon_{1t} + \beta \epsilon_{2t}$ and $u_{2t} = \epsilon_{2t}$.

EXAMPLE 2: Here we consider a purely stationary bivariate system with both variables following a threshold autoregressive process. Consider $\Delta y_{1t} = \rho_{11} y_{1t-1} I(q_{t-1} \leq \gamma) + \rho_{21} y_{1t-1} I(q_{t-1} > \gamma) + u_{1t}$ and $\Delta y_{2t} = \rho_{12} y_{2t-1} I(q_{t-1} \leq \gamma) + \rho_{22} y_{2t-1} I(q_{t-1} > \gamma) + u_{2t}$ with $\rho_{i1} < 0$ and $\rho_{i2} < 0$ for $i = 1, 2$. We can again reformulate this system as in (15.1) by writing

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} \rho_{11} & 0 \\ 0 & \rho_{12} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} \leq \gamma) + \begin{pmatrix} \rho_{21} & 0 \\ 0 & \rho_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} > \gamma) + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \tag{15.3}$$

In order to explore the properties of the Wald-type test for the above null hypothesis, it will be convenient to reformulate (15.1) in matrix form. In what follows, for clarity and simplicity of exposition we focus on a restricted version of (1) which sets the constant term as well as the coefficients on the lagged dependent variables equal to zero. Since our framework does not consider threshold effects in those parameters, it would be straightforward to concentrate (15.1) with respect to $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ using an appropriate projection matrix. This leads to no loss of generality since our distributional results presented in Propositions 1 and 2 below would remain unaffected. We now write

$$\Delta Y = \mathbf{\Pi}_1 Z_1 + \mathbf{\Pi}_2 Z_2 + U \tag{15.4}$$

where $\Delta Y, Z_1$ and Z_2 are all $p \times T$ matrices stacking the vectors $\Delta Y_t, Y_{t-1} I(q_{t-d} \leq \gamma)$ and $Y_{t-1} I(q_{t-d} > \gamma)$, respectively. Within the formulation (15.4) we have $\Delta Y = (\Delta y_1, \Delta y_2, \ldots, \Delta y_T)$ $Z_1 = (y_0 I(q_{0-d} \leq \gamma), \ldots, y_{T-1} I(q_{T-d} \leq \gamma))$ and $Z_2 = (y_0 I(q_{0-d} > \gamma), \ldots, y_{T-1} I(q_{T-d} > \gamma))$. Similarly U is a $p \times T$ matrix of random disturbances given by $U = (u_1, \ldots, u_T)$. We note that within our parameterization the regressor matrices Z_1 and Z_2 are orthogonal due to the presence of the two indicator functions. Their dependence on γ is omitted for notational parsimony. For later use we also introduce the $p \times T$ matrix $Z = (y_0, \ldots, y_{T-1})$, which is such that $Z = Z_1 + Z_2$.

The unknown parameters of the model (15.4) can be estimated via concentrated least squares, proceeding conditionally on a known γ. Indeed, since given γ the model is linear in its parameters, the least squares estimators of $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are $\widehat{\mathbf{\Pi}}_1(\gamma) = \Delta Y Z_1' (Z_1 Z_1')^{-1}$ and $\widehat{\mathbf{\Pi}}_2(\gamma) = \Delta Y Z_2' (Z_2 Z_2')^{-1}$. For later use we also introduce the vectorised versions of the parameter matrices, writing $\hat{\pi}_1 \equiv vec\, \widehat{\mathbf{\Pi}}_1$ and $\hat{\pi}_2 \equiv vec\, \widehat{\mathbf{\Pi}}_2$, and the null hypothesis of interest can be equivalently expressed as $H_0 : \pi_1 = \pi_2$ or $H_0 : R\pi = 0$ with $R = [I_{p^2}, -I_{p^2}]$ and $\pi = (\pi_1', \pi_2')'$.

The Wald statistic for testing the above null hypothesis takes the following form

$$W_T(\gamma) = (R\hat{\pi})' \Big[R((DD')^{-1} \otimes \hat{\mathbf{\Omega}}_u) R' \Big]^{-1} (R\hat{\pi}) \tag{15.5}$$

where $\otimes$ is the Kronecker product operator, $\hat{\pi}_1 = [(Z_1 Z_1')^{-1} Z_1 \otimes I_p]\ vec\ \Delta Y, \hat{\pi}_2 = [(Z_2 Z_2')^{-1} Z_2 \otimes I_p]\ vec\ \Delta Y$ and $D = [Z_1\ Z_2]$. The $p \times p$ matrix $\hat{\boldsymbol{\Omega}}_u$ refers to the least squares estimator of the covariance matrix defined as $\hat{\boldsymbol{\Omega}}_u = \hat{U}\hat{U}'/T$, with $\hat{U} = \Delta Y - \hat{\boldsymbol{\Pi}}_1(\gamma) Z_1 - \hat{\boldsymbol{\Pi}}_2(\gamma) Z_2$. Since Z_1 and Z_2 are orthogonal, it also immediately follows that $DD' = diag(Z_1 Z_1', Z_2 Z_2')$ and $(DD')^{-1} \otimes \hat{\boldsymbol{\Omega}}_u = diag[(Z_1 Z_1')^{-1} \otimes \hat{\boldsymbol{\Omega}}_u, (Z_2 Z_2')^{-1} \otimes \hat{\boldsymbol{\Omega}}_u]$. We can thus also reformulate the Wald statistic in (15.5) as

$$W_T(\gamma) = (\hat{\pi}_1 - \hat{\pi}_2)' \left[(Z_2 Z_2')(ZZ')^{-1}(Z_1 Z_1') \otimes \hat{\boldsymbol{\Omega}}_u^{-1} \right] (\hat{\pi}_1 - \hat{\pi}_2) \tag{15.6}$$

where $ZZ' = Z_1 Z_1' + Z_2 Z_2'$.

At this stage it is also important to reiterate the fact that, when implementing our test of the null hypothesis of linearity with, say, $\boldsymbol{\Pi}_1 = \boldsymbol{\Pi}_2 = \boldsymbol{\Pi}$, the corresponding characteristic polynomial $\Phi(z) = (1 - z) I_p - \boldsymbol{\Pi} z$ will be assumed to have all its roots either outside or on the unit circle and the number of unit roots present in the system will be given by $p - r$ with $0 \leq r \leq p$. Our analysis rules out instances of explosive behavior or processes that may be integrated of order two. This also allows us to have a direct correspondence between the stochastic properties of Y_t under the null hypothesis and the rank structure of the long-run impact matrix $\boldsymbol{\Pi}$. In the particular case where all the roots of the characteristic polynomial are outside the unit circle, the series will be referred to as $I(0)$.

15.2.2 Assumptions and limiting distributions

Throughout this section we will be operating under the following set of assumptions

- (A1) $u_t = (u_{1t}, \ldots, u_{pt})'$ is a zero mean iid sequence of p-dimensional random vectors with a bounded density function, covariance matrix $E[u_t u_t'] = \Omega_u > 0$ and with $E|u_{it}|^{2\delta} < \infty$ for some $\delta > 2$ and $i = 1, \ldots, p$;
- (A2) q_t is a strictly stationary and ergodic sequence that is independent of $u_{is} \forall t, s, i = 1, \ldots, p$ and has distribution function F that is continuous everywhere;
- (A3) the threshold parameter γ is such that $\gamma \in \Gamma = [\gamma_L, \gamma_U]$, a closed and bounded subset of the sample space of the threshold variable.

Assumption (A1) above is required for our subsequent limiting distribution theory. It will ensure, for instance, that the functional central limit theorem can be applied to the sample moments used in the construction of Wald and related tests. Assumption (A2) restricts the behavior of the scalar random variable that induces threshold effects in the model (15.1). Although it allows q_t to follow a very rich class of processes, it requires it to be external, in the sense of being independent of the u_t sequence, and also rules out the possibility of q_t being I(1) itself. Finally, assumption (A3) is standard in this literature. The threshold variable sample space Γ is typically taken to be $[\gamma_L, \gamma_U]$, with γ_L and γ_U chosen such that $P(q_{t-d} \leq \gamma_L) = \theta_1 > 0$ and $P(q_{t-d} \leq \gamma_U) = 1 - \theta_1$. The choice of θ_1 is commonly taken to be 10 or 15 percent. Restricting the parameter space of the threshold in this fashion ensures that there are enough observations in each regime and also guarantees the existence of nondegenerate limits for the test statistics of interest.

In what follows, we will be interested in obtaining the limiting behavior of $W_T(\gamma)$ defined in (15.6). In this context it will be important to explore the distinctive features of the limiting null distribution of the test statistic when the maintained model is either a pure multivariate unit root process with no cointegration (i.e., $\Delta Y_t = u_t$) or a VECM in the form $\Delta Y_t = \mathbf{\Pi} Y_{t-1} + u_t$ with $Rank(\mathbf{\Pi}) = r$ such that $0 < r \leq p$. The case where $r = p$ would correspond to a purely stationary specification. We note that under all these cases the null hypothesis of linearity holds. Before proceeding further it is also important to emphasize the fact that we are facing a nonstandard inference problem, since under the null hypothesis the threshold parameter γ is not unidentified. This is now a well-known and documented problem in the literature on testing for the presence of various forms of nonlinearities in regression models and is commonly referred to as the Davies problem. Under a stationary setting, where $Rank(\mathbf{\Pi}) = p$ and taking γ as fixed and given, we would expect $W_T(\gamma)$ to behave like a χ^2 random variable in large samples. Since we will not be assuming that γ is known, however, we will follow Davies (1977, 1987) and test the null hypothesis of linearity using $SupW = \sup_{\gamma \in \Gamma} W_T(\gamma)$. In what follows we also make use of the equality $I(q_{t-d} \leq \gamma) = I(F(q_{t-d}) \leq F(\gamma))$, which allows us to use uniform random variables (see Caner and Hansen, 2001, p. 1586). In this context we let $\lambda \equiv F(\gamma) \in \Lambda$ with $\Lambda = [\theta_1, 1 - \theta_1]$, and throughout this chapter we will be using λ and $F(\gamma)$ interchangeably.

In the following proposition we summarize the limiting behavior of the Wald statistic for testing the null hypothesis of linearity when it is assumed that the system is purely stationary.

Proposition 1 *Under assumptions A1–A3, $H_0 : \mathbf{\Pi}_1 = \mathbf{\Pi}_2$ and Y_t a p-dimensional I(0) vector we have*

$$SupW \;\Rightarrow\; Sup_{\lambda \in \Lambda} G(\lambda)' V(\lambda)^{-1} G(\lambda) \tag{15.7}$$

where $G(\lambda)$ is a zero mean p^2-dimensional Gaussian random vector with covariance $E[G(\lambda_1)G(\lambda_2)] = V(\lambda_1 \wedge \lambda_2)$ and $V(\lambda) = \lambda(1-\lambda)(Q \otimes \Omega_u)$ with $Q = E[ZZ']$.

