Partial Cointegration

Jaya Krishnakumar and David Neto

No 2005.04

Cahiers du département d'économétrie
Faculté des sciences économiques et sociales
Université de Genève

Août 2006

Département d'économétrie
Université de Genève, 40 Boulevard du Pont-d'Arve, CH -1211 Genève 4
http://www.unige.ch/ses/metri/

Partial Cointegration*

Jaya Krishnakumar David Neto
Department of Econometrics, University of Geneva†

August 2006

Abstract

This paper proposes a simple procedure to test the hypothesis of no cointegration against both threshold cointegration and an intermediate possibility that we call *partial cointegration*. Asymptotic theory is developed, the power of the proposed test is analysed through simulations and an empirical example is provided.

Keywords: Threshold cointegration, partial unit root, Wald test, term structure of interest rates.

JEL classification: C12, C32, E43

*Financial support from the Swiss National Fund for Scientific Research is gratefully acknowledged.

†Address: 40, Bd. du Pont d'Arve, CH-1211, Geneva 4, Switzerland. Email: jaya.krishnakumar@metri.unige.ch, david.neto@metri.unige.ch

1 Introduction

The traditional literature on cointegration and error correction models considers adjustment to long-run equilibrium as a continuous process. More recently, the joint analysis of nonlinearity and nonstationarity, introduced in economics by Balke and Fomby (1997), has been an area of intense research especially during the past decade. Indeed, there are many economic situations for which adjustment toward equilibrium is not continuous. For example, transaction costs in financial assets markets may lead to changing speeds of convergence to equilibrium of rates of return. This nonlinearity in the error correction mechanism can also be due to institutional constraints such as policy interventions in exchange rate or interest rates management in a given monetary context. Thus, central banks may manipulate two different interest rates so that the spread does not exceed a given fluctuation band. Similarly, for exchange rates, nonlinearities may arise in adjustment to equilibrium when monetary authorities use target zones device to apply an exchange rate policy. Indeed, such a target zone model was introduced by Krugman (1991) in which the long-run parity relationship is inactive within a given range of disequilibria while it becomes active when the system crosses the boundaries of allowed fluctuations.

These stylized facts have generated a wide interest in the use of specifications such as Threshold Autoregressive (TAR) models introduced by Tong (1983). An interesting application was proposed by Balke and Fomby (1997) who extend the definition of error correction representation provided by Engle and Granger (1987) to a threshold context.

Since the early paper of Balke and Fomby on threshold cointegration, most of the studies focus on asymmetric adjustment in the error correction mechanism described as a stationary system where the speed of adjustment differs according to the regime. For example, Enders and Siklos (2001) propose a testing procedure to detect an asymmetric adjustment effect in the equilibrium error, while Hansen and Seo (2002) and Gonzalo and Pitarakis (2006) provide statistics and asymptotic theory for testing the existence of a threshold but without testing for cointegration. More recently, Seo (2004) presents a complete theoretical investigation of the test of no cointegration against threshold cointegration in the particular case when the lagged disequilibrium error is the threshold variable. However, few papers focus on threshold cointegrated systems with unit root in one of the regimes as was initially considered by Balke and Fomby (1997). Following the terminology of Caner and Hansen (2001) who investigate the presence of unit root in a simple threshold autoregressive model,[1] we propose to call this case as one of *partial cointegration*. The reason for this lack could probably be traced back to Balke and Fomby (1997) itself where the authors mention the difficulties that *partial cointegration* gives rise to in its statistical inference, in particular when one wishes to test the no-cointegration hypothesis against threshold cointegration. Indeed, two pitfalls

[1] Caner and Hansen (2001) called such a TAR process, a *partial unit root* process.

are combined: in addition to the well known issue of the non-identification of the threshold parameter under the null hypothesis (*i.e.* linear adjustment) inherent in TAR models, the process is nonstationary under the null (corresponding to the no cointegration hypothesis) and locally nonstationary under the alternative of threshold cointegration with unit root in one regime. However, recent studies provide routes for testing unit root in a simple TAR model (see Caner and Hansen, 2001, Kapetanios and Shin, 2002) which may be used in threshold cointegration analysis. The aim of this paper is therefore to propose a testing procedure able to test no cointegration and at the same time discriminate under the alternative, between *partial cointegration* and cointegration with asymmetric adjustment in the error correction mechanism.

For this purpose, we consider a two-regime threshold long-run equilibrium model for which the transition variable is an appropriate stationary variable. We derive the error correction representation and consider the test for which the null hypothesis is no-cointegration against two alternatives. The first one corresponds to the traditional threshold cointegration hypothesis where the speed of adjustment differs according to the regime, and the other one covers the *partial cointegration* hypothesis. The difficulty of this test type is twofold (Hansen and Seo, 2002, Seo 2004, Gonzalo and Pitarakis, 2006). First, both the threshold parameter and the cointegrating vector are not identified under the null of no-threshold and no-cointegration. Secondly, as mentioned above, nonstationarity appears both in the null hypothesis and in the alternative when the we focus on the *partial cointegration* hypothesis.

Our contribution attempts to fill the gap in the existing literature on testing for threshold cointegration. In particular, it completes the investigation of Balke and Fomby (1997) by providing an asymptotic test theory and an empirical illustration. In addition to testing the no cointegration hypothesis, we will show that our testing procedure allows to discriminate *partial cointegration* from threshold cointegration.

The paper is organized as follows. In Section 2, threshold cointegration is briefly reviewed, partial cointegration is formally defined and error correction representation is derived for a two-regime model. Section 3 presents the test statistics and the testing procedure. Section 4 derives the asymptotic distributions both when the cointegrating vector is known and when it is estimated. Section 5 is devoted to Monte-Carlo experiments. In Section 6, we present an empirical application of the proposed procedure to the term structure of interest rates. The last section summarizes and concludes. Proofs of the theorems are presented in the appendix.

The following general notation is used. We denote by $D[0,1]$ the space of *cadlag* functions (*i.e.* the function which are left continuous and have right limits) on the unit interval $[0,1]$, and $D^n[0,1]$, the product space of n-copies of $D[0,1]$. A $(n \times n)$ identity matrix is denoted by I_n and the projection matrices by $P_A = A(A'A)^{-1}A'$. We will use $I(1)$ and $I(0)$ to signify time series that are integrated of order 1 and 0, respectively. Throughout the paper, the integrals are taken over the unit interval unless indicated otherwise and we use "$\Longrightarrow$" to denote the weak convergence of the associated probability measures as the sample size T tends to infinity.

2 Threshold cointegration

2.1 Asymmetric adjustment to long-run equilibrium

In order to introduce the threshold cointegration concept, we begin by studying the following n-dimensional cointegrated system, $y_t' = (y_{1t}, \ \ y_{2t}')$, with only one cointegrated relationship:

$$\begin{aligned} y_{1t} &= \beta' y_{2t} + z_t \\ \Delta y_{2t} &= \eta_t \end{aligned} \tag{1}$$

where y_{1t} is an univariate integrated process $I(1)$ and y_{2t}, a m-vector integrated process, and where $u_t = (z_t \ , \ \eta_t)'$ is a n-vector $(n = 1 + m)$ white-noise disturbance which satisfies the following assumption.

Assumption 1. *(i)* $E(u_t) = 0$ *for all* t; *(ii)* $\sup_t E\left(|u_{kt}|^\delta\right) < \infty$ *for* $\delta > 2$ *and* $k = 1, ..., n$; *(iii)* $E(u_t u_t') = \Psi > 0$; *(iv)* u_{kt}, *are strongly mixing, with mixing coefficients* $\{\alpha_{ki}\}$ *such that* $\sum_{i=1}^{\infty} \alpha_{ki}^{1-2/\delta} < \infty$, *for* $k = k = 1, ..., n$.

The first assumption ensures that all drawings of u_t have the same mean. The second one is a sufficient condition to ensure the existence of the variance and a higher non-integer moment of u_{kt}, $\forall t$. The third, is a convergence condition on the average variance of the partial sum $S_T = \sum_{\tau}^{t} u_{k\tau}$, and the last assumption controls the extent of the temporal dependence in the processes u_{kt} (See Phillips (1987) for details).

Phillips (1987) has shown that under conditions (i) to (iv), the suitably normalised process $B_T(r) = T^{-1/2} S_{[Tr]}$ (where $[x]$ denotes the largest integer which does not exceed x), such that $B_T \in D^n[0,1]$, converges weakly to a bivariate Brownian motion B with covariance matrix $\Psi : B_T(r) \Longrightarrow B(r)$. The continuous mapping theorem (Billingsley, 1968) states that if $B_T(r) \Longrightarrow B(r)$ and g is any continuous function on $D[0,1]$, then $g(B_T(r)) \Longrightarrow g(B(r))$.

The system implies that y_{2t} is not cointegrated and the first equation is a single cointegrating regression with cointegrating vectors $\alpha' = (1, -\beta')$. This is the so-called triangular representation (Phillips, 1991). Taking first differences in (1) and rearranging, the system has the following VECM representation:

$$\Delta y_t = \gamma \alpha' y_{t-1} + v_t \tag{2}$$

where $y_t' = (y_{1t}, \ y_{2t})$, $\gamma' = (-1, 0)$, and $v_t = \begin{pmatrix} 1 & \beta' \\ 0 & I_m \end{pmatrix} u_t$.

