Time Varying Cointegration*

Herman J. Bierens[†] and Luis F. Martins[‡]

February 9, 2010

Abstract

In this paper we propose a time varying vector error correction model in which the cointegrating relationship varies smoothly over time. The Johansen setup is a special case of our model. A likelihood ratio test for time-invariant cointegration is defined and its asymptotic chi-square distribution is derived. We apply our test to the purchasing power parity hypothesis of international prices and nominal exchange rates, and find evidence of time-varying cointegration.

Keywords: Time Varying Error Correction Model; Chebyshev Polynomials; Likelihood Ratio; Power; Trace Statistic

J.E.L. Classification: C32.

---

*The authors are grateful to Peter Phillips, Pentti Saikkonen and two anonymous referees for helpful comments.

[†]Department of Economics and CAPCP, Pennsylvania State University, 608 Kern Graduate Building, University Park, PA 16802, USA. E-mail: hbierens@psu.edu. Support for research within the Center for the Study of Auctions, Procurements, and Competition Policy (CAPCP) at Penn State has been provided by a gift from the Human Capital Foundation (http://www.hcfoundation.ru/en/).

[‡]UNIDE and Department of Quantitative Methods, ISCTE - Business School, Av. das Forças Armadas, 1649-026 Lisbon, Portugal. E-mail: luis.martins@iscte.pt. Financial support under grants SFRH/BD/814/2000 and PTDC/ECO/68367/2006 from the Fundação para a Ciência e Tecnologia is gratefully acknowledged. This paper is a substantial further elaboration of Martins' Ph.D. thesis at the Pennsylvania State University. A previous version of this paper was presented by Martins at the Econometric Society World Congress 2005 in London.

1 Introduction

Since the seminal papers by Granger (1987), Engle and Granger (1987) and Johansen (1988), the growth of literature on cointegration has been impressive. In the standard approach it is assumed that the cointegrating vectors do not change over time. However, this assumption is quite restrictive.

The literature on structural change and cointegration has focused on developing procedures to detect structural breaks and/or to estimate their dates. Papers addressing these issues in a single-equation framework include Hansen (1992), Quintos and Phillips (1993), Hao (1996), Andrews et al. (1996), Bai et al. (1998), and Kuo (1998), among others (see Maddala and Kim 1998 for a survey). Moreover, Lütkepohl et al. (2003), Inoue (1999) and Johansen et al. (2000) analyze the effects of breaks in the deterministic trend. In the context of a system of equations, which is the focus of our analysis, the main contributions are those by Seo (1998), who extends the tests of Hansen (1992). Hansen and Johansen (1999) and Quintos (1997) propose fluctuation tests (based on recursive sequences of eigenvalues and cointegrating vectors) for parameter constancy in cointegrated VAR's, but they do not parameterize the shifts. Regarding time-varying error correction models, Hansen (2003) generalizes reduced-rank methods to cointegration under sudden regime shifts with a known number of break points. Andrade et al. (2005) study a similar model as Hansen (2003) and develop tests on the cointegration rank and on the cointegration space under known and unknown break points.

Also, the Markov-switching approach of Hall et al. (1997), and the smooth transition model of Saikkonen and Choi (2004) provide an interesting way of modeling shifts in the cointegrating vectors. The first authors considering sudden shifts between two states, whereas the latter authors permit a gradual shift between regimes. Lütkepohl et al. (1999) and Terasvirta and Eliasson (2001) propose money demand functions modeled by single-equation error correction models in which a smooth transition stationary term is added. The transition function is driven by one of the processes of the long-run relationship.

Park and Hahn (1999) propose a cointegrating regression in the spirit of Engle and Granger (1987) with parameters that vary with time. They model the elements of a (single) cointegrating vector as smooth functions of time, via Fourier series expansions. They derive the asymptotic properties of the semi-nonparametric sieve estimators involved and propose several residualbased specification tests.

The latter approach is part of the growing literature on modeling nonlinear long run relationships. See for example Blake and Fomby (1997), de Jong (2001), Granger and Yoon (2002), Harris et al. (2002) and Juhl and Xiao (2005), among others.

In this paper we propose a likelihood ratio test for time varying cointegration, with time invariant cointegration as the null hypothesis, by allowing the cointegrating vectors in a vector error correction model (VECM) to be smooth functions of time, similar to Park and Hahn (1999). In particular, we propose to model these time varying cointegrating vectors via expansions in terms of Chebyshev time polynomials. The resulting extended VECM can be estimated similar to Johansen's (1988, 1991, 1995) ML approach. The null hypothesis of standard cointegration then corresponds to the hypothesis that the parameters in the VECM that are related to Chebyshev time polynomials are jointly zero. The latter hypothesis can be tested via a likelihood ratio test.