REMARK 1: It is interesting to note that the above limiting distribution is equivalent to a normalized squared Brownian Bridge process identical to the one arising when testing for the presence of structural breaks as in Andrews (1993, Theorem 3, p. 838). The same distribution also arises in particular parameterizations of self-exciting threshold autoregressive models when only the constant terms are allowed to be different in each regime (see Chan, 1990). We also note that, for known and given γ, the quantity $G(\lambda)' V(\lambda)^{-1} G(\lambda)$ reduces to a χ^2 random variable with p^2 degrees of freedom. Since $G(\lambda)$ is $(Q \otimes \Omega_u)^{\frac{1}{2}} N(0, \lambda(1-\lambda) I_{p^2}) \equiv (Q \otimes \Omega_u)^{\frac{1}{2}} [W(\lambda) - \lambda W(1)]$, with $W(.)$ denoting a p^2-dimensional standard Brownian Motion, the result follows from the above definition of $V(\lambda)$. We also note that the limiting process is free of nuisance parameters, solely depending on the number of parameters being tested under the null hypothesis and is tabulated in Andrews (1993, table 1, p. 840). For a more extensive set of p-values of the corresponding limiting distributions, see also Hansen (1997).

In the next proposition we summarize the limiting behavior of the same Wald test statistic when the system is assumed to be a p-dimensional pure I(1) process as $\Delta Y_t = u_t$ or, alternatively, I(1) but cointegrated as in $\Delta Y_t = \alpha\beta' Y_{t-1} + u_t$, with α and β having reduced ranks. In what follows, a standard Brownian Sheet $W(s, t)$ is defined as a zero mean two-parameter Gaussian process indexed by $[0, 1]^2$ and having a covariance function given by $Cov[W(s_1, t_1), W(s_2, t_2)] = (s_1 \wedge t_1)(s_2 \wedge t_2)$, while a Kiefer process K on $[0, 1]^2$ is given by $K(s, t) = W(s, t) - tW(s, 1)$. The Kiefer process is also a two-parameter Gaussian process with zero mean and covariance function $Cov[K(s_1, t_1), K(s_2, t_2)] = (s_1 \wedge s_2)(t_1 \wedge t_2 - t_1 t_2)$.

Proposition 2 *Under assumptions A1–A3, $H_0 : \mathbf{\Pi}_1 = \mathbf{\Pi}_2$ and Y_t a p-dimensional I(1) vector cointegrated or not*:

$$SupW \Rightarrow Sup_{\lambda\in\Lambda} \frac{1}{\lambda(1-\lambda)} tr\left(\int_0^1 W(r)dK(r,\lambda)'\right)'\left(\int_0^1 W(r)W(r)'\right)^{-1} \times \left(\int_0^1 W(r)dK(r,\lambda)'\right) \tag{15.8}$$

where $K(r, \lambda)$ is a Kiefer process given by $K(r, \lambda) = W(r, \lambda) - \lambda W(r, 1)$, with $W(.)$ denoting a p-dimensional standard Brownian Motion and $W(r, \lambda)$ a p-dimensional standard Brownian Sheet.

Looking at the expression of the limiting distribution in Proposition 2, we again observe that for given and known λ, the limiting random variable is $\chi^2(p^2)$, exactly as occurred under the purely stationary setup of proposition 1. This follows from the observation that $W(r)$ and $K(r, \lambda)$ are independent. Note that we have $E[W(r)K(r, \lambda)] = E[W(r)W(r, \lambda)] - \lambda E[W(r)^2]$ and since $E[W(r)W(r, \lambda)] = r\lambda$ and $E[W(r)^2] = r$ the result follows. It also follows that the limiting random variables in (15.7) and (15.8) are equivalent in distribution.

15.2.3 Simulation-based evidence

Having established the limiting behavior of the Wald statistic for testing the null of no threshold effects within the VECM type representation, we next explore the adequacy of the asymptotic approximations presented in Propositions 1–2 when dealing with finite samples. This will also allow us to explore the documented robustness of the above limiting distributions to the absence or presence of unit roots and cointegration, and to the stochastic properties of the threshold variable q_t when faced with limited sample sizes.

We initially consider a purely stationary bivariate DGP as the model under the null hypothesis, parameterised as $Y_t = \Phi Y_{t-1} + u_t$ with $\Phi = diag(0.5, 0.8)$ and

$u_t = NID(0, I_2)$. As a candidate threshold variable required in the construction of the Wald statistic, we consider two options; one in which q_t is taken to be a normal iid random variable (independent of $u_{it}, i = 1, 2$) and one where q_t follows a stationary AR(1) process given by $q_t = \theta q_{t-1} + \epsilon_t$, with $\theta = 0.5$ and $\epsilon_t = NID(0, 1)$ with $Cov(\epsilon_t, u_{is}) = 0 \ \forall t, s$ and $i = 1, 2$. Regarding the magnitude of the delay parameter, we set $d = 1$ throughout all our experiments, all conducted using samples of size $T = 200, 400, 2000$ across $N = 5000$ replications and with a 10 percent trimming of the sample space of the threshold variable. Another important purpose of our experiments is to construct a range of critical values for the distributions presented in (15.7)–(15.8) and to compare them with the corresponding tabulations in Andrews (1993, Table 1, p. 840). Results for the purely stationary system are presented in Table 15.1.

The critical values tabulated in Table 15.1 suggest that the finite sample distributions of the Wald statistic track their asymptotic counterpart (as judged by a sample of size T = 2,000) very accurately. As discussed in Remark 1 above, we can also observe that the critical values obtained in Andrews (1993) are virtually identical to the ones obtained using our DGPs and multivariate framework with thresholds (note that within our bivariate VAR we are testing for the presence of threshold effects across p^2 parameters).

In Tables 15.2 and 15.3 we concentrate on the limiting and finite sample behavior of the Wald statistic for testing the absence of threshold effects when the true DGP is a system of I(1) variables. Table 15.2 focuses on the case of a purely I(1) system with no cointegration, given by $\Delta Y_t = u_t$, while Table 15.3 focuses on a cointegrated system given $\Delta y_{1t} = u_{1t}$ and $y_{2t} = 0.8y_{2t-1} + u_{2t}$. In this latter case the bivariate system is characterized by the presence of one stationary relationship and the corresponding rank of the long-run impact matrix is one. The dynamics of q_t were maintained as above in both sets of experiments.

The empirical results presented in Tables 15.2–15.3 clearly illustrate the robustness of the limiting distributions to various parameterizations of the threshold variable. Our tabulations also corroborate our earlier observation that the limiting distributions are unaffected by the presence or absence of I(1) components.

Table 15.1 Critical values under an I(0) system and $p^2 = 4$

	T	90%	95%	99%
		$q_t : NID(0,\ 1)$		
SupW	200	14.946	16.909	21.246
SupW	400	14.606	16.686	21.239
SupW	2,000	14.762	16.596	20.741
		$q_t : AR(1)$		
SupW	200	15.135	17.252	21.331
SupW	400	14.836	17.024	21.323
SupW	2,000	14.829	16.737	20.854
Andrews	∞	14.940	16.980	21.040

Table 15.2 Critical values under a pure I(1) system and $p^2 = 4$

	T	90%	95%	99%
		$q_t : NID(0,\ 1)$		
SupW	200	14.970	17.023	22.098
SupW	400	14.858	18.578	21.205
SupW	2,000	15.012	16.947	20.967
		$q_t : AR(1)$		
SupW	200	15.369	17.197	22.164
SupW	400	14.948	18.527	21.358
SupW	2,000	14.904	16.840	21.212
Andrews	∞	14.940	16.980	21.040

Table 15.3 Critical values under a cointegrated system and $p^2 = 4$

	T	90%	95%	99%
		$q_t : NID(0,\ 1)$		
SupW	200	15.030	17.236	21.431
SupW	400	14.685	16.879	20.926
SupW	2,000	14.723	16.739	20.911
		$q_t : AR(1)$		
SupW	200	15.068	16.903	21.074
SupW	400	15.150	17.040	21.153
SupW	2,000	14.961	16.758	21.013
Andrews	∞	14.940	16.980	21.040

15.3 Estimation of the threshold parameter

Once inferences based on the Wald test reject the null hypothesis of a linear VECM, our next objective is to obtain a consistent estimator of the threshold parameter. The model under which we operate is now given by $\Delta Y = \mathbf{\Pi}_1 Z_1 + \mathbf{\Pi}_2 Z_2 + U$. We propose to obtain an estimator of γ based on the least squares principle. Letting $\hat{U}(\gamma) = \Delta Y - \hat{\mathbf{\Pi}}_1(\gamma)Z_1(\gamma) - \hat{\mathbf{\Pi}}_2(\gamma)Z_2(\gamma)$, we consider

$$\hat{\gamma} = \arg\min_{\gamma \in \Gamma} |\hat{U}(\gamma)\hat{U}(\gamma)'|. \tag{15.9}$$

Before establishing the large sample behavior of $\hat{\gamma}$ introduced in (15.9), it is important to highlight the fact that a VECM type of representation with threshold effects as in (15.4) is compatible with either a purely stationary Y_t or a system of I(1) variables that is cointegrated in a conventional sense and with threshold effects present in its adjustment process. Examples of such processes are provided in (15.2) and (15.3) above, while a formal discussion of the stationarity properties of Y_t generated from (15.4) is provided below.

The following proposition summarizes the limiting behavior of the threshold parameter estimator defined above, with γ_0 referring to its true magnitude.

Proposition 3 *Under assumptions (A1)–(A3) with Y_t I(0) or I(1) but cointegrated and generated as in* (15.4) *we have $\hat{\gamma} \xrightarrow{p} \gamma_0$ as $T \to \infty$.*

From the above proposition it is clear that the consistency property of the threshold parameter estimator remains unaffected by the presence of I(1) components. In order to empirically illustrate the above proposition, and to explore the behavior of $\hat{\gamma}$ in smaller samples, we conducted a Monte-Carlo experiment covering a range of parameterizations, including purely stationary and cointegrated systems. Our objective was to assess the finite sample performance of the least squares based estimator of γ_0 in moderate to large samples in terms of bias and variability.

For the purely stationary case we consider the specification introduced in (15.3), setting $(\rho_{11}, \rho_{21}) = (-0.8, -0.4)$ and $(\rho_{12}, \rho_{22}) = (-0.2, -0.6)$. Regarding the choice of threshold variable, we consider the case of a purely Gaussian iid process as well as an AR(1) specification given by $q_t = 0.5q_{t-1} + u_t$ with $u_t = NID(0, 1)$. The true threshold parameter is set to $\gamma_0 = 0.25$ under the AR(1) dynamics and to $\gamma_0 = 0$ when q_t is iid. The delay parameter is fixed at $d = 1$. For the cointegrated case we consider a system given by $y_{1t} = 2y_{2t} + z_t$ with $\Delta y_{2t} = \epsilon_{2t}$ and $z_t = 0.2z_{t-1}I(q_{t-1} \leq \gamma_0) + 0.8z_{t-1}I(q_{t-1} > \gamma_0) + v_t$, while retaining the same dynamics for q_t and the same threshold parameter configurations as above. Both ϵ_{2t} and v_t are chosen as NID(0,1) random variables.