While traditional cointegration theory assumes linearity and symmetry in adjustment to long-run equilibrium, threshold cointegration describes the disequilibrium error process z_t as a Threshold Autoregressive (TAR) process such that

$$z_t = \begin{cases} \phi_1 z_{t-1} + \epsilon_t \text{ if } s_{t-d} \leq \theta \\ \phi_2 z_{t-1} + \epsilon_t \text{ if } s_{t-d} > \theta \end{cases} \tag{3}$$

where ϵ_t is a white-noise disturbance of variance σ^2, and s_{t-d} is called the transition variable where $d \geq 0$ is the threshold lag or delay, and θ is a threshold, which corresponds to the long-run equilibrium value. The transition variable s_t could be an exogenous variable or could even be the the lagged dependent variable, for instance the disequilibrium term z_{t-1} (Balke and Fomby, 1997, Enders and Siklos, 2001). This latter particular case is called Self-Exciting Threshold Autoregressive (SETAR).

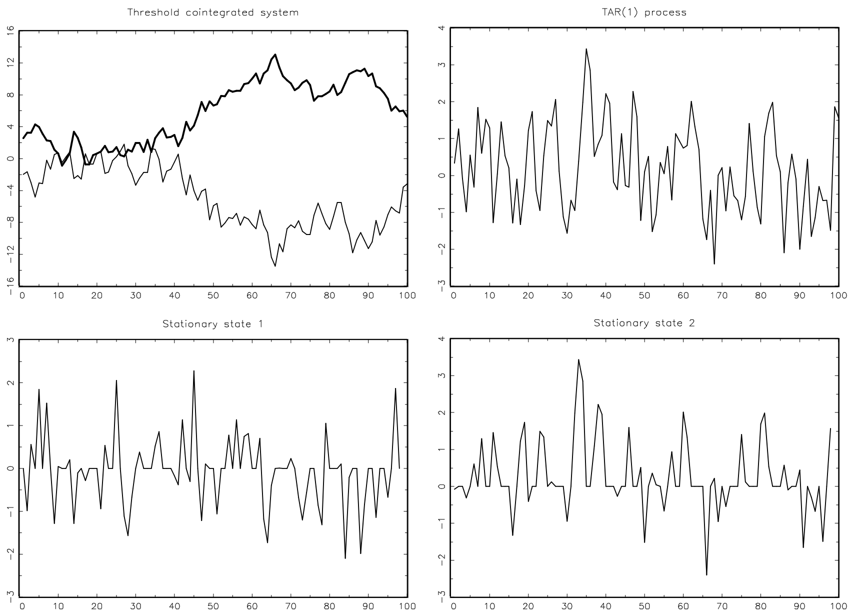
Another perhaps more convenient choice for the transition variable is Δs_t or more generally $\Delta^d s_t$ because it ensures the stationarity of the transition variable (See Enders and Siklos, 2001, and Caner and Hansen, 2001). When this transition variable corresponds to the lagged dependent variable, Δz_{t-1}, such a TAR model is called a Momentum-TAR (M-TAR).

The basic TAR model was introduced by Tong (1983) for which inference was widely studied by Tsay (1989, 1998), Chan (1993) and Hansen (1996, 1997) among others. Chan *et al.* (1985) have shown that this two-regime TAR process is stationary and ergodic if $\phi_1 < 1$, $\phi_2 < 1$ and $\phi_1\phi_2 < 1$. Under these conditions and according to the definition of cointegration, y_{1t} and y_{2t} are cointegrated with cointegrating vector α. The equilibrium error z_t is said to be mean reverting. However, unlike conventional cointegrated systems considered by Engle and Granger (1987) where the adjustment is symmetric (*i.e.* $\phi_1 = \phi_2$), here the movement towards long-run equilibrium is asymmetric in that the speeds of mean reversion?, ϕ_1 and ϕ_2, differ according to the value of s_{t-d}. Thus the Engle-Granger definition appears as a special case of threshold cointegration. Note that when the special case $\phi_1 = \phi_2 = 1$ occurs, then z_t is $I(1)$ and y_{1t} and y_{2t} are not cointegrated.

Figure 1 displays an example of threshold cointegrated system as defined by (1) and (3), where $s_{t-d} = \Delta z_{t-1}$ and where the error-correction term is an asymmetric mean reverting process $\phi_1 \neq \phi_2 \neq 1$ (we use this 'short-cut' notation to both are different and different from 1). Both regimes are stationary but when Δz_{t-1} exceeds the threshold (taken to be zero here), the persistence of z_t in the second regime is stronger: the parameter values used for simulation are $\phi_1 = 0.3$, $\phi_2 = 0.6$, and $\beta = -0.9$.

The formulation (1) and (3) describing a threshold cointegrated system is now commonly used in empirical and theoretical studies to take into account nonlinear adjustments (Enders and Siklos, 2001, Lo and Zivot, 2001, among others). In the above example, we have only considered two regimes in the

Figure 1: Threshold cointegrated system



TAR model, but multi-band models could also be developed. See for example Balke and Fomby (1997), or more recently Kapetanios and Shin (2003) that consider TAR models with three regimes.

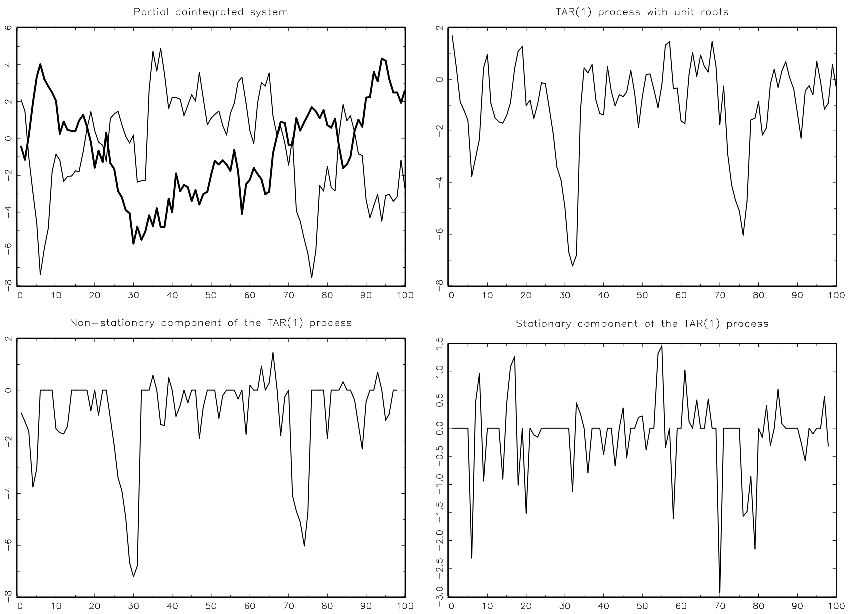
2.2 Partial cointegration

A growing interest in the recent literature on TAR processes has focused on the presence of a unit root in one of the regimes. Indeed, some macroeconomic variables may behave in a non-stationary manner only in a particular context. This local nonstationary feature occurs when $\phi_1 = 1$ and $\phi_2 < 1$ or $\phi_1 < 1$ and $\phi_2 = 1$ in eq. (3) such that the TAR behaves like a unit root process in one regime.[2] In the terminology of Caner and Hansen (2001), this intermediate case is called *partial unit root*. By analogy with this terminology, we call the same configuration in threshold cointegration, *partial cointegration*. Figure 2 illustrates such a cointegrated system where the non-stationary state is only active under the first regime. The parameter values used for this simulated system are $\phi_1 = 1$, $\phi_2 = 0.3$, and $\beta = -0.9$.

In spite of the lack of literature in this area, one should note that when Balke and Fomby (1997) published their work, it was this case of *partial cointegration* that they initially considered. However, emphasizing the difficulties in testing *partial cointegration*, the authors did not provide either a formal test theory or any empirical illustration.

[2]Note that, in spite of the local nonstationarity of the process, it remains globally ergodic (See Tweedie, 1975 and Chan *et al.*, 1985)

Figure 2: Partial cointegrated system



The purpose of this paper therefore is to complete the Balke and Fomby investigations exploiting some of the results from the literature on TAR processes with unit root. In particular we extend the works of Caner and Hansen (2001) to develop a testing procedure to detect the presence of a partial unit root in the long-run disequilibrium term. Since our procedure makes use of the error correction representation of a threshold cointegration system, we briefly recall its formulation in the following subsection.

2.3 Error correction representation

To derive the Threshold Vector Error Correction Model (TVECM henceforth) associated with the data generating process (DGP) of Section 2, it is convenient to rewrite (3) as follows. Using the notations $z_t^{(1)} = z_t \mathbb{I}_{\{s_{t-d} \leq \theta\}}$ and $z_{t-1}^{(2)} =$

$z_t \mathbb{I}_{\{s_{t-d} > \theta\}}$, where $\mathbb{I}_{\{.\}}$ denotes the indicator function, and by noting that $z_t^{(1)} + z_t^{(2)} = z_t$, we can write the TAR(1) process as:

$$\Phi_1 (L) z_t^{(1)} + \Phi_2 (L) z_t^{(2)} = \epsilon_t \tag{4}$$

where $\Phi_j (L) = 1 - \phi_j L$, for $j = 1, 2$.

Proposition 1. *The Threshold VECM associated with the system described by* (1) *and* (2) *with* η_t denoting a *q-order autocorrelated process, is given by*

$$\Delta y_t = \pi^{(1)} y_{t-1}^{(1)} + \pi^{(2)} y_{t-1}^{(2)} + \Gamma (L) \Delta y_{t-1} + \varepsilon_t \tag{5}$$

where $y_{t-1}^{(1)} = y_{t-1}\mathbb{1}_{\{s_{t-d} \leq \theta\}}$, $y_{t-1}^{(2)} = y_{t-1}\mathbb{1}_{\{s_{t-d} > \theta\}}$,
$\pi^{(j)} = \gamma_j \alpha'$, ε_t *with covariance matrix* Σ
$\gamma_j = \begin{pmatrix} -\Phi_j(1) \\ 0 \end{pmatrix}$, $\Gamma(L) = \begin{pmatrix} 0 & \beta' g(L) \\ 0 & g(L) \end{pmatrix}$ *with*
$g(L) = \sum_{i=1}^{q} g_i \, L^{i-1}$, g_i *being* $(m \times m)$ *coefficient matrices* .