The remainder of the paper is organized as follows. In Section 2 we introduce the time varying (TV) VECM using Chebyshev time polynomials. In Section 3 we propose a likelihood ratio test to distinguish Johansen's standard cointegration from our time-varying alternative, for the case without drift, and show that the asymptotic null distribution is chi-square. In Section 4 the asymptotic power of the test is derived analytically and via Monte Carlo simulations. In Section 5 we show that our results carry over to the drift case. In Section 6 we illustrate the merits of our approach by testing for TV cointegration of international prices and nominal exchange rates. In Section 7 we make some concluding remarks. The proofs of the lemmas and theorems can be found in either the Appendix at the end of this paper or in Bierens and Martins (2009).

As to some notations, "$\Rightarrow$" denotes weak convergence, "$\xrightarrow{d}$" denotes convergence in distribution, and $\mathbf{1}(.)$ is the indicator function.

2 Definitions and Representations

For the $k \times 1$ vector time series Y_t, we assume that for some t there are fixed $r < k$ linearly independent columns of the time-varying $k \times r$ matrix $\beta_t = (\beta_{1t}, \beta_{2t}, ..., \beta_{rt})$ of cointegrating vector. Thus, these columns form the basis of the time-varying space of cointegrating vectors, $S_t^c = \text{span}(\beta_{1t}, \beta_{2t}, ..., \beta_{rt}) \subset$

$\mathbb{R}^k$, $t = 1, 2, ...$ The remaining $k - r$ orthogonal vectors, expressed by a $k \times (k - r)$ matrix $\beta_{t_\perp}$, are such that $\beta'_{t_\perp} Y_{t-1}$ does not represent a cointegrating relationship. The matrices β_t will be modeled using Chebyshev time polynomials.

2.1 Time Varying VECM Representation

Consider the time-varying VECM(p) with Gaussian errors, without intercepts and time trends,

$$\triangle Y_t = \Pi_t' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \triangle Y_{t-j} + \varepsilon_t, \ t = 1, ..., T, \tag{1}$$

where $Y_t \in \mathbb{R}^k$, $\varepsilon_t \sim i.i.d.\ N_k[0, \Omega]$ and T is the number of observations. Our objective is to test the null hypothesis of time-invariant (TI) cointegration, $\Pi_t' = \Pi' = \alpha\beta'$, where α and β are fixed $k \times r$ matrices with rank r, against TV cointegration of the type

$$\Pi_t' = \alpha\beta_t',$$

where α is the same as before but now the β_t's are time-varying $k \times r$ matrices with constant rank r. In both cases Ω and the Γ_j's are fixed $k \times k$ matrices, and $1 \leq r < k$.

Admittedly, this form of TV cointegration is quite restrictive, as only the β_t's are assumed to be time dependent. A more general form of TV cointegration is the case $Y_t = C_t Z_t$, where C_t is a sequence of nonsingular $k \times k$ matrices and $Z_t \in \mathbb{R}^k$ is a time-invariant cointegrated $I(1)$ process with a VECM(p) representation. Then Y_t has a VECM(p) representation, but where all the parameters are functions of t.

2.2 Chebyshev Time Polynomials

Chebyshev time polynomials $P_{i,T}(t)$ are defined by

$$\begin{aligned} P_{0,T}(t) &= 1, \ P_{i,T}(t) = \sqrt{2}\cos(i\pi(t - 0.5)/T), \\ t &= 1, 2, ..., T, \ i = 1, 2, 3, ... \end{aligned}$$

See for example Hamming (1973). Bierens (1997) uses them in his unit root test against nonlinear trend stationarity. Chebyshev time polynomials are orthonormal, in the sense that for all integers i,j, $\frac{1}{T}\sum_{t=1}^{T}P_{i,T}(t)\,P_{j,T}(t)=\mathbf{1}(i=j)$. Due to this orthonormality property, any function $g(t)$ of discrete time, $t=1,...,T$, can be represented by

$$g(t)=\sum_{i=0}^{T-1}\xi_{i,T}P_{i,T}(t)\,,\ \text{where } \xi_{i,T}=\frac{1}{T}\sum_{t=1}^{T}g(t)\,P_{i,T}(t)\,.$$

In this expression, $g(t)$ is decomposed linearly in components $\xi_{i,T}P_{i,T}(t)$ of decreasing smoothness. Therefore, if $g(t)$ is smooth (to be made more precise in Lemma 1 below), it can be approximated quite well by

$$g_{m,T}(t)=\sum_{i=0}^{m}\xi_{i,T}P_{i,T}(t)$$

for some fixed natural number $m<T-1$.