Results for these two classes of DGPs are presented in Table 15.4, which displays the empirical mean and standard deviation of $\hat{\gamma}$ estimated as in (15.9) using samples of size $T = 200$ and $T = 400$ across $N = 5{,}000$ replications.

From both of the above experiments we note that $\hat{\gamma}$ as defined in (15.9) displays a reasonably small and negative finite sample bias of approximately 0.5 percent under both configurations of the dynamics of the threshold variable and system properties. At the same time, however, we note that $\hat{\gamma}$ is characterized by a substantial variability across all model configurations. Its empirical standard deviation is virtually twice the magnitude of γ_0 under $T = 200$ and, although clearly

Table 15.4 Empirical mean and standard deviation of $\hat{\gamma}$

	q_t *AR*(1), $\gamma_0 = 0.15$		q_t *iid*, $\gamma_0 = 0$	
	$E(\hat{\gamma})$	$Std(\hat{\gamma})$	$E(\hat{\gamma})$	$Std(\hat{\gamma})$
	Stationary system			
$T = 200$	0.142	0.278	−0.014	0.247
$T = 400$	0.145	0.108	−0.004	0.100
	Cointegrated system			
$T = 200$	0.140	0.266	−0.006	0.229
$T = 400$	0.144	0.101	−0.003	0.091

declining with sample size, remains substantial even under $T = 40$. Similar features of threshold parameter estimators have also been documented in Gonzalo and Pitarakis (2002).

Taking the presence of threshold effects as given, together with the availability of a consistent estimator of the unknown threshold parameter, our next concern is to explore further the stochastic properties of the p-dimensional vector Y_t.

15.4 Stochastic properties of the system and rank configuration of the VECM with threshold effects

So far the test developed in the previous sections allows us to decide whether the inclusion of threshold effects into a VECM-type specification is supported by the data. Given the simplicity of its implementation, and the fact that the limiting distribution of the test statistic is unaffected by the stationarity properties of the variables being modeled, the proposed Wald-based inferences can be viewed as a useful pre-test before implementing a formal analysis of the integration and cointegration properties of the system. If the null hypothesis is not rejected, we can proceed with the specification of a linear VECM using the methodology developed in Johansen (1995 and references therein).

Our next concern is to explore the implications of the rejection of the null hypothesis of linearity for the stability and, when applicable, cointegration properties of Y_t, whose dynamics are now known to be described by the specification (15.4). Although rejecting the hypothesis that $\mathbf{\Pi}_1 = \mathbf{\Pi}_2$ rules out the scenario of a purely I(1) system with no cointegration as traditionally defined, since having $\mathbf{\Pi}_1 \neq \mathbf{\Pi}_2$ is trivially incompatible with the specification $\Delta Y = U$, as shown below, it remains possible that the system is either purely covariance stationary or I(1) with cointegration in a sense to be made clear (see, for example, the formulation in (15.2) under Example 1).

15.4.1 Stability properties of the system

In the context of our specification in (15.4), and maintaining the notation $\mathbf{\Phi}_1 = I_p + \mathbf{\Pi}_1$ and $\mathbf{\Phi}_2 = I_p + \mathbf{\Pi}_2$, so that the system can be formulated as $Y_t = \mathbf{\Phi}_t Y_{t-1} + u_t$ with $\mathbf{\Phi}_t = \mathbf{\Phi}_1 I(q_{t-d} \leq \gamma) + \mathbf{\Phi}_2 I(q_{t-d} > \gamma)$, the stability properties of the system are summarized in the following proposition, where for a square matrix $\mathbf{M}$ the notation $\boldsymbol{\rho}(\mathbf{M})$ refers to its spectral radius.

Proposition 4 *Under assumptions (A1)–(A3), Y_t generated from (15.4) is covariance stationary iff* $\boldsymbol{\rho}(F(\gamma)(\mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1) + (1 - F(\gamma))(\mathbf{\Phi}_2 \otimes \mathbf{\Phi}_2)) < 1$.

From the above proposition it is interesting to note that, even if one of the two regimes has a root on the unit circle, the model could still be covariance stationary. In fact, the system could even be characterized by an explosive behavior in one of its regimes while still being covariance stationary, if, for example, the magnitudes of the transition probabilities are such that switching occurs very often. Note also that the condition ensuring the covariance stationarity of Y_t is also equivalent to requiring the eigenvalues of $E[\mathbf{\Phi}_t \otimes \mathbf{\Phi}_t]$ to have moduli less than one.

EXAMPLE 3: We can here consider the example of a bivariate process given by $Y_t = I_2 Y_{t-1} I(q_{t-d} \leq \gamma) + \mathbf{\Phi}_2 Y_{t-1} I(q_{t-d} > \gamma) + u_t$ and let $\Phi_2 = \phi I_2$ with $|\phi| < 1$, where I_2 denotes a two-dimensional identity matrix. This system can be seen to be characterized by a random walk type of behavior in one regime and covariance stationarity in the second regime. In matrix form we have

$$\begin{aligned}\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} \leq \gamma) \\ &\quad + \begin{pmatrix} \phi - 1 & 0 \\ 0 & \phi - 1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} > \gamma) + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.\end{aligned} \tag{15.10}$$

Letting $\mathbf{M} = F(\gamma)(\mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1) + (1 - F(\gamma))(\mathbf{\Phi}_2 \otimes \mathbf{\Phi}_2))$, it is straightforward to establish that, in the case of (15.10), we have $\boldsymbol{\rho}(\mathbf{M}) = F(\gamma) + \phi^2(1 - F(\gamma)) < 1$, since $\phi^2 < 1$, thus implying that $Y_t = (y_{1t}, y_{2t})'$ is covariance stationary.

EXAMPLE 4: Another example of a covariance stationary system is given by

$$\begin{aligned}\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & \phi - 1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} \leq \gamma) \\ &\quad + \begin{pmatrix} \phi - 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} I(q_{t-1} > \gamma) + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}\end{aligned} \tag{15.11}$$

for which we have $\boldsymbol{\rho}(\mathbf{M}) = (1 - F(\gamma))(1 - \phi)^2 < 1$ if $F(\gamma) < 0.5$, and $\rho(M) = F(\gamma)(1 - \phi)^2 < 1$ if $F(\gamma) > 0.5$. On the other hand, if we concentrate on the specification given in (15.2), it is straightforward to establish that $\rho(M) = 1$, thus violating the requirement for Y_t to be covariance stationary.

For later use it is also important at this stage to observe the correspondence between the ranks of the long-run impact matrices presented in the above example and the covariance stationarity of each system. In example 3, for examples, we note that $r_1 \equiv Rank(\mathbf{\Pi}_1) = 0$ and $r_2 \equiv Rank(\mathbf{\Pi}_2) = 2$, while in model (15.11) we have $(r_1, r_2) = (1, 1)$. This highlights the fact that within a nonlinear specification, as in (15.4), the correspondence between the rank structure of the long-run impact matrices and the stability/cointegration properties of the system will be less clearcut than within a simple linear VECM. Before exploring further this issue, it will be important to clarify the type of threshold nonlinearities that are compatible with an I(1) system and its VECM representation in (15.4).

15.4.2 I(1)ness and cointegration within a nonlinear VECM

The recent literature on the inclusion of nonlinear features in models with I(1) variables and cointegration can typically be categorized into two strands. Single equation approaches aim to detect the presence of nonlinearities in regressions with I(1) processes known to be cointegrated (see Saikkonen and Choi, 2004;

Hong, 2003; Arai, 2004). In Saikkonen and Choi (2004), the authors included a smooth transition type of function $g(.)$ within a postulated cointegrating regression model of the form $y_{1t} = \beta y_{2t} + \theta y_{2t} g(y_{2t}; \gamma) + u_t$ and proposed a methodology for testing the null hypothesis of no such effects, given here by $H_0 : \theta = 0$. The presence of such nonlinearities within a cointegrating relationship implies some form of switching equilibria, in the sense that the cointegrating vector is allowed to be different depending on the magnitude of y_{2t}. In both Hong (2003) and Arai (2004), the authors focused on a similar setup without an explicit choice of functional form. This was achieved through the inclusion of additional polynomial terms in the y_2 variable on the right-hand side of a cointegrating regression.

Another strand of the same literature focused on the treatment of nonlinearities within a multivariate error correction framework. The motivation underlying this research was again to detect the presence of nonlinear cointegration, but here defined as a nonlinear adjustment towards the long-run equilibrium while maintaining the assumption that the cointegration relationship is itself linear. Another important maintained assumption in this line of research is the existence of a single cointegrating vector (see Balke and Fomby, 1997; Seo and Hansen, 2002; Seo, 2004). Regarding the theoretical properties of multivariate models with nonlinearities, Bec and Rahbek (2004) have explored the strict stationarity and ergodicity properties of multivariate error correction models with general cointegrating rank and nonlinearities in their adjustment process.

One aspect that seems not to have been emphasized in the literature is the fact that, when operating within a VECM-type framework, an important aspect of restricting the presence of nonlinearity to occur solely in the adjustment process stems from representation concerns. More specifically, it can be shown that two I(1) variables that are linearly cointegrated but with a nonlinear adjustment process continue to admit a "nonlinear" VECM representation similar to (15.4) above. If we also wish to explore the possibility of nonlinearities in the cointegrating relationship itself, however, it becomes difficult to justify the existence of a VECM representation à la Granger.

To highlight this point, let us consider the following simple *nonlinear* cointegrating relationship, which is characterized by the presence of a threshold type of nonlinearity

$$\begin{aligned} y_{1t} &= \beta y_{2t} + \theta y_{2t} I(q_{t-1} > \gamma) + z_t \\ \Delta y_{2t} &= \epsilon_{2t} \\ \Delta z_t &= \rho z_{t-1} + u_t \end{aligned} \tag{15.12}$$

with $\rho < 0$ and z_t representing the stationary equilibrium error.

If we were in a linear setup with $\theta = 0$, it would be straightforward to reformulate the above specification as $\Delta y_{1t} = \rho z_{t-1} + v_t$, with $v_t = u_t + \beta \epsilon_{2t}$, and we would have a traditional VECM representation with ρ playing the role of the adjustment coefficient to equilibrium and $z_{t-1} = (y_{1t-1} - \beta y_{2t-1})$ denoting the previous period's equilibrium error. At this stage it is important to note that a key aspect of the linear setup that allows us to move toward an ECM type representation is the fact that taking y_{2t} to be an I(1) variable, as in (15.12), directly implies that y_{1t} is also difference stationary, since taking the first difference of both sides of the first equation gives $\Delta y_{1t} = \beta \Delta y_{2t} + \Delta z_t$ and both the left- and right-hand sides are characterized by the same integration properties.