Proof. *See Appendix A.1.* ■

The threshold effect is active only if $0 < \mathbb{P}(s_{t-d} \leq \theta) < 1$, otherwise the model is reduced to the standard VECM, *i.e.* $\pi^{(1)} = \pi^{(2)}$. Indeed, if there is no threshold effect, the disequilibrium in (1) is an autoregressive process of order 1 and the earlier VECM becomes

$$\Delta y_t = \pi y_{t-1} + \Gamma(L) \Delta y_{t-1} + \varepsilon_t$$

¿From a TVECM, Seo (2004) proposes a test of no-cointegration whose alternative is the threshold cointegration, while Hansen and Seo (2002) provide a test of linear cointegration hypothesis against the threshold cointegration hypothesis conditioning on the existence of cointegration. The main difficulty of these tests is that under the null hypothesis there is no threshold effect and so the threshold parameter is not identified. Moreover, the issue considered by Seo is more complicated because there are added unidentified parameters under the null namely those of the cointegrating vector. Under the null of no-cointegration, *i.e.* $\phi_1 = \phi_2 = 1$, so $\gamma_1 = \gamma_2 = 0$, the disequilibrium in (1) is an $I(1)$ process, so the TVECM of the proposition 1 becomes a simple VAR process.

However, these tests do not consider the alternative that (3) has a unit root. As we have seen in the previous section, the hypothesis of partial cointegration in (3) can be written as:

$$H_0 : \phi_1 = \phi_2 = 1 \quad \text{vs.} \quad H_1 : \begin{cases} \phi_1 = 1 \\ \phi_2 < 1 \end{cases} \quad \text{or} \quad \begin{matrix} \phi_1 < 1 \\ \phi_2 = 1 \end{matrix}$$

Now, consider the TVECM specification (5). It is straightforward to show that the previous alternative hypothesis is equivalent to the following:

$$H_a : \begin{cases} \gamma_1 = 0 \\ \gamma_2 \neq 0 \end{cases} \quad \text{or} \quad \begin{matrix} \gamma_1 \neq 0 \\ \gamma_2 = 0 \end{matrix}$$

Under H_a, the process described by the TVECM is locally nonstationary. In this context, $\pi^{(1)} = 0$ in our TVECM:

$$\Delta y_t = \pi^{(2)} y_{t-1}^{(2)} + \Gamma(L) \Delta y_{t-1} + \varepsilon_t$$

While the test proposed by Seo (2004) could be used to test the null of no-cointegration against the alternatives $\gamma_1 \neq \gamma_2 \neq 0$ *and* H_a above, it is unable to discriminate between these two hypotheses. Moreover, Caner and Hansen (2001) point out the difficulty of testing H_0 against H_1 based on a TAR model because of the one-sided feature of H_1. They suggest examining the significance of the individual estimators $\hat{\phi}_1$ and $\hat{\phi}_2$ considering negative values of the t-statistics to improve the power of the test. In cointegration analysis, it is convenient to conduct this test using the TVECM representation because the alternative becomes two-sided. Hence, we will derive the Wald test from the TVECM which is easier to perform than the one-sided test in model (3). Considering the individual Wald statistics of both vector estimates $\hat{\gamma}_1$ and $\hat{\gamma}_2$, if only one is significant, then the test would be consistent with the hypothesis H_1 but not with the alternative $\gamma_1 \neq \gamma_2 \neq 0$.

We now add some assumptions on the transition variable required for the inference in the generalised context.

Assumption 2. (i) *The transition variable s_{t-d} is a strictly stationary and ergodic sequence that is independent of ε_{kt}, for $k = 1, ..., n$, and whose distribution F is continuous everywhere;* (ii) *the threshold θ is such that $\theta \in \Theta = \left[\,\underline{\theta}\,, \overline{\theta}\,\right]$, a closed and bounded subset of the sample space of the transition variable s_{t-d}.*

The support for the threshold parameter is required to be trimmed such that $\mathbb{P}\left(s_{t-d} \leq \underline{\theta}\right) = \lambda > 0$ and $\mathbb{P}\left(s_{t-d} \leq \overline{\theta}\right) = 1 - \lambda$. In general, empirical investigations choose λ to be 15 percent. Notice that the choice of this percentage is somewhat arbitrary. It is chosen in order to ensure enough observations in each regime and that the limits for the test statistics used in inference are nondegenerate (See Hansen, 1996). Let $U_t = F\left(s_{t-d}\right) \sim U_{[0,1]}$. Due to the equality $\mathbb{1}_{\{s_{t-d} \leq \theta\}} = \mathbb{1}_{\{F(s_{t-d}) \leq F(\theta)\}}$, we will use $u = F(\theta) \in [\,\lambda\,, 1 - \lambda\,]$ (See Caner and Hansen, 2001).

Let "$\Longrightarrow$" denote weak convergence with respect to the uniform metric on $[0,1]^2$, under Assumptions 1 and 2 we have the following Lemma from Caner and Hansen (2001), that will serve for the subsequent results developed in the paper.

Lemma 1. *Let Assumption 2 hold and consider the earlier integrated n-vector process y_t such as $\Delta y_t = e_t$, where e_t satisfies Assumption 1 and whose covariance is Ω. Let $\Omega^{1/2}$ be the matrix $\mathbf{P}$ such that $\Omega = \mathbf{PP}'$. Then, as $T \longrightarrow \infty$ we have:*

(i) $B_T\left(r, u\right) = T^{-1/2} \sum_{t=1}^{[Tr]} \mathbb{1}_{\{U_t \leq u\}} e_t \Longrightarrow \Omega^{1/2} B(r, u)$ *on* $(r, u) \in [0,1]^2$, with $B\left(r, u\right)$ *a n-vector standard two-parameter Brownian motion, such that*
$B(r, u) \sim N\left(0, ru\right), B\left(r, 1\right) = B\left(r\right).$
Symmetrically we have $T^{-1/2} \sum_{t=1}^{[Tr]} \mathbb{1}_{\{U_t > u\}} e_t \Longrightarrow B\left(r, 1 - u\right).$

(ii) $T^{-1}\sum_{t=1}^{T}\mathbb{1}_{\{U_t\leq u\}}y_{t-1}e_t'\Longrightarrow\Omega^{1/2}\int B(r)\,dB(r,u)'$

(iii) $T^{-2}\sum_{t=1}^{T}\mathbb{1}_{\{U_t\leq u\}}z_{t-1}z_{t-1}'\Longrightarrow F(\theta)\int V(r)\,V(r)'\,dr$

(iv) $T^{-2}\sum_{t=1}^{T}\mathbb{1}_{\{U_t>u\}}z_{t-1}z_{t-1}'\Longrightarrow S(\theta)\int V(r)\,V(r)'\,dr$

where $z_t=\alpha' y_t$ with α is a nonrandom matrix and $S(\theta)=1-F(\theta)$.

Proof. *See Caner and Hansen (2001, Theorems 1-3, pp. 1560-1561).* ■

3 Testing no-cointegration

Testing no-cointegration against partial cointegration

In this section we develop a test that can discriminate against the alternative of partial cointegration where the equlibrium error follows a TAR process with a unit root in one regime.

Letting $Var(\varepsilon_t)=\Sigma$ and $z_t^{(j)}=\alpha' y_t^{(j)}$ in the n-dimensional TVECM (5), we write

$$\Delta y_t=\gamma_1 z_{t-1}^{(1)}+\gamma_2 z_{t-1}^{(2)}+\Gamma(L)\,\Delta y_{t-1}+\varepsilon_t \tag{6}$$

where γ_j is a n-vector. In order to consider the test of no-cointegration with the alternative of partial cointegration H_a, it is convenient to vectorize the formulation of the previous model stacking the observations in order to obtain:

$$vec(\Delta Y)=\gamma_1\otimes Z_1+\gamma_2\otimes Z_2+\sum_{i=1}^{q}(\Gamma_i\otimes I_T)\,vec(\Delta Y_{-i})+\xi$$

where ΔY is a $(T\times n)$ matrix, $\xi=vec(\varepsilon)$ with ε is the $(T\times n)$ matrix of stacked errors, Z_j is a T-vector, Γ_i are $(n\times n)$ matrices, and $\Delta Y_{-i}=[\Delta Y_{1,-i}\quad \Delta Y_{n,-i}]$, where $\Delta Y_{k,-i}$ is the matrix $(T\times n)$ which stacks the observations of the kth component of Δy_{t-i}, for $k=1,...,n$ and $i=1,...q$. Using partitioned regression, the Wald statistic under the null is given by

$$\mathcal{W}_{T,j}=\hat\gamma_j(u)'\,Var(\hat\gamma_j(u))^{-1}\,\hat\gamma_j(u)$$

$$\mathcal{W}_{T,j}=\left[\left(I_n\otimes\left(Z_j'M_XZ_j\right)^{-1}Z_j'M_X\right)\xi\right]'\left[\hat\Sigma^{-1}\otimes\left(Z_j'M_XZ_j\right)\right]\left[\left(I_n\otimes\left(Z_j'M_XZ_j\right)^{-1}Z_j'M_X\right)\xi\right]$$

$$=vec\left(\left[\left(Z_j'M_XZ_j\right)^{-1}Z_j'M_X\right]\varepsilon\right)'\left[\hat\Sigma^{-1}\otimes\left(Z_j'M_XZ_j\right)\right]vec\left(\left[\left(Z_j'M_XZ_j\right)^{-1}Z_j'M_X\right]\varepsilon\right)$$

for $j=1,2$, where $M_X=I_T-X\left(X'X\right)^{-1}X'$ is the idempotent matrix associated with the projection onto the orthogonal space of the sub-space generated by $X=\left[\mathbf{R}\;\; Z_{j'}\right]$ of order $(T\times qn+1)$ with $j\neq j'$, and where