When we introduce nonlinearities in the relationship linking y_{1t} and y_{2t}, however, the stochastic properties of the system become less obvious. Specifically, taking y_{2t} to be I(1), or equivalently difference stationary, no longer implies that y_{1t} is also difference stationary. Indeed, it becomes straightforward to show that, although the I(1)ness of y_{2t} makes y_{1t} nonstationary, this nonstationarity of y_{1t} can no longer be removed by first differencing. Put differently, although the variance of y_{1t} behaves in a manner similar to the variance of a random walk, first differencing y_{1t} will no longer make it stationary. More formally, if we take the first difference of the first equation in (15.12) and use the notation $I_t \equiv I(q_t > \gamma)$, we have

$$\begin{aligned} \Delta y_{1t} &= \beta \Delta y_{2t} + \theta \Delta (y_{2t} I_{t-1}) + \Delta z_t \\ &= \rho z_{t-1} + \theta y_{2t-1} \Delta I_{t-1} + v_t \end{aligned} \tag{15.13}$$

where $v_t = \theta \epsilon_{2t} I_{t-1} + \beta \epsilon_{2t} + u_t$. Clearly, the presence of the term $y_{2t-1} \Delta I_t$ in the right-hand side of (15.13) precludes the possibility of a traditional ECM-type representation à la Granger. If we take q_t to be an iid process, then it is straightforward to establish that $V(y_{2t-1} \Delta I_t) = 2F(\gamma)(1 - F(\gamma))(t - 1)$. Similarly, y_{1t} cannot really be viewed as a difference stationary process, as would have been the case within a linear framework. As dicussed in Granger, Inoue and Morin (1997), where the authors introduced a specification similar to (15.13), the correct but not directly operational form of the error correction model could be formulated as

$$\Delta y_{1t} - \theta y_{2t-1} \Delta I_{t-1} = \rho z_{t-1} + v_t$$

where now both the left- and right-hand side components are stationary. Practical tools and their theoretical properties for handling models such as the above are developed in Gonzalo and Pitarakis (2005a).

Our specification in (15.13) has highlighted the difficulties of handling switching phenomena within the cointegrating relationship itself if we want to operate within the traditional VECM framework. It is also worth emphasizing that similar conceptual difficulties will arise in non-VECM-based approaches to the treatment of nonlinearities in cointegrating relationships. Writing $y_{1t} = \beta_t y_{2t} + u_t$, with y_{2t} an I(1) variable and u_t an I(0) error term, defines a stationary relationship between y_{1t} and y_{2t} which is not invalid per se. However, it would be inaccurate to refer to it as a cointegrating relationship linking two I(1) variables since y_{1t} cannot be difference stationary due to the time varying nature of β_t.

In summary, a system such as (15.12), which has a switching cointegrating vector, cannot admit a VECM representation as in (15.4) in which both the left- and right-hand sides are balanced in the sense of both being stationary. Equivalently, for an I(1) vector to admit a formal VECM representation as in (15.4), it must be the case that the threshold effects are only present in the adjustment process.

15.4.3 Rank configuration under alternative stochastic properties of Y_t

Our objective here is to explore further the correspondence between the rank characteristics of $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ and the stability properties of Y_t, akin to the well-known relationship between the rank of the long-run impact matrix of a linear VECM specification and its cointegration properties. We are interested in the rank configurations of $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ that are consistent with covariance stationarity of Y_t. Similarly, we also wish to explore the correspondence between the presence of threshold effects in the adjustment process of a cointegrated I(1) system and the rank configurations of the two long-run impact matrices that are compatible with such a system.

Within a linear VECM specification, whose corresponding lag polynomial has roots either on or outside the unit circle, it is well known that having a matrix $\mathbf{\Pi}$ that has full rank also implies that the underlying process is I(0). Although from our result in proposition 4 it is straightforward to see that, if both *or either* of $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ have full rank, then Y_t generated from (15.4) is going to be covariance stationary as well, it is also true that the full rank condition is not necessary for covariance stationarity. Our examples in (15.2) and (15.11), for instance, have illustrated the fact that two identical rank configurations, say $(r_1, r_2) = (1, 1)$ may be compatible with either a purely I(1) system as in (15.2) or a covariance stationary system as in (15.11). Similarly, Example 3 with $(r_1, r_2) = (0, 2)$ illustrated the possibility of having a covariance stationary DGP in which either $\mathbf{\Pi}_1$ or $\mathbf{\Pi}_2$ have zero rank. These observations highlight the difficulties that may arise when attempting to clearly define the meaning of "nonlinear cointegration" when operating within an Error Correction type of model.

Drawing from our analysis in section 15.4.2, if we take the a priori view that Y_t is I(1) and (15.4) is the correct specification, then it must be the case that the rejection of the null hypothesis of linearity $H_0 : \mathbf{\Pi}_1 = \mathbf{\Pi}_2$ *directly* implies that we have threshold cointegration, here understood to mean that the adjustment process has a threshold type nonlinearity driven by the external variable q_t while the cointegrating relationship itself is stable over time. Differently put, we can formulate $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ as $\mathbf{\Pi}_1 = \alpha_1 \boldsymbol{\beta}'$ and $\mathbf{\Pi}_2 = \alpha_2 \boldsymbol{\beta}'$.

At this stage it is also important to note that, even under the maintained assumption that the cointegrating relationship itself is linear and is not characterized by threshold effects, this does not necessarily imply that $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ must have identical ranks. This feature of the system can be illustrated by considering our earlier example in (15.2), in which we set $\rho_1 = 0$ and $\rho_2 < 0$. This specific parameterization implies, for instance, that $r_1 \equiv Rank(\mathbf{\Pi}_1) = 0$ and $r_2 \equiv Rank(\mathbf{\Pi}_2) = 1$. Alternatively, we could also have set $\rho_2 = 0$ and $\rho_1 < 0$, implying the rank configuration $(r_1, r_2) = (1, 0)$ within the same example. Obviously our system could also be characterized by a parameterization such as

$\rho_1 < 0$ and $\rho_2 < 0$ with a corresponding rank configuration given by $(r_1, r_2) = (1, 1)$ as in example 1.

Using our result in Proposition 4 and our discussion above, it is straightforward to observe that within a system whose characteristic roots may lie either on or outside the complex unit circle (excluding roots that induce explosive behavior), I(1)ness with cointegration characterized by threshold adjustment may only occur if the rank configuration of $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ is such that $(r_1, r_2) \in \{(0, 1), (1, 0), (1, 1)\}$. Note, however, that the scenario whereby $(r_1, r_2) = (1, 1)$ may also be compatible with a purely stationary Y_t, as in example 2 above with $\rho_{11} = 0$ and $\rho_{12} = 0$, among other possible configurations. At this stage it is also important to recall that, within our operating framework, cases involving processes that are integrated of order greater than one are ruled out. The above observations are summarized more formally in the following proposition.

Proposition 5 *Letting $r_j \equiv Rank(\mathbf{\Pi}_j)$ for $j = 1, 2$ and assuming that $p = 2$, we have that (i) Y_t is covariance stationary if either r_1 or r_2 is equal to 2, (ii) Y_t is I(1) with threshold cointegration if $(r_1, r_2) = (0, 1)$ or $(r_1, r_2) = (1, 0)$, (iii) Y_t is either covariance stationary or I(1) with threshold cointegration if $r_1 = r_2 = 1$.*

According to the above proposition, even if at most one of the two long-run impact matrices characterizing the model in (15.4) is found to have full rank, it must be that Y_t itself is covariance stationary. On the other hand, if we have a rank configuration such as $(r_1, r_2) = (0, 1)$ or $(r_1, r_2) = (1, 0)$ then this would imply that Y_t described by (4) is I(1) and the model is characterized by threshold effects in its adjustment process towards its long-run equilibrium. Intuitively, such a rank configuration captures the idea of an adjustment process that shuts off when the threshold variable q_t crosses above or below a certain magnitude given by γ. Finally, the case whereby $(r_1, r_2) = (1, 1)$ is compatible with either a purely covariance stationary system or an I(1) system with an underlying adjustment process characterized by different speeds of adjustment depending on the magnitude of q_t.

15.4.4 Estimation of r_1 and r_2

Having established the correspondence between alternative rank configurations and the stochastic properties of Y_t, our next objective is to estimate each individual rank r_1 and r_2. In what follows we will take the view that Y_t is known to be I(1), so that the rejection of the null hypothesis of linearity directly implies threshold effects in the adjustment process towards equilibrium. Furthermore, for the simplicity of the exposition, we will assume that the system under consideration is bivariate, setting $p = 2$ in (15.4). Thus we wish to decide whether $(r_1, r_2) = (0, 1)$, $(r_1, r_2) = (1, 0)$ or $(r_1, r_2) = (1, 1)$ in the true specification. Note that any other configuration of (r_1, r_2) would imply that Y_t is covariance stationary and is therefore ruled out by our operating framework.

Before introducing our proposed methodology for estimating r_1 and r_2, we define the following sample quantities. We let $\widehat{\Delta Y_1} = \Delta Y * I(q \leq \hat{\gamma})$, $\widehat{\Delta Y_2} = \Delta Y * I(q > \hat{\gamma})$ and $\hat{Z}_1$ and $\hat{Z}_2$ are as in (15.4) with γ replaced with its estimated counterpart $\hat{\gamma}$. The residual vector is obtained as $\hat{U} = \Delta Y - \hat{\Pi}_1 \hat{Z}_1 - \hat{\Pi}_2 \hat{Z}_2$ and we also define $\hat{U}_1 = \widehat{\Delta Y_1} - \hat{\Pi}_1 \hat{Z}_1$ and $\hat{U}_2 = \widehat{\Delta Y_2} - \hat{\Pi}_2 \hat{Z}_2$, from which we note the equality $\hat{\Omega} = \hat{\Omega}_1 + \hat{\Omega}_2$, where $\hat{\Omega}_1 = \hat{U}_1 \hat{U}_1'/T, \hat{\Omega}_2 = \hat{U}_2 \hat{U}_2'/T$ and $\hat{\Omega} = \hat{U}\hat{U}'/T$. For later use we also introduce the following moment matrices corresponding to each regime j

$$\begin{aligned} S_{11}^j &= \frac{\hat{Z}_j \hat{Z}_j'}{T}, \\ S_{00}^j &= \frac{\widehat{\Delta Y_j} \widehat{\Delta Y_j}'}{T}, \\ S_{01}^j &= \frac{\widehat{\Delta Y_j} \hat{Z}_j'}{T}, \\ S_{10}^j &= (S_{01}^j)' \end{aligned} \tag{15.14}$$

with $j = 1, 2$. Using (15.14) we can now reformulate the estimated covariance matrices as $\hat{\Omega}_j = S_{00}^j - S_{01}^j (S_{11}^j)^{-1} S_{10}^j, j = 1, 2$, and for later use it will also be useful to note that the eigenvalues of $(S_{00}^j)^{-1} S_{01}^j (S_{11}^j)^{-1} S_{10}^j$ are the same as those of $I - (S_{00}^j)^{-1} \hat{\Omega}_j$ for $j = 1, 2$.

We now propose to estimate the unknown ranks of $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ using a model selection approach as introduced and investigated in Gonzalo and Pitarakis (1998, 1999, 2002). We view the problem of the estimation of r_1 and r_2 from a model selection perspective in which our main task is to select the optimal model among a portfolio of nested specifications. The selection is made via the optimization of a penalized objective function. The latter has one component which decreases as the number of estimated parameters increases (e.g., as r_j increases) and another component that increases to penalize overfitting. The use of a model selection based approach for inferences similar to the above has been advocated in numerous related areas of the econometric literature. In Gonzalo and Pitarakis (2002), for example, the authors explore the properties of a model selection based approach for estimating the number of regimes of a stationary time series characterized by threshold effects. In Cragg and Donald (1997), the authors used AIC and BIC type criteria for estimating the rank of a normally distributed matrix. Similarly, in Phillips and Chao (1999) the authors developed a new information theoretic criterion used to determine the rank and short-run dynamics of error correction models.

Formally, letting $\hat{\Omega}_j(r_j)$ denote the sample covariance matrices obtained from each regime characterizing (15.4) under the restriction that $rank(\mathbf{\Pi}_j) = r_j$, our estimator of r_j is defined as

$$\hat{r}_j = \arg \min_{r_j} IC_j(r_j) \tag{15.15}$$

where

$$IC(r_j) = \ln |\hat{\Omega}_j(r_j)| + \frac{c_T}{T} m(r_j) \tag{15.16}$$

with $m(r_j)$ denoting the number of estimated parameters (here $m(r_j) = 2pr_j - r_j^2$) and c_T a deterministic penalty term. Next, using the fact that

$$\ln|\hat{\Omega}_j(r_j)| = \ln|S_{00}^j| + \sum_{i=1}^{r_j}(1 - \hat{\lambda}_i^j) \tag{15.17}$$

and noting that S_{00}^j is independent of the magnitude of r_j, we can instead focus on the optimization of the following modified criterion

$$\overline{IC}(r_j) = \sum_{i=1}^{r_j} \ln(1 - \hat{\lambda}_i^j) + \frac{c_T}{T}(2pr_j - r_j^2). \tag{15.18}$$

A clear advantage of using (15.18) stems from the simplicity of its empirical implementation, requiring solely the availability of the eigenvalues of $I - (S_{00}^j)^{-1}\hat{\Omega}_j$ for $j = 1, 2$. It is also interesting to observe the close similarity between conducting inferences using (15.18) and, for example, a formal likelihood ratio-based testing procedure. Focusing on the estimation of r_1, our model selection-based approach involves selecting $\hat{r}_1 = 0$ as the optimal choice if $\overline{IC}(r_1 = 0) < \overline{IC}(r_1 = 1)$ and $\hat{r}_1 = 1$ if $\overline{IC}(r_1 = 1) < \overline{IC}(r_1 = 0)$. Equivalently, the model selection-based approach points to $\hat{r}_1 = 1$ if $-T\ln(1 - \hat{\lambda}_1^1) > 3c_T$ and to $\hat{r}_1 = 0$ otherwise under a bivariate setting. This is equivalent to the formulation of a likelihood ratio statistic for testing the null $H_0 : r_1 = 0$ against $H_1 : r_1 = 1$, except that here the decision rule is dictated by the magnitude of the penalty term and the number of estimated parameters. A formal distribution theory for an LR test-based approach for the determination of r_1 and r_2 à la Johansen can be found in Gonzalo and Pitarakis (2005a). We next summarize the asymptotic properties of the model selection approach in the following proposition.

Proposition 6 *Letting r_j^0 denote the true rank of Π_j for $j = 1, 2$, $\hat{r}_j$ defined as in (15.15), with c_T such that (i) $c_T \to \infty$ and (ii) $c_T/T \to 0$ as $T \to \infty$, we have $\hat{r}_j \xrightarrow{p} r_j^0$.*

The above proposition establishes the weak consistency of the rank estimators obtained through the model selection-based approach. A possible candidate for the choice of the penalty term satisfying both (i) and (ii) is $c_t = \ln T$, corresponding to the well-known BIC-type criterion. It is clear, however, that other functionals of the sample size may be equally valid (e.g. $c_T = 2\ln\ln T$), making it difficult to argue in favor of a universally optimal criterion.

Having established the limiting properties of our rank estimators, we next concentrate on their finite and large sample performance across a wide range of possible model configurations. Following Gonzalo and Pitarakis (2002), we implement our experiments using $c_T = \ln T$ as the penalty term in (15.18).

We initially consider the DGP given in (15.2) under example 1. We have a bivariate system that is I(1) with a single cointegrating vector $(1, -\beta)$. We set $\beta = 2$ and consider $(\rho_1, \rho_2) = (0, -0.4)$, so that the system is characterized by a true rank configuration given by $(r_1, r_2) = (0, 1)$. In a second set of experiments we set

$(\rho_1, \rho_2) = (-0.2, -0.6)$, so that this second system has $(r_1, r_2) = (1, 1)$. Our results are summarized in Table 15.5, which presents the decision frequencies for each possible magnitude of r_j. Throughout all our experiments q_t is assumed to follow the AR(1) process given by $q_t = 0.5q_{t-1} + \epsilon_t$ with $\epsilon_t = iid(0, 1)$ and the true threshold parameter is set at $\gamma_0 = 0$. As in our earlier experiments the delay parameter is set at $d = 1$ throughout.

From the decision frequencies presented in Table 15.5 it is clear that the proposed model selection procedure performs remarkably well across the three alternative specifications. As expected from our result in Proposition 6, it is selecting the true magnitude of each rank 100% of the times under $T = 1000$, while maintaining very high correct decision frequencies even under $T = 200$. Under the specification in (2), for instance, with $(r_1^0, r_2^0) = (0, 1)$, the procedure picked $r_1 = 0$ about 85% of the times and $r_2 = 1$ 100% of the times under $T = 200$, with the correct decision frequency increasing to about (93%, 100%) under $T = 400$.

To provide further empirical support for our proposed approach, we next consider a set of threshold DGPs that restrict Y_t to being covariance stationary. For this purpose we have focused on the specification given in (15.3) under Example 2 and considered two alternative rank configurations. First, imposing $(\rho_{11}, \rho_{12}) = (0, 0)$ and $(\rho_{21}, \rho_{22}) = (-0.2, -0.4)$, we have a covariance stationary system with $(r_1, r_2) = (0, 2)$. Second, setting $(\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}) = (-0.4, 0.0, 0.0, -0.2)$, we have another covariance stationary system, this time with $(r_1, r_2) = (1, 1)$. All simulation results are presented in Table 15.6.

From the empirical decision frequencies presented in Table 15.6 it is again the case that the various estimators of r_1 and r_2 point to their true counterparts as T is allowed to increase. Although the accuracy of the estimators is somehow determined by the DGP specific parameters, it is also clear that, under both experiments, the frequency of selecting the true rank is high, reaching levels of between 90 and 100% accuracy.

Table 15.5 Decision frequencies in an I(1) system

	$\hat{r}_1 = 0$	$\hat{r}_1 = 1$	$\hat{r}_1 = 2$	$\hat{r}_2 = 0$	$\hat{r}_2 = 1$	$\hat{r}_2 = 2$
			$(r_1^0 = 0, r_2^0 = 1), \beta = 2, (\rho_1, \rho_2) = (0.0, -0.4)$			
$T = 200$	85.26	14.74	0.00	0.00	100.00	0.00
$T = 400$	93.42	6.58	0.00	0.00	100.00	0.00
$T = 1{,}000$	100.00	0.00	0.00	0.00	100.00	0.00
			$(r_1^0 = 1, r_2^0 = 1), \beta = 2, (\rho_1, \rho_2) = (-0.2, -0.6)$			
$T = 200$	34.76	65.24	0.00	0.02	99.98	0.00
$T = 400$	10.16	89.84	0.00	0.00	100.00	0.00
$T = 1{,}000$	0.00	100.00	0.00	0.00	100.00	0.00
			$(r_1^0 = 1, r_2^0 = 0), \beta = 2, (\rho_1, \rho_2) = (-0.4, 0.0)$			
$T = 200$	0.02	99.98	0.00	84.76	15.24	0.00
$T = 400$	0.00	100.00	0.00	93.50	0.00	0.00
$T = 1{,}000$	0.00	100.00	0.00	100.00	0.00	0.00

Table 15.6 Decision frequencies in a stationary system

	$\hat{r}_1=0$	$\hat{r}_1=1$	$\hat{r}_1=2$	$\hat{r}_2=0$	$\hat{r}_2=1$	$\hat{r}_2=2$
		$(r_1^0=0, r_2^0=2), (\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}) = (0.0, 0.0, -0.2, -0.4)$				
$T=200$	88.36	10.24	1.40	0.00	0.00	100.00
$T=400$	94.16	5.32	0.52	0.00	0.00	100.00
$T=1{,}000$	100.00	0.00	0.00	0.00	0.00	100.00
		$(r_1^0=1, r_2^0=1), (\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}) = (-0.4, 0.0, 0.0, -0.2)$				
$T=200$	0.00	86.90	13.10	0.56	86.94	12.50
$T=400$	0.00	90.38	0.00	0.00	91.00	9.00
$T=1{,}000$	0.00	92.64	7.36	0.00	92.96	7.04

15.5 A nonlinear permanent and transitory decomposition

Having established the threshold cointegration properties of Y_t, we next investigate how this vector process of interest can be decomposed into a permanent and transitory component following the methodology developed in Gonzalo and Granger (1995).

Recall that in the linear case with Y_t following a VECM of the form $\Delta Y_t = \alpha\beta' Y_{t-1} + u_t$, we are interested in decomposing the p-dimensional vector Y_t into two sets of components as

$$Y_t = A_1 f_t + \tilde{Y}_t \tag{15.19}$$

where A_1 is the $p \times (p-r)$ loading matrix, f_t the $(p-r) \times 1$ common I(1) factors and $\tilde{Y}_t$ is the I(0) component. The above decomposition of Y_t is such that the factors f_t are linear combinations of Y_t and $A_1 f_t$ and $\tilde{Y}_t$ form a Permanent-Transitory decomposition (see Gonzalo and Granger, 1995 for the detailed definitions of each component).

As shown in Gonzalo and Granger (1995), the above two conditions are sufficient to identify the permanent and transitory components. Formally we can write

$$Y_t = A_1 f_t + A_2 z_t \tag{15.20}$$

with $f_t = \alpha_\perp Y_t,\ z_t = \beta' Y_t$ and $A_1 = \beta_\perp(\alpha'_\perp \beta_\perp)^{-1},\ A_2 = \alpha(\beta'\alpha)^{-1}$. Note that $\alpha'_\perp \alpha = \beta'_\perp \beta = 0$.

Now, let us consider the following VECM with threshold effects

$$\Delta Y_t = \alpha_1 \beta' Y_{t-1} I(q_{t-d} \leq \gamma) + \alpha_2 \beta' Y_{t-1} I(q_{t-d} > \gamma) + u_t.$$

Following the same reasoning as in Gonzalo and Granger (1995), it is now straightforward to establish the following Threshold Permanent-Transitory decomposition for Y_t

$$Y_t = A_1 f_{1t} I(q_{t-d} \leq \gamma) + A_2 f_{2t} I(q_{t-d} > \gamma) + A_3 I(q_{t-d} \leq \gamma) + A_4 I(q_{t-d} > \gamma) z_t \quad (15.21)$$

where $f_{1t} = \alpha'_{1\perp} Y_t, f_{2t} = \alpha_{2\perp} Y_t$ and $z_t = \beta' Y_t$. The corresponding loading matrices are then given by $A_1 = \beta_\perp (\alpha'_{1\perp} \beta_\perp)^{-1}, A_2 = \beta_\perp (\alpha'_{2\perp} \beta_\perp)^{-1}$ and, similarly, $A_3 = \alpha_1 (\beta' \alpha_1)^{-1}$ and $A_4 = \alpha_2 (\beta' \alpha_2)^{-1}$. Given our estimator of the threshold parameter γ defined in (15.9), together with the corresponding sample moment matrices introduced in (15.14), the practical implementation of the above Threshold Permanent and Transitory decomposition becomes straightforward (see Gonzalo and Pitarakis, 2005b) and is obtained following the same approach as in Gonzalo and Granger (1995).