$\mathbf{R} = [\Delta Y_{-1} \cdots \Delta Y_{-q}]$ a $(T \times qn)$ matrix of stationary elements of X under the null hypothesis with $\hat{\Sigma}$ a consistent estimator of Σ. Using the result that, for suitable matrices A, B and C,

$$tr\left\{AB'CB\right\} = vec\left(B'\right)'\left(C' \otimes A\right) vec\left(B'\right) = vec\left(B\right)'\left(A \otimes C'\right) vec\left(B\right)$$

we have:

$$\mathcal{W}_{T,j}\left(u\right) = tr\left\{\hat{\Sigma}^{-1}\left(Z_j' M_X \varepsilon\right)'\left(Z_j' M_X Z_j\right)^{-1}\left(Z_j' M_X \varepsilon\right)\right\}$$

Since the threshold parameter is not identified under the null, we consider the supremum Wald statistic (denoted sup-Wald) over a grid set of possible values of threshold to eliminate this nuisance parameter. For stationary data, this statistic was investigated by Davies (1987), Andrews and Ploberger (1994) and Hansen (1996), while sup-Wald was discussed by Caner and Hansen (2001) and Seo (2004) for integrated data. Thus, the appropriate statistic is:

$$\sup \mathcal{W}_j = \sup_{u \in [\lambda, 1-\lambda]} \mathcal{W}_{T,j}\left(u\right), \quad \text{for } j = 1, 2$$

Testing no-cointegration against threshold cointegration

The previous section proposed a new test but this section re-examines a test that already exists in the literature *viz.*, the test for no-cointegration against threshold cointegration (without unit root) proposed by Seo (2004) and generalises it for any transition variable s_{t-d} which satisfies Assumption 2. Seo considers the case in which the transition variable is the lagged disequilibrium error process, z_{t-1}, and proposes a Wald statistic to test no-cointegration based on a TVECM similar to our (5), deriving its asymptotic distribution, while Kapetanios and Shin (2003) investigate the same test in a TAR process with three regimes. In both cases the alternative is global stationarity.

Keeping the same notations and stacking the observations, the model (6) could be written compactly as

$$\begin{aligned} vec\left(\Delta Y\right) &= \mathcal{Z}\, vec\left(\gamma\right) + \underline{\Delta Y}\ \Gamma + \xi \\ Var\left(\xi\right) &= \Sigma \otimes I_T \end{aligned}$$

where $\mathcal{Z} = I_n \otimes Z$ is a $(nT \times 2n)$, with $Z = [Z_1\ Z_2]$, $\gamma = [\gamma_1\ \ \gamma_2]$ is a $2n$-vector, $\underline{\Delta Y} = I_n \otimes \mathbf{R}$, and Γ the qn^2-vector of the coefficients of the lags.

Using partitioned regression, under $H_0 : vec\left(\gamma\right) = 0$ as before, the Wald statistic is given by:

$$\mathcal{W}_{T,0}\left(u\right) = \xi'\left(I_n \otimes \left(Z' M_R Z\right)^{-1} Z' M_R\right)'\left[\hat{\Sigma}^{-1} \otimes \left(Z' M_R Z\right)\right]\left(I_n \otimes \left(Z' M_R Z\right)^{-1} Z' M_R\right)\xi$$

Once again we rewrite this statistic as:

$$\mathcal{W}_{T,0}(u) = tr\left\{\hat{\Sigma}^{-1}\left(Z'M_R\varepsilon\right)'\left(Z'M_RZ\right)^{-1}\left(Z'M_R\varepsilon\right)\right\}$$

where M_R denotes the idempotent matrix associated with the projection onto the orthogonal space of the sub-space generated by R. Thus the test statistic will be $\sup\mathcal{W}_0 = \sup_{u\in[\lambda,1-\lambda]} \mathcal{W}_{T,0}(u)$.

In the next section, we derive the large sample distribution of our statistics under the null, taking into account that the threshold is unidentified under the null (i.e. $\gamma_1 = \gamma_2 = 0$).

4 Asymptotic distributions

In this section, asymptotic distributions of statistics are derived. As traditionally, two cases are considered: when the cointegrated vector is known and when it is unknown *a priori*, hence estimated.

Economic theory often provides equilibrium relationships that the econometrician would like to verify in practice. This is typically the case for the hypothesis of purchasing power parity where the long-run equilibrium relation is fully known. Therefore, we begin by assuming β known to derive the asymptotic distributions. The following theorem gives the asymptotic distributions for $\sup\mathcal{W}_0$, $\sup\mathcal{W}_1$ and $\sup\mathcal{W}_2$.

Theorem 1. *Let Assumptions 1 and 2 hold. Then, under H_0, the distribution of $\sup\mathcal{W}_0$, $\sup\mathcal{W}_1$ and $\sup\mathcal{W}_2$ as $T\to\infty$ are:*

$$\sup\mathcal{W}_j \Longrightarrow \sup_{u\in[\lambda,1-\lambda]} T_j(u) \quad , \text{for } j = 0,1,2$$

where $T_0(u)$, $T_1(u)$ and $T_2(u)$ are defined as

$$T_0(u) = tr\left\{\mathbf{Q}(r,u)'\left(\mathbf{M}(u)\textstyle\int V^2(r)\,dr\right)^{-1}\mathbf{Q}(r,u)\right\}$$

$$T_1(u) = tr\left\{\textstyle\int dB(r,u)\,V(r)'\left(u\textstyle\int V^2(r)\,dr\right)^{-1}\textstyle\int V(r)\,dB(r,u)'\right\}$$

$$T_2(u) = tr\left\{\textstyle\int dB(r,1-u)\,V(r)'\left((1-u)\textstyle\int V^2(r)\,dr\right)^{-1}\textstyle\int V(r)\,dB(r,1-u)'\right\}$$

where $\mathbf{Q}(r,u) = \begin{pmatrix} \int V(r)\,dB(r,u)' \\ \int V(r)\,dB(r,1-u)' \end{pmatrix}$, $\mathbf{M}(u) = \begin{pmatrix} u & 0 \\ 0 & 1-u \end{pmatrix}$, $V(r)$ *is a univariate Brownian motion, $B(r,u)$ and $B(r,1-u)$ are standard two-parameter Brownian motions on $(r,F(\theta))$ and $(r,S(\theta))$ respectively.*

Proof. *See Appendix A.2.* ■

In many cases the cointegrating vector is unknown. Since these parameters are only identified under the alternative hypothesis, the cointegrated vector is a nuisance parameter, and so β must be estimated. Exploiting the super-consistency of the OLS estimator in the long-run equilibrium relationship, we conduct a two-step method like the Engle-Granger procedure (1987) to estimate the TVECM. Let now $\hat{\beta}_T = (y_2' y_2)^{-1} y_2' y_1$ be the least squares estimator of β based on a sample of T observations. The results of Phillips and Ouliaris (1990) allow us to derive the asymptotic distribution of the sup-Wald statistics given $\hat{\alpha}_T' = (1, \ -\hat{\beta}_T')$.

Theorem 2. *Let Assumption 1 and 2 hold. Under H_0 and for a consistent estimator $\hat{\alpha}_T$, the asymptotic distribution of our test statistics* $\sup\mathcal{W}_0$*,* $\sup\mathcal{W}_1$ *and* $\sup\mathcal{W}_2$ *are the same as in Theorem 1 where the Brownian motion $V(r)$ should be replaced by:*

$$V(r) = U_1(r) - \int U_1(r) U_m(r)' dr \left(\int U_m(r) U_m(r)' dr \right)^{-1} U_m(r)$$

with the following partition of the n-vector Brownian motion $U(r) = \left(U_1(r), \ \ U_m(r)' \right)'$ *with $U_m(r)$ is m-dimensional.*

Proof. *See Appendix A.3.* ■

Having derived the asymptotic distributions, we go on to tabulate them by Monte Carlo simulations. The stochastic integrals are evaluated at 10,000 points over the argument r and 100 steps over the argument u. The critical values are computed as the empirical quantiles from 100,000 replications and are reported in Table 1 for the bivariate case ($n = 2$). Note that the critical values for $\sup \mathcal{W}_1$ and $\sup \mathcal{W}_2$ are the same due to the symmetry of both statistics. We report those of $\sup \mathcal{W}_1$ only for various ranges $[\lambda, 1-\lambda]$ because the choice of λ affects the power of the tests. Caner and Hansen (2001) discuss the inconsistency of the tests for $\lambda = 0$. Indeed, the critical values of the statistics increase as λ decreases. This implies that the rejection of null hypothesis requires a larger value of the statistics as λ tends to 0. It follows that λ should be set in the interior of $(0, 1)$. However, as discussed by Andrews (1993) and Caner and Hansen (2001) the choice remains somewhat arbitrary and so various values must be tried in empirical applications to verify the robustness of the results to the selected value of λ.

So far, we assumed that the DGP had zero mean in order to arrive at the results without complicating notations. However it is not a very realistic assumption in practice where it is reasonable to include an intercept in the TVECM as follows:

$$\Delta y_t = \mu + \pi^{(1)} y_{t-1}^{(1)} + \pi^{(2)} y_{t-1}^{(2)} + \Gamma(L) \Delta y_{t-1} + \varepsilon_t$$

While the presence of an intercept in the model does not cause trouble for the estimation procedure, the asymptotic distributions of the estimators of long-run parameters and of the test statistics need minor analytical transformations. The Brownian motions V in Theorem 1 and 2 should be replaced by the corresponding demeaned Brownian motion defined as $V^* = V - \int V dr$. For Theorem 2, we will have $V = U_1 - \int U_1^* U_m^{*\prime} dr \left(\int U_m^* U_m^{*\prime} dr \right)^{-1} U_m$, where $U^* = (U_1^*, \quad U_m^{*\prime})' = \left(U_1 - \int U_1, \quad (U_m - \int U_m)' \right)'$. This latter modification comes from the presence of a drift in the estimation of the long-run relationship by OLS. The critical values for demeaned case are also reported in Table 1.