Despite the representational complications that would arise if we were to also allow the cointegrating vector β to be characterized by the presence of threshold effects, as say $\beta_t = \beta_1 I(q_{t-d} \leq \gamma) + \beta_2 I(q_{t-d} > \gamma)$ (see our discussion in section 15.4.2), the above threshold-based decomposition would translate naturally to such a framework by reformulating it as $Y_t = A_1 f_{1t} I(q_{t-d} \leq \gamma) + A_2 f_{2t} I(q_{t-d} > \gamma) + A_3 z_{1t} I(q_{t-d} \leq \gamma) + A_4 z_{2t} I(q_{t-d} > \gamma)$, with $z_{1t} = \beta'_1 Y_t, z_{2t} = \beta'_2 Y_t$. The corresponding loading matrices would then be given by $A_1 = \beta_{1\perp} (\alpha'_{1\perp} \beta_{1\perp})^{-1}, A_2 = \beta_{2\perp} (\alpha'_{2\perp} \beta_{2\perp})^{-1}$, $A_3 = \alpha_1 (\beta'_1 \alpha_1)^{-1}$ and $A_4 = \alpha_2 (\beta'_2 \alpha_2)^{-1}$.

15.6 Conclusions

This chapter has focused on the issue of introducing and testing for threshold-type nonlinear behavior into the conventional multivariate error correction model. The threshold nonlinearities we considered were driven by a stationary and external random variable triggering the regime switches. Within this context we obtained the limiting properties of a Wald-type test statistic for testing for the presence of such threshold effects characterizing the long-run impact matrix of the VECM. An interesting property of the proposed test is its robustness to the presence or absence of unit roots in the system, displaying the same limiting null distribution under a wide range of stochastic properties of the system.

We subsequently proceeded with the interpretation and further analysis of the system following a rejection of the null hypothesis of linearity. We showed that cointegration as traditionally defined was compatible with such an error correction type specification only if the nonlinearities are present in the adjustment process rather than the long-run equilibrium itself. We then introduced a model selection-based approach designed to gain further insight into the stochastic properties of the system through the determination of the rank structure of the long-run impact matrices characterizing each regime. This then allowed us to extend the permanent and transitory decomposition of Gonzalo and Granger (1995) into a nonlinear permanent and transitory decomposition.

Much remains to be done in the area of nonlinear multivariate specifications such as the VAR/VECMs considered here. In this chapter we restricted our analysis to models with no deterministic trends. Similarly, our results also ignored the possibility of having such components together with the lagged dependent variables and cointegrating vectors displaying threshold switching behavior.

Extensions along these lines, together with a formal representation theory for such models, are topics currently being investigated by the authors.

Appendix

Lemma A1: *Under assumptions A1–A3 and Y_t a p-dimensional vector of I(0) variables we have as $T \to \infty$*

(a) $\frac{ZZ'}{T} \xrightarrow{p} Q \equiv E[ZZ']$,
(b) $\frac{Z_1 Z_1'}{T} \xrightarrow{p} F(\gamma)Q$, $\frac{Z_2 Z_2'}{T} \xrightarrow{p} (1 - F(\gamma))Q$,
(c) $\frac{UZ'}{T} \xrightarrow{p} 0$, $\frac{UZ_j'}{T} \xrightarrow{p} 0$ for $j = 1, 2$,
(d) $\hat{\Omega}_u \xrightarrow{p} \Omega_u$.

where Q denotes a positive definite $p \times p$ matrix.

Proof: Under the stated assumptions parts (a) and (d) follow directly from the ergodic theorem. Parts (b) and (d) follow from Lemma 1 in Hansen (1996) and part (e) is obvious.

Lemma A2: *Letting $H_T(\gamma) \equiv \frac{1}{\sqrt{T}}(Z_1 \otimes I) vec\ U$, under assumptions A1–A3 and Y_t a p-dimensional vector of I(0) variables we have $H_T(\gamma) \Rightarrow H(\gamma)$ as $T \to \infty$, where $H(\gamma)$ is a zero mean Gaussian process with covariance kernel $F(\gamma_1 \wedge \gamma_2)(Q \otimes \Omega_u)$.*

Proof: The use of the central limit therem for martingale differences applied to the sequence $\{Y_{t-1} u_t I(q_{t-d} \leq \gamma)\}$ leads to the required Gaussianity for each $\gamma \in \Gamma$. This, combined with the componentwise tightness of $H_T(\gamma)$, which follows from Hansen (1996, Theorem 1), leads to the desired result.

Proof of Proposition 1: From Lemma A1 it directly follows that

$$(Z_2 Z_2'/T)(ZZ'/T)^{-1}(Z_1 Z_1'/T) \otimes \hat{\Omega}_u^{-1} \xrightarrow{p} F(\gamma)(1 - F(\gamma))Q \otimes \Omega_u^{-1} \tag{15.22}$$

and the Wald statistic in (15.6) can be formulated as

$$W_T(\gamma) = F(\gamma)(1 - F(\gamma))\sqrt{T}(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2)'(Q \otimes \Omega_u^{-1})\sqrt{T}(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2) + o_p(1). \tag{15.23}$$

Standard least squares algebra together with Lemma A1 also imply

$$\begin{aligned} \sqrt{T}(\hat{\boldsymbol{\pi}}_1 - \boldsymbol{\pi}) &= \sqrt{T}[(Z_1 Z_1')^{-1} Z_1 \otimes I_p] vec\, U \\ &= \left[\left(\frac{Z_1 Z_1'}{T}\right)^{-1} \otimes I_p\right] \frac{1}{\sqrt{T}}(Z_1 \otimes I_p) vec\, U \\ &= \frac{1}{F(\gamma)}(Q^{-1} \otimes I_p)\frac{1}{\sqrt{T}}(Z_1 \otimes I_p) vec\, U + o_p(1) \end{aligned} \tag{15.24}$$

and

$$\sqrt{T}(\hat{\pi}_2 - \pi) = \left[\left(\frac{Z_2 Z_2'}{T}\right)^{-1} \otimes I_p\right] \frac{1}{\sqrt{T}}(Z_2 \otimes I_p) vec\, U$$
$$= \frac{1}{(1 - F(\gamma))}(Q^{-1} \otimes I_p)\frac{1}{\sqrt{T}}(Z_2 \otimes I_p) vec\, U + o_p(1). \tag{15.25}$$

Combining (15.24) and (15.25) above and using the fact that $Z_2 = Z - Z_1$, we have

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2) = \frac{(Q^{-1} \otimes I)}{F(\gamma)(1 - F(\gamma))}\left[\frac{1}{\sqrt{T}}(Z_1 \otimes I) vec\, U - F(\gamma)\frac{1}{\sqrt{T}}(Z \otimes I) vec\, U\right] + o_p(1). \tag{15.26}$$

We can now write the Wald statistic as

$$W_T(\gamma) = \left[\frac{1}{\sqrt{T}}(Z_1 \otimes I) vec\, U - F(\gamma)\frac{1}{\sqrt{T}}(Z \otimes I) vec\, U\right]' V(\gamma)^{-1}$$
$$\times \left[\frac{1}{\sqrt{T}}(Z_1 \otimes I) vec\, U - F(\gamma)\frac{1}{\sqrt{T}}(Z \otimes I) vec\, U\right] + o_p(1) \tag{15.27}$$

where $V(\gamma) = F(\gamma)(1 - F(\gamma))(Q \otimes \Omega_u)$. Next letting $G_T(\gamma) \equiv [(Z_1 \otimes I) vec\, U - F(\gamma)(Z \otimes I) vec\, U]/\sqrt{T}$, Lemmas A1–A2, together with the fact that $\frac{1}{\sqrt{T}}(Z \otimes I) vec\, U \xrightarrow{d} N(0, Q \otimes \Omega_u)$, which follows directly from the CLT, imply $G_T(\gamma) \Rightarrow G(\gamma)$, where $G(\gamma)$ is a zero mean Gaussian random vector with covariance $E[G(\gamma_1)G(\gamma_2)] = V(\gamma_1 \wedge \gamma_2) \equiv F(\gamma_1 \wedge \gamma_2)(1 - F(\gamma_1 \wedge \gamma_2))(Q \otimes \Omega_u)$. It now follows that the limiting distribution of the Wald statistic $W_T(\gamma)$ is given by $W_T(\gamma) \Rightarrow G(\gamma)' V(\gamma)^{-1} G(\gamma)$ and the final result follows from the continuous mapping theorem.

Lemma A3: *Under assumptions A1–A3 and Y_t a p-dimensional vector of I(1) variables with $\Delta Y = U$ we have as $T \to \infty$*

(a) $\frac{ZZ'}{T^2} \Rightarrow \int_0^1 W(r)W(r)'dr$,
(b) $\frac{Z_1 Z_1'}{T^2} \Rightarrow F(\gamma)\int_0^1 W(r)W(r)'dr$,
(c) $\frac{Z_2 Z_2'}{T^2} \Rightarrow (1 - F(\gamma))\int_0^1 W(r)W(r)'dr$

where $W(r)' = (W_1(r), \ldots, W_p(r))$ is a p-dimensional standard Brownian motion.

Proof: Part (a) follows directly from Phillips and Durlauf (1986). For part (b) we first write

$$\frac{Z_1 Z_1'}{T^2} = F(\gamma)\frac{ZZ'}{T^2} + \frac{W_1 W_1'}{T^2} \tag{15.28}$$

where $W_1 W_1'$ stacks the elements of the form $Y_{t-1}Y_{t-1}'(I(q_{t-d} \leq \gamma) - F(\gamma))$. It now suffices to show that $\frac{W_1 W_1'}{T^2} = o_p(1)$. We let $S_t = \sum_{i=1}^{t}(I(q_{t-1} \leq \gamma) - F(\gamma))$ and with no loss of generality set $d = 1$ and take zero initial conditions. Using summation by parts we can write $\sum_{t=1}^{T}(I(q_{t-1} \leq \gamma) - F(\gamma))Y_{t-1}Y'_{t-1} = S_{T-1}Y_TY'_T - \sum_{t=1}^{T-1} S_t(Y_{t+1}Y'_{t+1} - Y_tY'_t)$. Next, using the fact that $Y_{t+1}Y'_{t+1} = Y_tY'_t + Y_tu'_{t+1} + u_{t+1}Y'_t + u_{t+1}u'_{t+1}$, we also have

$$\begin{aligned}\frac{1}{T^2}W_1W_1' = \frac{S_{T-1}}{T}\frac{Y_TY_T'}{T} - \frac{1}{T^2}\sum_{t=1}^{T-1} Y_tu'_{t+1}S_t - \frac{1}{T^2}\sum_{t=1}^{T-1} u_{t+1}Y'_tS_t \\ - \frac{1}{T^2}\sum_{t=1}^{T-1}(u_{t+1}u'_{t+1} - \Omega_u)S_t - \frac{1}{T^2}\Omega_u\sum_{t=1}^{T-1} S_t.\end{aligned} \tag{15.29}$$

Under the maintained assumptions the ergodic theorem ensures that $S_{T-1}/T \xrightarrow{p} 0$. Since $Y_TY'_T/T$ is stochastically bounded, it thus follows that the first term in the right-hand side of (15.29) is $o_p(1)$. Next, we consider the components $y_{it}u_{jt+1}S_t$. We have $E[y_{it}u_{jt+1}S_t] = 0$ and it is also straightforward to establish that

$$\lim_{T\to\infty} E\left[\frac{1}{T^2}\sum_{t=1}^{T-1} y_{it-1}u_{jt}S_tS_t\right]^2 = 0$$

and both the second and third terms in the right-hand side of (15.29) are also $o_p(1)$. Proceeding similarly, the third and fourth components can also be seen to be $o_p(1)$ and the final result follows from (a). Part (c) can be shown to hold in exactly the same manner as part (b).