Table 1: Critical Values for Wald Statistics for Bivariate Models (*i.e.* $n = 2$)

	$[\lambda\ ,\ 1-\lambda]$	90%	95%	99%	90%	95%	99%
			$(\mu = 0)$			$(\mu \neq 0)$	
	[0.15 , 0.85]	10.666	12.631	16.821	13.008	15.106	19.745
$\sup \mathcal{W}_1$, $\sup \mathcal{W}_2$	[0.10 , 0.90]	10.983	12.911	17.120	13.328	15.415	20.018
	[0.05 , 0.95]	11.327	13.211	17.365	13.647	15.702	20.295
	[0.15 , 0.85]	14.947	17.117	21.620	17.735	20.104	24.998
$\sup \mathcal{W}_0$	[0.10 , 0.90]	15.374	17.501	22.059	18.137	20.488	25.409
	[0.05 , 0.95]	15.855	18.011	22.554	18.615	20.981	26.046

Note : calculated from 100,000 simulations.

5 Finite sample size and power

In this section we perform Monte-Carlo experiments to evaluate the finite-sample properties of the proposed tests, when the threshold is unidentified. For this purpose, we simulate the bivariate TVECM according to the DGP (1), setting $\gamma_1 = \gamma_2 = 0$ and assuming an AR(1) process for η_t in (1) i.e.

$$\Delta y_t = \Gamma \Delta y_{t-1} + \varepsilon_t$$

where $\Gamma = \begin{pmatrix} 0 & 0 \\ 0 & \Gamma_{22} \end{pmatrix}$ with Γ_{22} taking values from $\{-0.5, 0, 0.5\}$. We generate 100 independent bivariate observations of $\varepsilon_t' = (\varepsilon_{1t}, \quad \varepsilon_{2t})$ according to *iid* $N(0, I_2)$.

We first examine the size based on asymptotic distributions. Table 2 reports rejection frequencies of the null hypothesis at 5% level from 5000 Monte-Carlo replications for the tests $\sup \mathcal{W}_1$, $\sup \mathcal{W}_2$, and $\sup \mathcal{W}_0$. It shows that the asymptotic tests are strongly biased in small samples for all the three tests. Note that for the test based on $\sup \mathcal{W}_1$ and $\sup \mathcal{W}_2$, the null is rejected if either one, $\sup \mathcal{W}_1$ or $\sup \mathcal{W}_2$, rejects the null.

Let us now turn to the bootstrap approximation. Caner and Hansen (2001) discuss it in the context of TAR inference and find that the finite sample inference is improved when the bootstrap-approximated distribution is used. We follow the same procedure to construct our bootstrap critical values and p-values. Recall that under the null $\gamma_1 = \gamma_2 = 0$ in the TVECM(1). Let $\left(\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\Gamma}, \tilde{F}\right)$ be estimates of parameters $(\gamma_1, \gamma_2, \Gamma, F(\varepsilon_t))$ where $\tilde{F}$ is the empirical distribution of the residual $\{\hat{\varepsilon}_t\}$. Next, we generate $\Delta y_t^b = \tilde{\Gamma}\Delta y_{t-1}^b + \varepsilon_t^b$ from $\tilde{F}$ (*i.e.* ε_t^b is a drawing from $\tilde{F}$). Compute $\sup \mathcal{W}_j^b$ and $\sup \mathcal{W}_0^b$. The estimated p-values are the percentage of simulated $\sup \mathcal{W}_j^b$ and $\sup \mathcal{W}_0^b$ that exceed the observed $\sup \mathcal{W}_j$ and $\sup \mathcal{W}_0$ respectively. Our bootstrap 5% critical values for $T = 100$, are 19.674 for $\sup \mathcal{W}_j$ and 28.790 for $\sup \mathcal{W}_0$.[3] Based on these values, we repeat the Monte-Carlo experiments to obtain the size of our no-cointegration tests. The results are given in Table 3. It also reports the rejection rates for Johansen's trace test (1995). The results show that, while the statistic $\sup \mathcal{W}_0$ dominates, both joint tests work better than the procedure based on $\sup \mathcal{W}_1$ and $\sup \mathcal{W}_2$.

Table 2: Size of Asymptotic Tests

	$\sup \mathcal{W}_1$, $\sup \mathcal{W}_2$	$\sup \mathcal{W}_0$
Γ_{22}		
-0.5	0.300	0.247
0.0	0.297	0.243
0.5	0.293	0.238

Note : $T = 100$. Nominal size 5%. Rejection rates from 5000 replications.

Table 3: Size of Bootstrap No-Cointegration Tests

	$\sup \mathcal{W}_1$, $\sup \mathcal{W}_2$	$\sup \mathcal{W}_0$	λ_{tr}
Γ_{22}			
-0.5	0.083	0.040	0.055
0.0	0.082	0.037	0.057
0.5	0.078	0.033	0.058

Note : Nominal size 5%. Rejection rates from 5000 replications.
λ_{tr} denotes the trace statistic of Johansen.

It is interesting to analyze the power of the proposed testing procedures when the DGP is a partial cointegrated system. For this purpose, 1000 partially cointegrated processes were generated according the following values of parameters: $\beta = -0.9$, $\phi_1 = 1$, $u = 0.5$, Γ_{22} once again taking values from $\{-0.5, 0, 0.5\}$ and ϕ_2 from $\{0.2, 0.5, 0.7, 0.9\}$. The associated TVECM is

$$\Delta y_t = \gamma_2 z_{t-1}^{(2)} + \Gamma \Delta y_{t-1} + \varepsilon_t$$

[3]These values are obtained form 5000 bootstrap replications with bound $\lambda = 0.15$.

Note that in this case we have $\gamma_2 = \begin{pmatrix} \phi_2 - 1 \\ 0 \end{pmatrix}$ and $\Gamma = \begin{pmatrix} 0 & \beta\Gamma_{22} \\ 0 & \Gamma_{22} \end{pmatrix}$. The percentage of times that the null hypothesis is correctly rejected defines the power of the tests. The results of experiments (based on finite-sample critical values), with estimated cointegrated parameter β, are reported in Table 4. Our experiments show that the Wald statistics do a much better job than Johansen's test in general. Moreover, as expected, when the DGP is a partial cointegrated system, the testing procedure based on $\sup \mathcal{W}_1$ and $\sup \mathcal{W}_2$ has better power than the joint test ($\sup \mathcal{W}_0$ and Johansen's test) because it allows to discriminate between the unit root in z_t and the partial unit root. Note that, as always the case, the power of all tests falls as z_t tends toward a unit root process, *i.e.* when $\phi_2 \longrightarrow 1$. Thus, the tests suffer from a very low power when the process is near unit root.

Table 4: Power of Tests

		$\sup \mathcal{W}_1$, $\sup \mathcal{W}_2$		
ϕ_2	0.2	0.5	0.7	0.9
Γ_{22}				
-0.5	1.00	0.967	0.656	0.137
0.0	1.00	0.963	0.666	0.133
0.5	1.00	0.969	0.653	0.131
		$\sup \mathcal{W}_0$		
-0.5	0.996	0.852	0.391	0.063
0.0	0.995	0.852	0.400	0.059
0.5	0.997	0.854	0.388	0.056
		λ_{tr}		
-0.5	0.959	0.754	0.388	0.087
0.0	0.950	0.752	0.379	0.097
0.5	0.960	0.768	0.398	0.103

Note: $T = 100$. Nominal size 5%. Rejection rates from 1000 replications.

6 Application : the term structure of interest rates

We illustrate the partial cointegration test by analyzing the relationship between various long-term (r_t) and short-term (R_t) U.S. interest rates. According to the theory of the term structure of interest rates (Shiller, 1990), the long-term interest rates should be cointegrated with the interest rates on shorter maturity bonds (Campbell and Shiller, 1987, 1988). The basic idea behind threshold cointegration in the spread between the two interest rates is that the nonlinearity in the adjustment process stems from interventions of monetary authority. In order to test this hypothesis, we use the monthly interest rate series of McCulloch and Kwon (1993) which are constructed from the prices of U.S. Treasury securities and expressed as continuously compounded zero-coupon bonds. Following Hansen and Seo (2001), we estimate and test models of threshold cointegration using a selection of bond rates with maturities ranging from 1 to 120 months for the period January 1951 - February 1991 (482 observations). It is generally agreed that interest-rates series are $I(1)$. For this reason the unit root test results on each variable, which confirm I(1), are not reported here to save space.

Following the Engle-Granger methodology, we estimate the long-run equilibrium relationship by OLS: $R_t = \mu + \beta r_t + z_t$. Once the cointegrating vector is estimated, the residuals are used to estimate the TVECM by Least Squares over 340 grid points on the threshold parameter. Note that the transition variable is taken be Δz_{t-1} to ensure its stationarity in the case of no-cointegration or partial cointegration. In other words, the threshold process implicitly assumed for the disequilibrium term is a M-TAR. The grid set corresponds to 70% of estimated values of the transition variable $\Delta \hat{z}_{t-1}$ which are possible values of the threshold. The discarded values are the largest and smallest 15% of $\Delta \hat{z}_{t-1}$. As discussed earlier, this choice of the percentage of discarded observations is somewhat arbitrary. In practice, it is chosen in order to ensure a sufficient number of observations to conduct regression in each regime and Andrews (1993) suggests $\lambda = 0.15$ is a reasonable choice. The TVECM to be estimated is formally given by:

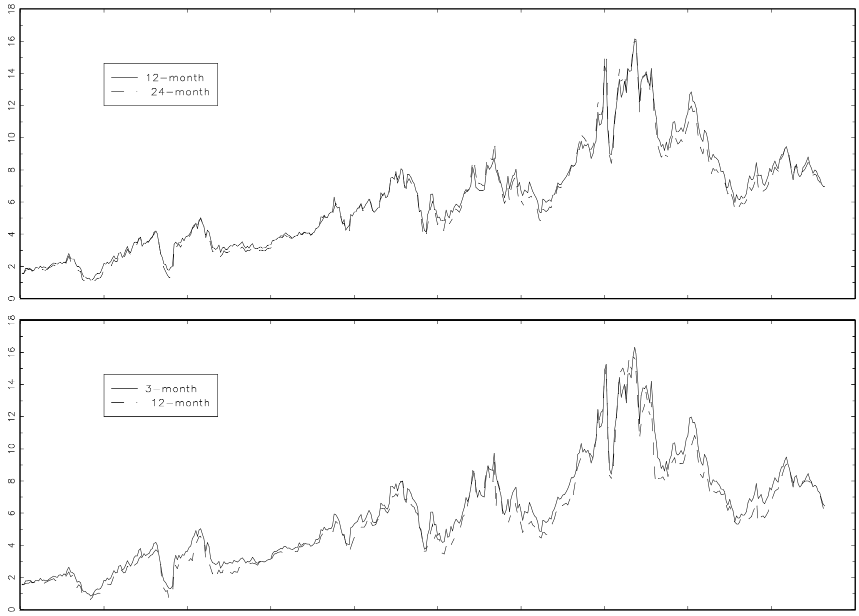
$$\begin{pmatrix} \Delta r_t \\ \Delta R_t \end{pmatrix} = \mu + \gamma_1 z_{t-1} \mathbb{I}_{\{F(\Delta z_{t-1}) \leq F(\theta)\}} + \gamma_2 z_{t-1} \mathbb{I}_{\{F(\Delta z_{t-1}) \leq F(\theta)\}} + \Gamma \begin{pmatrix} \Delta r_{t-1} \\ \Delta R_{t-1} \end{pmatrix} + \varepsilon_t$$

The final results, such that $F(\hat{\theta}) = \arg\min \log \left| \hat{\Sigma} \left(\hat{\beta}, F(\theta) \right) \right|$, are reported below for two interesting pairs of interest rates: 3-month / 12-month and 12-month / 24-month. These series are depicted in Figure 3. Note that the order of TVECM was selected to minimize the AIC and BIC criteria.

For each model, we compute the statistics $\sup \mathcal{W}_{T,1}$, $\sup \mathcal{W}_{T,2}$ and $\sup \mathcal{W}_{T,0}$ over 340 grid points. Since the interest rates are highly heteroskedastic, standard deviations of parameters (in brackets) are computed using the Newey-West (1987) method in order to obtain an heteroskedasticity-consistent value. That is, Wald statistics are given by $\tilde{\mathcal{W}}_{T,j} = (\hat{\gamma}_j(u))' \tilde{\Omega} (\hat{\gamma}_j(u))^{-1} (\hat{\gamma}_j(u))$, for $j = 0, 1, 2$ where $\tilde{\Omega}(\hat{\gamma}_j(u))$ is the Newey-West covariance matrix. Thus the adequate test statistic is: $\sup \tilde{\mathcal{W}}_{T,j}(u)$. We also report the heteroscedasticity-robust Ljung Box statistics $\left(Q^W\right)$ on residuals for 6 and 12 lags. Their values show that the residuals do not exhibit autocorrelation.

Examining the test results, one can conclude that there is evidence of threshold cointegration in the first model whereas for the pair 3-month / 12-month, the pure threshold cointegration is not an obvious fact, because $\sup \mathcal{W}_2$ is not significant at 1%. Therefore, the null hypothesis could not be accepted and the alternative is consistent with the partial cointegration hypothesis. As expected, the statistic $\sup \mathcal{W}_0$ rejects the null of no-cointegration hypothesis for both models but it cannot discriminate between the stationary threshold cointegration case and the partial cointegration case.

Figure 3: Treasury Bond Rates



12-month / 24-month

$$R_t = \underset{(0.074)}{-0.231} + \underset{(0.005)}{1.007} r_t + z_t$$

$$\Delta R_t = \underset{(0.021)}{0.028} + \underset{(0.074)}{0.066} z_{t-1}^{(1)} \underset{(0.128)}{-0.617}\, z_{t-1}^{(2)} \underset{(0.146)}{-0.121}\, \Delta R_{t-1}^{(1)} + \underset{(0.160)}{0.379} \Delta r_{t-1}^{(1)} + \varepsilon_{1t}$$

$$u = 0.848$$

$$Q^W(6) = 2.019 \quad ; \quad Q^W(12) = 9.164$$

$$\Delta r_t = \underset{(0.018)}{0.026} + \underset{(0.064)}{0.135} z_{t-1}^{(1)} \underset{(0.110)}{-0.406}\, z_{t-1}^{(2)} + \underset{(0.126)}{0.004} \Delta R_{t-1}^{(1)} + \underset{(0.138)}{0.151} \Delta r_{t-1}^{(1)} + \varepsilon_{2t}$$

$$Q^W(6) = 1.951 \quad ; \quad Q^W(12) = 8.294$$

$$\sup \mathcal{W}_1 = 31.751^{a,b,c,d} \quad ; \quad \sup \mathcal{W}_2 = 37.810^{a,b,c,d} \quad ; \quad \sup \mathcal{W}_0 = 58.527^{a,b,c,d}$$

3-month / 12-month

$$R_t = \underset{(0.103)}{-0.310} + \underset{(0.006)}{0.980} r_t + z_t$$

$$\Delta R_t = \underset{(0.022)}{0.014} + \underset{(0.087)}{0.051} z^{(1)}_{t-1} \underset{(0.065)}{-0.133}\, z^{(2)}_{t-1} \underset{(0.082)}{-0.172}\, \Delta R^{(1)}_{t-1} + \underset{(0.084)}{0.344} \Delta r^{(1)}_{t-1} + \varepsilon_{1t}$$

$$u = 0.175$$

$$Q^W(6) = 2.532 \quad ; \quad Q^W(12) = 8.857$$

$$\Delta r_t = \underset{(0.018)}{0.021} + \underset{(0.073)}{0.388} z^{(1)}_{t-1} + \underset{(0.054)}{0.015} z^{(2)}_{t-1} \underset{(0.069)}{-0.157}\, \Delta R^{(1)}_{t-1} + \underset{(0.071)}{0.287} \Delta r^{(1)}_{t-1} + \varepsilon_{2t}$$

$$Q^W(6) = 3.732 \quad ; \quad Q^W(12) = 11.152$$

$$\sup \mathcal{W}_1 = 81.370^{a,b,c,d} \quad ; \quad \sup \mathcal{W}_2 = 19.053^{a,c} \quad ; \quad \sup \mathcal{W}_0 = 96.743^{a,b,c,d}$$

[a] means that H_0 could not be accepted at 5% based on asymptotic critical values.
[b] means that H_0 could not be accepted at 1% based on asymptotic critical values.
[c] means that H_0 could not be accepted at 5% based on finite-sample critical value: 16.417 for $\sup \mathcal{W}_j$ and 21.968 for $\sup \mathcal{W}_0$.
[d] means that H_0 could not be accepted at 1% based on finite-sample critical value: 21.460 for $\sup \mathcal{W}_j$ and and 27.811 for $\sup \mathcal{W}_0$.

7 Conclusion

This paper investigates a partial cointegration case of cointegration in one regime and no cointegration in another within the context of threshold error correction models developed by Balke and Fomby (1997). It is called partial cointegration following the same terminology as Caner and Hansen (2001) who examined the case of a presence of a unit root in a TAR process. We propose a test of the hypothesis of no cointegration against the alternatives of threshold cointegration and partial cointegration. Asymptotic properties of the test statistics are derived, critical values are provided and the finite sample size and power are analysed. An empirical application is carried out for illustrating the practical relevance of such a test.

Appendix

A.1 PROOF OF PROPOSITION 1:

We use the simplified notations $I_t = \mathbb{1}_{\{U_t \leq u\}}$ and $(1 - I_t) = \mathbb{1}_{\{U_t > u\}}$. Using $z_t = \phi_1 \alpha' y_{t-1}^{(1)} + \phi_2 \alpha' y_{t-1}^{(2)} + \epsilon_t$ in VECM (2) we have

$$\Delta y_t = \gamma \alpha' y_{t-1} + \begin{pmatrix} 1 & \beta' \\ 0 & I_m \end{pmatrix} \begin{pmatrix} \phi_1 \alpha' y_{t-1}^{(1)} + \phi_2 \alpha' y_{t-1}^{(2)} + \epsilon_t \\ \eta_t \end{pmatrix}$$

$$\Delta y_t = \gamma \alpha' y_{t-1} + \begin{pmatrix} \phi_1 I_t + \phi_2 (1 - I_t) \\ 0 \end{pmatrix} \alpha' y_{t-1} + \begin{pmatrix} \epsilon_t + \beta' \eta_t \\ \eta_t \end{pmatrix},$$

so, we get:

$$\Delta y_t = \begin{pmatrix} -1 + \phi_1 I_t + \phi_2 (1 - I_t) \\ 0 \end{pmatrix} \alpha' y_{t-1} + \varepsilon_t,$$

which can be rewritten as $\Delta y_t = \pi^{(1)} y_{t-1}^{(1)} + \pi^{(2)} y_{t-1}^{(2)} + \varepsilon_t$,

with $\pi^{(j)} = \begin{pmatrix} -\Phi_j (1) \\ 0 \end{pmatrix} \alpha'$, with $\Phi_j (L) = 1 - \phi_j L$.