Lemma A4: *Under assumptions A1–A3 and Y_t a p-dimensional vector of I(1) variables with $\Delta Y = U$ we have as $T \to \infty$*

(a) $\frac{1}{T}(Z \otimes I_p)vec\ U \Rightarrow vec\left[\int_0^1 dW(r)W(r)'\right]$,
(b) $\frac{1}{T}(Z_1 \otimes I_p)vec\ U \Rightarrow vec\left[\int_0^1 dW(r, F(\gamma))W(r)'\right]$

Proof: Part (a) follows directly from Phillips and Durlauf (1986). For part (b), the result follows from $L_T(\gamma) \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} u_tI(q_{t-1} \leq \gamma) \Rightarrow W(r, F(\gamma))$, where $W(r, F(\gamma))$ denotes a standard Brownian Sheet (see Theorem 1 in Diebolt, Laib and Wandji, 1997; and Theorem 2 in Caner and Hansen, 2001).

Proof of Proposition 2 We assume that the underlying null model is a pure unit root process as $\Delta Y = U$. Within the present I(1) framework we consider the following normalization of the Wald statistic

$$T(\hat{\pi}_1 - \hat{\pi}_2)'\left[\left(\frac{Z_2Z_2'}{T^2}\right)\left(\frac{ZZ'}{T^2}\right)^{-1}\left(\frac{Z_1Z_1'}{T^2}\right) \otimes \hat{\Omega}_u^{-1}\right]T(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2).$$

and with no loss of generality in what follows we will impose $\Omega_u = I_p$. Next, from Lemma A3 it follows that

$$\left[\left(\frac{Z_2 Z_2'}{T^2}\right)\left(\frac{ZZ'}{T^2}\right)^{-1}\left(\frac{Z_1 Z_1'}{T^2}\right) \otimes \hat{\Omega}_u^{-1}\right] \Rightarrow F(\gamma)(1-F(\gamma)) \int_0^1 W(r)W(r)'dr \otimes I_p \quad (15.30)$$

and we formulate the test statistic of interest as

$$W_T(\gamma) = F(\gamma)(1-F(\gamma))T(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2)'\left[\int_0^1 W(r)W(r)'dr \otimes I_p\right]T(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2) + o_p(1).$$

We next focus on the large sample behavior of $T(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2)$ when the true DGP is given by $\Delta Y = U$. We have

$$\begin{aligned} T\hat{\boldsymbol{\pi}}_1 &= \left[\left(\frac{Z_1 Z_1'}{T^2}\right)^{-1} \otimes I_p\right]\frac{1}{T}(Z_1 \otimes I_p)vec\ U \\ &= \frac{1}{F(\gamma)}\left[\left(\int_0^1 W(r)W(r)'dr\right)^{-1} \otimes I_p\right]\frac{1}{T}(Z_1 \otimes I_p)vec\ U + o_p(1). \end{aligned} \quad (15.31)$$

Proceeding similarly for $\hat{\pi}_2$ and rearranging as above we have

$$T(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_2) = \frac{1}{F(\gamma)(1-F(\gamma))}\left[\left(\int_0^1 WW'\right)^{-1} \otimes I_p\right]\left[\frac{1}{T}(Z_1 \otimes I_p)vec\, U - F(\gamma)\frac{1}{T}(Z \otimes I_p)vec\, U\right]$$

Next, using Lemma A4 it follows that

$$\begin{aligned} \frac{1}{T}(Z_1 \otimes I)vec\, U - F(\gamma)\frac{1}{T}(Z \otimes I)vec\, U \Rightarrow\ & vec\left[\int_0^1 dW(r, F(\gamma))W(r)'\right] \\ & - F(\gamma)vec\left[\int_0^1 dW(r,1)W(r)'\right] \\ &= vec\left[\int_0^1 [dW(r, F(\gamma)) - F(\gamma)dW(r,1)]W(r)'\right] \\ &= vec\left[\int_0^1 dK(r, F(\gamma))W(r)'\right] \end{aligned} \quad (15.32)$$

where we let $K(r, F(\gamma)) = W(r, F(\gamma)) - F(\gamma)W(r, 1)$. Using the above in the expression of the Wald test statistic and rearranging we obtain the required result. The case for a cointegrated system follows along the same lines.

Proof of Proposition 3 From $\hat{U}(\gamma) = \Delta Y - \hat{\mathbf{\Pi}}_1 Z_1 - \hat{\mathbf{\Pi}}_2 Z_2 + U$ we can write

$$\begin{aligned} \hat{U}(\gamma)\hat{U}(\gamma)' &= (\Delta Y - \hat{\mathbf{\Pi}}_1 Z_1 - \hat{\mathbf{\Pi}}_2 Z_2)(\Delta Y' - Z_1'\hat{\mathbf{\Pi}}_1' - Z_2'\hat{\mathbf{\Pi}}_2') \\ &= \Delta Y \Delta Y' - \Delta Y Z_1'(Z_1 Z_1')^{-1} Z_1 \Delta Y' - \Delta Y Z_2'(Z_2 Z_2')^{-1} Z_2 \Delta Y' \end{aligned} \quad (15.33)$$

where we made use of the fact that $Z_i Z_j' = 0\ \forall i \neq j$ and $i, j = 1, 2$. Next, letting γ_0

denote the true threshold parameter, we write the model evaluated at γ_0 as $\Delta Y = \mathbf{\Pi}_1 Z_1^0 + \mathbf{\Pi}_2 Z_2^0 + U$, where $Z_1^0 = (y_0 I(q_{0-d} \le \gamma_0), \ldots, y_{T-1} I(q_{T-d} \le \gamma_0))$ and $Z_2^0 = Z - Z_1^0$ with $Z_1^0 Z_2^{\prime 0} = 0$. Inserting into (15.33) and rearranging gives

$$\begin{aligned}\hat{U}(\gamma)\hat{U}(\gamma)' = {} & \mathbf{\Pi}_1 Z_1^0 Z_1^0 \mathbf{\Pi}_1' + \mathbf{\Pi}_2 Z_2^0 Z_2^0 \mathbf{\Pi}_2' + 2\mathbf{\Pi}_1 Z_1^0 U' + 2\mathbf{\Pi}_2 Z_2^0 U' + UU' - \mathbf{\Pi}_1 Z_1^0 M_1 Z_1^{0\prime} \mathbf{\Pi}_1' \\ & - \mathbf{\Pi}_2 Z_2^0 M_1 Z_2^{\prime 0} \mathbf{\Pi}_1' - 2\mathbf{\Pi}_1 Z_1^0 M_1 Z_2^{\prime 0} \mathbf{\Pi}_2' - 2\mathbf{\Pi}_1 Z_1^0 M_1 U' - 2\mathbf{\Pi}_2 Z_2^0 M_1 U' - UM_1 U' \\ & - \mathbf{\Pi}_1 Z_1^0 M_2 Z_1^{\prime 0} \mathbf{\Pi}_1' - \mathbf{\Pi}_2 Z_2^0 M_2 Z_2^{\prime 0} \mathbf{\Pi}_2' - 2\mathbf{\Pi}_1 Z_1^0 M_2 Z_2^{\prime 0} \mathbf{\Pi}_2' - 2\mathbf{\Pi}_1 Z_1^0 M_2 U' \\ & - 2\mathbf{\Pi}_2 Z_2^0 M_2 U' - UM_2 U'\end{aligned}$$

where $M_1 = Z_1'(Z_1 Z_1')^{-1} Z_1$ and $M_2 = Z_2'(Z_2 Z_2')^{-1} Z_2$. We next evaluate the limiting behavior of the above quantity for $\gamma < \gamma_0, \gamma = \gamma_0$ and $\gamma > \gamma_0$. Applying appropriate normalizations, we obtain the following uniform convergence in probability result over $\gamma \in \Gamma$ for the case $\gamma < \gamma_0$

$$\begin{aligned}\frac{\hat{U}(\gamma)\hat{U}(\gamma)'}{T} & \xrightarrow{p} (\mathbf{\Pi}_1 - \mathbf{\Pi}_2)[(G(\gamma_0) - G(\gamma))(G - G(\gamma))^{-1}(G - G(\gamma_0))](\mathbf{\Pi}_1 - \mathbf{\Pi}_2)' + \Omega_u \\ & \equiv \frac{(F(\gamma_0) - F(\gamma))(1 - F(\gamma_0))}{1 - F(\gamma)} (\mathbf{\Pi}_1 - \mathbf{\Pi}_2) Q (\mathbf{\Pi}_1 - \mathbf{\Pi}_2)' + \Omega_u.\end{aligned} \tag{15.34}$$

Proceeding similarly for the case $\gamma > \gamma_0$ we have

$$\begin{aligned}\frac{\hat{U}(\gamma)\hat{U}(\gamma)'}{T} & \xrightarrow{p} (\mathbf{\Pi}_1 - \mathbf{\Pi}_2)[G(\gamma_0) G(\gamma)^{-1}(G(\gamma) - G(\gamma_0))](\mathbf{\Pi}_1 - \mathbf{\Pi}_2)' + \Omega_u \\ & \equiv \frac{F(\gamma_0)(F(\gamma) - F(\gamma_0))}{F(\gamma)} (\mathbf{\Pi}_1 - \mathbf{\Pi}_2) Q (\mathbf{\Pi}_1 - \mathbf{\Pi}_2)' + \Omega_u.\end{aligned}$$

Finally with

$$\frac{\hat{U}(\gamma_0)\hat{U}(\gamma_0)'}{T} \xrightarrow{p} \Omega_u$$

we have that the objective function converges uniformly in probability to a nonstochastic limit that is uniquely minimized at $\gamma = \gamma_0$ and the required result follows from Theorem 2.1 in Newey and McFadden (1994).