Now, assume that η_t has an finite q-autoregressive representation, $G(L) \eta_t = \zeta_t$, where ζ_t, is an independently and identically distributed (iid) process with 0 mean and such that $G(L) = I_m - L\ g(L)$, with $g(L) = \sum_{i=1}^{q} g_i L^{i-1}$, where g_i are matrices $(m \times m)$ of coefficients.

The previous VECM becomes

$$\Delta y_t = \pi^{(1)} y_{t-1}^{(1)} + \pi^{(2)} y_{t-1}^{(2)} + \begin{pmatrix} \beta' L g(L) \eta_{t-1} \\ L g(L) \eta_{t-1} \end{pmatrix} + \begin{pmatrix} \beta' \zeta_t + \epsilon_t \\ \zeta_t \end{pmatrix}$$

Denoting $\varepsilon_t = \begin{pmatrix} \beta' \zeta_t + \epsilon_t \\ \zeta_t \end{pmatrix}$, we have:

$$\Delta y_t = \pi^{(1)} y_{t-1}^{(1)} + \pi^{(2)} y_{t-1}^{(2)} + \begin{pmatrix} 0 & \beta' g(L) \\ 0 & g(L) \end{pmatrix} \Delta y_{t-1} + \varepsilon_t. \quad \blacksquare$$

A.2 PROOF OF THEOREM 1:

Recall that under the null hypothesis, we have $\Delta y_t = \underline{\Gamma}(L)^{-1} \varepsilon_t$,
where $\underline{\Gamma}(L) = I_n - L\, \Gamma(L)$ such that $\underline{\Gamma}(L)$ has roots lying outside the unit circle, $i.e.$ $\underline{\Gamma}(z) \neq 0$, for all $|z| \leq 1$.

Then we have $T^{-1/2} y_t \Longrightarrow W(r) = \underline{\Gamma}(1)^{-1} \Sigma^{1/2} B(r)$,
where $B(r)$ is a standard n-vector Brownian.

Consider the Wald statistic which tests the unit root in the first regime:

$$(A2.1) \qquad \mathcal{W}_{T,1} = tr\left\{\hat{\Sigma}^{-1}\tfrac{1}{T}\left(Z_1'M_X\varepsilon\right)'T^2\left(Z_1'M_XZ_1\right)^{-1}\tfrac{1}{T}\left(Z_1'M_X\varepsilon\right)\right\}$$
$$= tr\left\{\tfrac{1}{T}\left(Z_1'M_X\varepsilon\hat{\Sigma}^{-1/2}\right)'T^2\left(Z_1'M_XZ_1\right)^{-1}\tfrac{1}{T}\left(Z_1'M_X\varepsilon\hat{\Sigma}^{-1/2}\right)\right\}$$

and using the result that for conformable matrices F and G such that $H = [F\ G]$,
$P_H = H\left(H'H\right)^{-1}H' = P_F + M_FG\left(G'M_FG\right)^{-1}G'M_F$, with $M_F = I - P_F$,
we get for the first term of A2.1

$$(A2.2) \qquad Z_1'M_X\varepsilon = Z_1'M_R\varepsilon + Z_1'M_RZ_2\left(Z_2'M_RZ_2\right)^{-1}Z_2'M_R\varepsilon$$

Consider the first term of $(A2.2)$

$$T^{-1}Z_1'M_R\varepsilon = T^{-1}Z_1'\varepsilon - T^{-3/2}Z_1'\mathbf{R}\left(T^{-1}\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{R}'\,T^{-1/2}\varepsilon$$

Under the null z_t being $I(1)$, we can use Lemma 1 to obtain the limit of $T^{-1}Z_1'\varepsilon$:

$$T^{-1}Z_1'\varepsilon = T^{-1}\alpha'\textstyle\sum^T y_{t-1}\mathbb{I}_{\{U_t\le u\}}\varepsilon_t' \Longrightarrow \Sigma^{1/2}\int V(r)\,dB(r,u)',$$

where $V(r)$ is an univariate Brownian motion such that $V(r) = \alpha'B(r)$.

Since $T^{-3/2}Z_1'\mathbf{R}$ is $o_p(1)$, we have:

$$T^{-1}Z_1'M_R\varepsilon \Longrightarrow \Sigma^{1/2}\int V(r)\,dB(r,u)'.$$

Consider now, the second term of $(A2.2)$.

$$T^{-1}Z_1'M_RZ_2\left(Z_2'M_RZ_2\right)^{-1}Z_2'M_R\varepsilon = T^{-2}A\left(T^{-2}D^{-1}\right)T^{-1}C.$$

where

$$T^{-2}A = T^{-2}Z_1'Z_2 - T^{-1}Z_1'\boldsymbol{R}\left(T^{-1}\mathbf{R}'\mathbf{R}\right)^{-1}\left(T^{-3/2}\right)T^{-1/2}\mathbf{R}'Z_2 \quad ,$$
$$T^{-1}C = T^{-1}Z_2'\varepsilon - T^{-3/2}Z_2'\mathbf{R}\left(T^{-1}\mathbf{R}'\mathbf{R}\right)^{-1}T^{-3/2}\mathbf{R}'Z_2,$$
$$T^{-2}D = T^{-2}Z_2'Z_2 - T^{-3/2}Z_2'\mathbf{R}\left(T^{-1}\mathbf{R}'\mathbf{R}\right)^{-1}T^{-3/2}\mathbf{R}'Z_2 \ ,$$

Let us look at them one by one.

By construction $Z_1'Z_2 = Z_2'Z_1 = 0$ and due to over normalization by $T^{-3/2}$, we have $T^{-2}A = o_p(1)$.

Since the second term of $T^{-1}C$ is $o_p(1)$, using Lemma 1, we get:

$$T^{-1}C = T^{-1}Z_2'\varepsilon - o_p(1) \Longrightarrow \Sigma^{1/2}\int V(r)\,dB(r,1-u)'$$

where $B(r,1-u)$ is a two-parameter Brownian motion on $(r, S(\theta))$,

as $T^{-1}Z_2'\varepsilon \Longrightarrow \Sigma^{1/2}\int V(r)\,dB(r,1-u)'$.

Finally $T^{-2}D \Longrightarrow S(\theta)\int V^2(r)\,dr$.

Therefore the limiting distribution of $(A2.2)$ is given by:

$\frac{1}{T}Z_1'M_X\varepsilon \Longrightarrow \Sigma^{1/2}\int V(r)\,dB(r,u)'$.

In the same way, for the middle term of the Wald statistic $(A2.1)$, we have:

$Z_1'M_XZ_1 = Z_1'Z_1 - Z_1'P_RZ_1 + Z_1'M_RZ_2\left(Z_2'M_RZ_2\right)^{-1}Z_2'M_RZ_1$

$T^{-2}Z_1'M_XZ_1 =$
$T^{-2}Z_1'Z_1 - T^{-3/2}Z_1'\mathbf{R}\left(T^{-1}\mathbf{R}'\boldsymbol{R}\right)^{-1}T^{-3/2}\mathbf{R}'Z_1 + \left(T^{-1}\right)T^{-2}A\left(T^{-2}D^{-1}\right)T^{-1}C$

Hence the limiting distribution of the middle term is $T^{-2}Z_1'M_XZ_1 \Longrightarrow F(\theta)\int V^2(r)\,dr$.

Putting the above terms appropriately together yields the expression of limiting distribution in Theorem 1.

A similar derivation can be carried out to obtain the asymptotic distribution of the sup-Wald statistic for the parameter estimators of regime 2.

For the statistic $\mathcal{W}_{T,0}$ which jointly tests γ_1 and γ_2 , the algebra are similar but easier because $X = \mathbf{R}$ with only stationary variables, so $M_X = M_R$. Let $Z = [Z_1 \;\; Z_2]$ be the $(T \times 2)$ matrix. Then we have:

$\mathcal{W}_{T,0} = tr\left\{\hat{\Sigma}^{-1}\left(T^{-1}Z'M_R\varepsilon\right)'\left(T^{-2}Z'M_RZ\right)^{-1}\left(T^{-1}Z'M_R\varepsilon\right)\right\}$

Let us take the first term of the above expression.

$T^{-1}Z'M_R\varepsilon = T^{-1}Z'\varepsilon - T^{-3/2}Z'\mathbf{R}\left(T^{-1}\mathbf{R}'\mathbf{R}\right)^{-1}T^{-1/2}\mathbf{R}'\varepsilon,$

Using Lemma 1, we get:

$T^{-1}Z'M_R\varepsilon \Longrightarrow \Sigma^{1/2}\mathbf{Q}(r,u)$

where $\mathbf{Q}(r,u) = \begin{pmatrix} \int V(r)\,dB(r,u)' \\ \int V(r)\,dB(r,1-u)' \end{pmatrix}$ is a $(2 \times n)$ matrix.

Similarly, for the middle term :

$T^{-2}Z'M_RZ = T^{-2}Z'Z - T^{-3/2}Z'\mathbf{R}\left(T^{-1}\mathbf{R}'\mathbf{R}\right)^{-1}T^{-3/2}\mathbf{R}'Z$, so

$T^{-2}Z'Z \Longrightarrow \begin{pmatrix} F(\theta) & 0 \\ 0 & S(\theta) \end{pmatrix}\int V^2(r)\,dr.$ ■

A.3 PROOF OF THEOREM 2:

This could be proved starting with the results of Phillips and Ourialis (1990).

Consider two processes ν_t iid $(0, I_n)$ and y_t iid $(0, \Sigma)$ such that: $T^{-1/2}\nu_t \Longrightarrow U(r)$ and $T^{-1/2}y_t \Longrightarrow W(r)$ are two n-vector Brownian motions.

Let us partition y_t as $\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$; $W(r)$ as $\begin{pmatrix} W_1(r) \\ W_m(r) \end{pmatrix}$ and $\Sigma as \begin{pmatrix} \sigma_{11} & \sigma'_{21} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}$ where Σ_{22} is a positive matrix $(m \times m)$.

Let us decompose Σ as $L'L$ with $L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{pmatrix}$, such that $W(r) \equiv L'U(r)$. We can then write $T^{-1/2}y_t = T^{-1/2}L'\xi_t \Longrightarrow W(r)$.

We also have :

$T^{-2}\nu'\nu \Longrightarrow \int U(r)U(r)'dr = \begin{pmatrix} s_{11} & s'_{21} \\ s_{21} & S_{22} \end{pmatrix}$ and

$T^{-2}y'y = T^{-2}L'\nu'\nu L \Longrightarrow \int W(r)W(r)'dr = \begin{pmatrix} \lambda_{11} & \lambda'_{21} \\ \lambda_{21} & \Lambda_{22} \end{pmatrix} = L'SL.$

Thus, $\hat{\beta}_T = (y'_2 y_2)^{-1} y'_2 y_1 \Longrightarrow \Lambda_{22}^{-1}\lambda_{21}$ and $\hat{\alpha}' = (1, -\hat{\beta}') \Longrightarrow \alpha'_l = \left(1, \quad -\lambda'_{21}\Lambda_{22}^{-1}\right)$,

Hence $T^{-1/2}\hat{\alpha}'_T y_t \Longrightarrow \alpha'_l W(r) = \alpha'_l L'U(r)$, with $\alpha'_l L' = \left(l_{11}, \quad l'_{21} - \lambda'_{21}\Lambda_{22}^{-1}L'_{22}\right)$.

By noting that $l'_{21} - \lambda'_{21}\Lambda_{22}^{-1}L'_{22} = -l_{11}s'_{21}S_{22}^{-1}$, we have $L\alpha_l = l_{11}\begin{pmatrix} 1 \\ -S_{22}^{-1}s_{21} \end{pmatrix}$, and

$$\begin{aligned} \alpha'_l L'U(r) &= l_{11}\left(1, \quad - s'_{21}S_{22}^{-1}\right)\begin{pmatrix} U_1(r) \\ U_m(r) \end{pmatrix} \\ &= l_{11}\left[U_1(r) - \left(\int U_1(r)U_m(r)'dr\right)\left(\int U_m(r)U_m(r)'dr\right)^{-1} U_m(r)\right]. \end{aligned}$$

We can finally write $T^{-1/2}\hat{\alpha}'_T y_t = T^{-1/2}\hat{z}_t \Longrightarrow \sigma U(r)$,

Noting $V(r) = U_1(r) - \left(\int U_1(r)U_m(r)'dr\right)\left(\int U_m(r)U_m(r)'dr\right)^{-1} U_m(r)$,

we obtain the proof of Theorem 2 using Lemma 1 with $\hat{z}_t \sim I(1)$ and a similar derivation to that of Appendix A.2 . ■

References

[1] Andrews, D.W.K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61, 821-856.

[2] Andrews, D.W.K., and W. Ploberger (1994): "Optimal Tests when a Nuisance Parameter is Present only under the Alternative", *Econometrica*, 62, 1383-1414.

[3] Balke, N.S., and T.B. Fomby (1997): "Threshold Cointegration", *International Economic Review*, 38, 627-645.

[4] Bellingsley, P. (1968): Convergence of Probability Measures. New York: John Wiley.

[5] Campbell, J.Y., and R.J. Shiller (1987): "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, 95(5), 1062-1088.

[6] Campbell, J.Y., and R.J. Shiller (1988): "Interpreting Cointegrated Models", *Journal of Economic Dynamics and Control*, 12(2), 505-522.

[7] Caner, M., and B.E. Hansen (2001): "Threshold Autoregression with a Unit Root", *Econometrica*, 69(6), 1555-1596.

[8] Chan, K.S., Petruccelli, J.D., Tong, H., and S.W. Woolford (1985): "A Multiple Threshold AR(1) Model", *Journal of Applied Probability*, 22, 267-279.

[9] Chan, K.S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model", *The annals of Statistics*, 21(1), 520-533.

[10] Davies, R.B. (1987): "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative", *Biometrika*, 74, 33-43.

[11] Enders, W., and P.L. Siklos (2001): "Cointegration and Threshold Adjustment", J*ournal of Business & Economic Statistics*, 19(2), 166-176.

[12] Engle, R.F., and C.W. Granger (1987): "Co-integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55(2), 251-276.

[13] Gonzalo, J., and P.Y. Pitarakis (2006): "Threshold Effects in Multivariate Error Correction Models", Palgrave Handbook of Econometrics, Vol I, Chapter 15.

[14] Hansen, B.E. (1996): "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis", *Econometrica*, 64(2), 413-430.

[15] Hansen, B.E. (1997): "Inference in TAR Models", *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1-14.

[16] Hansen, B.E., and B. Seo (2002): "Testing for Two-regime Threshold Cointegration in Vector Error-correction Models", *Journal of Econometrics*, 110, 293-318.

[17] Johansen, S. (1995): Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Advanced Texts in Econometrics. Oxford University Press, Oxford, UK.

[18] Kapetanios, G., and Y. Shin (2002): "Unit Root Tests in Three-Regime SETAR Models", Working Paper of Department of Economics, Queen Mary, University of London.

[19] Krugman, P.R. (1991): "Target Zones and Exchange Rate Dynamics", *Quaterly Journal of Economics*, 106(3), 669-682.

[20] Lo, M.C., and E. Zivot (2001): "Threshold Cointegration and Nonlinear Adjustment to the Law of One Price", *Macroeconomic Dynamics*, 5, 533-576.

[21] McCulloch, J.H., and H.C. Kwon (1993): "US Term Structure Data, 1947-1991", Ohio State University Working Paper No. 93-6.

[22] Newey, W.K., and K.D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55(3), 703-708.

[23] Phillips, P.C.B. (1987): "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.

[24] Phillips, P.C.B., and S. Ouliaris (1990): "Asymptotic Properties of Residual Based Tests for Cointegration", *Econometrica*, 58(1), 165-193.

[25] Phillips, P.C.B. (1991): "Optimal Inference in Cointegrated Systems", *Econometrica*, 59(3), 283-306.

[26] Seo, M. (2004): "Testing for the Presence of Threshold Cointegration in a Threshold Vector Error Correction Model", Working Paper, University of Wisconsin.

[27] Shiller, R.J. (1990): "The Term Structure of Interest Rates", Handbook of Monetary Economics, volume 1, Elsevier.

[28] Tong, H. (1983): Threshold Models in Non-linear Time Series Analysis, Lecture Notes in Statistics, 21. Berlin: Springer.

[29] Tsay, R.S. (1989): "Testing and Modeling Threshold Autoregressive Processes", *Journal of the American Statistical Association*, 84(405), 231-240.

[30] Tsay, R.S. (1998): "Testing and Modeling Multivariate Threshold Models", *Journal of the American Statistical Association*, 93(443), 1188-1202.

[31] Tweedie, R.L. (1975): "Sufficient Conditions for Ergodicity and Recurrence of Markov on a General State Space", *Stochastic Processes Appl.*, 3, 385-403.

Publications récentes du Département d'économétrie

pouvant être obtenues à l'adresse suivante :

Université de Genève
UNI MAIL
A l'att. de Mme Caroline Schneeberger
Département d'économétrie
40, Bd du Pont-d'Arve
CH - 1211 Genève 4
ou sur
INTERNET : http//www.unige.ch/ses/metri/cahiers

2005.03 VAN BAALEN Brigitte, Tobias MÜLLER, Social Welfare effects of tax-benefit reform under endogenous participation and unemployment, Février 2005, 42 pages.

2005.02 CZELLAR Véronika, G. Andrew KAROLYI, Elvezio RONCHETTI, Indirect Robust Estimation of the Short-term Interest Rate Process, Mars 2005, 29 pages.

2005.01 E. CANTONI, C. FIELD, J. MILLS FLEMMING, E. RONCHETTI, Longitudinal Variable Selection by Cross-Validation in the Case of Many Covariates, Février 2005, 17 pages.

2004.15 KRISHNAKUMAR Jaya, Marc-Jean MARTIN, Nils SOGUEL, Application of Granger Causality Tests to Revenue and Expenditure of Swiss Cantons, Décembre 2004, 27 pages.

2004.14 KRISHNAKUMAR Jaya, Gabriela FLORES, Sudip Ranjan BASU, Spatial Distribution of Welfare Across States and Different Socio-Economic Groups in Rural and Urban India, Mai 2004, 66 pages.

2004.13 KRISHNAKUMAR Jaya, Gabriela FLORES, Sudip Ranjan BASU, Demand 2004System Estimations and Welfare Comparisons : Application to Indian Household Data, Mai 2004, 70 pages.

2004.12 KRISHNAKUMAR Jaya, Going beyond functionings to capabilities: an econometric model to explain and estimate capabilities, Août 2004, 29 pages.

2004.11 MÜLLER Tobias, RAMSES Abul Naga, KOLODZIEJCZYK Christophe, The Redistributive Impact of Altrnative Income Maintenance Schemes : A Microsimulation Study using Swiss Data, Août 2004, 40 pages.

2004.10 GHIGLINO Christian, DUC François, Expectations in an OG Economy, Août 2004, 21 pages.

2004.09 GHIGLINO Christian, OLSZAK-DUQUENNE Marielle, On the Impact of Heterogeneity on Indeterminacy, Août 2004, 29 pages.