Proof of Proposition 4 We are interested in the covariance stationarity of the stochastic recurrence given by $Y_t = \mathbf{\Phi}_1 Y_{t-1} I_{1t-d} + \mathbf{\Phi}_2 Y_{t-1} I_{2t-d} + u_t$, where we use the notation $I_{1t-d} \equiv I(q_{t-d} \le \gamma)$ and $I_{2t-d} \equiv I(q_{t-d} > \gamma)$. Note first that, given Assumption A2, we have $E[Y_{t-1} I_{1t-d}] = E[I_{1t-d}] E[Y_{t-1}] = F(\gamma) E[Y_{t-1}]$ and $E[Y_{t-1} I_{2t-1}] = (1 - F(\gamma)) E[Y_{t-1}]$. With Y_t denoting a solution to the stochastic recurrence, we have $\forall t \; E[Y_t] = 0$ and

$$\begin{aligned}E[Y_t Y_t'] & = E[\mathbf{\Phi}_1 Y_{t-1} Y_{t-1}' \mathbf{\Phi}_1 I_{1t-1}] + E[\mathbf{\Phi}_2 Y_{t-1} Y_{t-1}' \mathbf{\Phi}_2 I_{2t-1}] + E[u_t u_t'] \\ & = F(\gamma) \mathbf{\Phi}_1 E[Y_{t-1} Y_{t-1}'] \mathbf{\Phi}_1 + (1 - F(\gamma)) \mathbf{\Phi}_2 E[Y_{t-1} Y_{t-1}'] \mathbf{\Phi}_2 + \Omega_u.\end{aligned}$$

Letting $V_t = E[Y_t Y_t']$, the above stochastic difference equation can be written more compactly as $V_t = F(\gamma)\mathbf{\Phi}_1 V_{t-1}\mathbf{\Phi}_1' + (1 - F(\gamma))\mathbf{\Phi}_2 V_{t-1}\mathbf{\Phi}_2' + \Omega_u$. Next, vectorizing both sides and letting $v_t \equiv vec(V_t)$ and $\omega \equiv vec(\Omega_u)$, we have $v_t = [F(\gamma)(\mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1) + (1 - F(\gamma))(\mathbf{\Phi}_2 \otimes \mathbf{\Phi}_2)]v_{t-1} + \omega$. For Y_t to be covariance stationary, it is thus necessary that V_t converges and this is ensured by the requirement that $\rho(F(\gamma)(\mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1) + (1 - F(\gamma))(\mathbf{\Phi}_2 \otimes \mathbf{\Phi}_2)) < 1$. Following the same line of proof as in Brandt (1986) and Karlsen (1990), it is also straightforward to establish that, if $\rho(F(\gamma)(\mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1) + (1 - F(\gamma))(\mathbf{\Phi}_2 \otimes \mathbf{\Phi}_2)) < 1$, then the above stochastic recurrence admits a unique covariance stationary solution. We can thus conclude that the above threshold VAR admits a unique covariance stationary solution *if and only if* $\rho(F(\gamma)(\mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1) + (1 - F(\gamma))(\mathbf{\Phi}_2 \otimes \mathbf{\Phi}_2)) < 1$.

Proof of Proposition 6 We first consider the case $r_j > r^0$ and establish that under the stated conditions, $P[\overline{IC}(r_j) < \overline{IC}(r^0)] \to 0$ as $T \to \infty$. From the definition of $\overline{IC}(.)$ in (15.18) we have $P[\overline{IC}(r_j) < \overline{IC}(r^0)] = P[-T\sum_{i=r^0+1}^{r_j} \ln(1 - \hat{\lambda}_i^j) > c_T(2pr_j - 2pr^0 - r_j^2 + (r^0)^2)]$. Since $-T\sum_{i=r^0+1}^{r_j} \ln(1 - \hat{\lambda}_i^j)$ is $O_p(1)$ and the right-hand side diverges towards infinity we have that $\lim_{T\to\infty} P[\overline{IC}(r_j) < \overline{IC}(r^0)] = 0$ and thus the procedure does not overrank asymptotically. For the case $r_j < r^0$ we have $P[\overline{IC}(r_j) < \overline{IC}(r^0)] = P[\sum_{i=r_j+1}^{r^0} \ln(1 - \hat{\lambda}_i^j) < \frac{c_T}{T}(2pr^0 - (r^0)^2 + r_j^2 - 2pr_j)]$. Since $-\sum_{i=r_j+1}^{r^0} \ln(1 - \hat{\lambda}_i^j) \xrightarrow{p} \theta > 0$ and $\frac{c_T}{T} \to 0$, it follows that, for $r_j < r^0$, $\lim_{T\to\infty} P[\overline{IC}(r_j) < \overline{IC}(r^0)] = 0$ as required.

References

Andrews, D.W.K. (1993) Tests for parameter stability and structural change with unknown change point. *Econometrica* 61, 821–56.

Arai, Y. (2004) Testing for linearity in regressions with I(1) processes. CIRJE Discussion Paper No. F-303, University of Tokyo.

Balke, N. and T. Fomby (1997) Threshold Cointegration. *International Economic Review* 38, 627–45.

Baum, C.F., J.T Barkoulas and M. Caglayan (2001) Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. *Journal of International Money and Finance* 20, 379–99.

Beaudry, P. and G. Koop (1993) Do recessions permanently change output? *Journal of Monetary Economics* 31, 149–64.

Bec, F., M. Ben-Salem and M. Carrasco (2001) Tests for unit-root versus threshold specification with an application to the PPP. Journal of Business and Economic Statistics, forthcoming.

Bec, F. and A. Rahbek (2004) Vector equilibrium correction models with nonlinear discontinuous adjustments. *Econometrics Journal* 7, 1–24.

Borenstein, S.A., A.C. Cameron and R. Gilbert (1997) Do gasoline prices respond asymmetrically to crude oil prices? *Quarterly Journal of Economics* 112, 305–39.

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–20.

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, 520–33.

Cragg, J.G. and S.G. Donald (1997) *Inferring the rank of a matrix. Journal of Econometrics* 76, 223–50.

Diebolt, J., N. Laib and J. Ngatchou-Wandji (1997) Limiting distributions of weighted processes of residuals. Application to parametric nonlinear autoregressive models. *Comptes Rendus de l'Académie des Sciences de Paris*, Series 1, 325, 535–40.

Durlauf, S.N. and P.A. Johnson (1995) Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics* 10, 365–84.

Enders, W. and B. Falk (1998) Threshold-autoregressive, median-unbiased, and cointegration tests of purchasing power parity. *International Journal of Forecasting* 14, 171–86.

Enders, W. and C.W.J. Granger (1998) Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business and Economic Statistics* 16, 304–11.

Enders, W. and P.L. Siklos (2001) Cointegration and threshold adjustment. *Journal of Business and Economic Statistics* 19, 166–76.

Engle, R.F. and C.W.J. Granger (1987) Cointegration and error correction: representation, estimation and testing. *Econometrica* 55, 251–76.

Gonzalez, M. and J. Gonzalo (1997) Threshold unit root processes. Unpublished manuscript, Department of Statistics and Econometrics, Universidad Carlos III de Madrid.

Gonzalo, J. and C.W.J. Granger (1995) Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics* 13, 27–35.

Gonzalo, J. and R. Montesinos (2000) Threshold stochastic unit root models. Unpublished manuscript, Department of Statistics and Econometrics, Universidad Carlos III de Madrid.

Gonzalo, J. and J-Y. Pitarakis (1998) Specification via model selection in vector error correction models. *Economics Letters*, 60, 321–8.

Gonzalo, J. and J-Y. Pitarakis (1999) Dimensionality effect in cointegration analysis. In R.F. Engle and H. White (eds). *Cointegration, Causality and Forecasting: a Festschrift in Honour of C.W.J. Granger*. Oxford: Oxford University Press, pp. 212–29.

Gonzalo, J. and J-Y. Pitarakis (2002) Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110, 319–52.

Gonzalo, J. and J-Y. Pitarakis (2005a) Threshold nonlinearities in cointegrated systems. Unpublished manuscript.

Gonzalo, J. and J-Y. Pitarakis (2005b) Estimation of common long-memory components in threshold cointegrated systems. Unpublished manuscript.

Granger, C.W.J., T. Inoue and N. Morin (1997) Nonlinear stochastic trends. *Journal of Econometrics* 81, 65–92.

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–30.

Hansen, B.E. (1997) Approximate asymptotic p-values for structural change tests. *Journal of Business and Economic Statistics* 15, 60–7.

Hansen, B.E. (1999a) Testing for linearity. *Journal of Economic Surveys* 13, 551–76.

Hansen, B.E. (1999b) Threshold effect in nondynamic panels: estimation, testing and inference. *Journal of Econometrics* 93, 345–68.

Hansen, B.E. and B. Seo (2002) Testing for two-regime threshold cointegration in vector error correction models. *Journal of Econometrics* 110, 293–318.

Hong, S. (2003) Testing linearity in cointegrating relations: application to PPP. Mimeo, Yale University, unpublished manuscript.

Johansen, S. (1988) Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control* 12, 231–54.

Johansen, S. (1991) Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Johansen, S. (1995) *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Karlsen, H.A. (1990) Existence of moments in a stationary stochastic difference equation. *Advances in Applied Probability* 22, 129–46.

Koop, G. and S.M. Potter (1999) Dynamic asymmetries in US unemployment. *Journal of Business and Economic Statistics* 17, 298–312.

Lo, M.C. and E. Zivot (2001) Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics* 5, 533–76.

Michael, P., A.R. Nobay and D.A. Peel (1997) Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation. *Journal of Political Economy* 105, 862–79.

Newey, W.K. and D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics*, vol. 4. New York: Elsevier, pp. 2113–2245.

Obstfeld, M. and A.M. Taylor (1997) Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. *Journal of the Japanese and International Economies* 11, 441–79.

O'Connell, P.G.J. and S. Wei (1997) The bigger they are the harder they fall: how price differences across US cities are arbitraged. *Journal of International Economics* 56, 21–53.

Peltzman, S. (2000) Prices rise faster than they fall. *Journal of Political Economy* 108, 466–502.

Phillips, P.C.B. and J.C. Chao (1999) Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91(2), 227–71.

Phillips, P.C.B. and S. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–95.

Pippenger, M.K. and G.E. Goering (1993) A note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics* 55, 473–81.

Potter, S.M. (1995) A nonlinear approach to US GNP. *Journal of Applied Econometrics* 2, 109–25.

Saikkonen, P. and I. Choi (2004) Cointegrating smooth transition regressions, *Econometric Theory* 20, 301–40.

Seo, M. (2004) Bootstrap testing for the presence of threshold cointegration in a threshold vector error correction model. Unpublished manuscript, Department of Economics, University of Wisconsin-Madison.

Tong, H. and K.S. Lim (1980) Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society* 4, Series B, 245–92.

Tong, H. (1983) Threshold models in non-linear time series analysis. *Lecture Notes in Statistics*, vol. 21. Berlin: Springer-Verlag.

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

Tsay, R.S. (1989) Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* 84, 231–40.

Tsay, R.S. (1998) Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188–1202.

Wohar, M.E. and N.S. Balke (1998) Nonlinear dynamics and covered interest rate parity. *Empirical Economics* 23, 535–59.