Lemma 1. *Let $g(t)=\varphi(t/T)$, where $\varphi(x)$ is a square integrable real function on $[0,1]$. Then*

$$\lim_{m\to\infty}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(g(t)-g_{m,T}(t)\right)^2=0.$$

Moreover, if $\varphi(x)$ is $q\geq 2$ times differentiable, where q is even, with $\varphi^{(q)}(x)=d^q\varphi(x)/(dx)^q$ satisfying $\int_0^1\left(\varphi^{(q)}(x)\right)^2dx<\infty$, then for $m\geq 1$,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(g(t)-g_{m,T}(t)\right)^2\leq\frac{\int_0^1\left(\varphi^{(q)}(x)\right)^2dx}{\pi^{2q}\left(m+1\right)^{2q}}.$$

Proof. See Bierens and Martins (2009).

Consequently, we may without loss of generality write β_t for $t=1,...,T$ as $\beta_t=\sum_{i=0}^{T-1}\xi_{i,T}P_{i,T}(t)\,,$ where $\xi_{i,T}=\frac{1}{T}\sum_{t=1}^{T}\beta_tP_{i,T}(t)\,,\ i=0,...,T-1$, are unknown $k\times r$ matrices. Then the null hypothesis of TI cointegration corresponds to $\xi_{i,T}=O_{k\times r}$ for $i=1,...,T-1$, and the alternative of TV cointegration corresponds to $\lim_{T\to\infty}\xi_{i,T}\neq O_{k\times r}$ for some $i\geq 1$. To make the latter operational, we will confine our analysis to TV alternatives for which $\lim_{T\to\infty}\xi_{i,T}\neq O_{k\times r}$ for some $i=1,...,m$, and $\xi_{i,T}=O_{k\times r}$ for all

$i > m$, where m is chosen in advance. Effectively this means that under the alternative β_t is specified as

$$\beta_t = \beta_m \left(t/T\right) = \sum_{i=0}^{m} \xi_{i,T} P_{i,T}\left(t\right) \tag{2}$$

for some fixed m. Because low-order Chebyshev polynomials are rather smooth functions of t, we allow β_t to change gradually over time under the alternative of TV cointegration, contrary to Hansen's (2003) sudden change assumption.

This specification of the matrix of time varying cointegrating vectors is related to the approach of Park and Hahn (1999). They consider a TV cointegrating relationship of the form $Z_t = \alpha_t' X_t + U_t$, where $Z_t \in \mathbb{R}$, X_t is a k-variate $I(1)$ process and U_t is a stationary process. Thus, with $Y_t = (Z_t, X_t')'$ and $\beta_t = (1, -\alpha_t')'$, $\beta_t' Y_t = U_t$ is stationary. Park and Hahn (1999) assume that the elements of α_t are of the form $\varphi\left(t/T\right)$, where $\varphi\left(x\right)$ has a Fourier flexible functional form.

2.3 Modeling TV Cointegration via Chebyshev Time Polynomials

Substituting $\Pi_t' = \alpha \beta_t' = \alpha \left(\sum_{i=0}^{m} \xi_i P_{i,T}\left(t\right)\right)'$ in (1) yields

$$\triangle Y_t = \alpha \left(\sum_{i=0}^{m} \xi_i P_{i,T}\left(t\right)\right)' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \triangle Y_{t-j} + \varepsilon_t$$

for some $k \times r$ matrices ξ_i, which can be written more conveniently as

$$\triangle Y_t = \alpha \xi' Y_{t-1}^{(m)} + \Gamma X_t + \varepsilon_t, \tag{3}$$

where $\xi' = \left(\xi_0', \xi_1', ..., \xi_m'\right)$ is an $r \times (m+1)k$ matrix of rank r, $Y_{t-1}^{(m)}$ is defined by

$$Y_{t-1}^{(m)} = \left(Y_{t-1}', P_{1,T}\left(t\right) Y_{t-1}', P_{2,T}\left(t\right) Y_{t-1}', ..., P_{m,T}\left(t\right) Y_{t-1}'\right)' \tag{4}$$

and

$$X_t = \left(\triangle Y_{t-1}', ..., \triangle Y_{t-p+1}'\right)'.$$

The null hypothesis of TI cointegration corresponds to $\xi' = \left(\beta', O_{r,k.m}\right)$, where β is the $k \times r$ matrix of TI cointegrating vectors, so that then $\xi' Y_{t-1}^{(m)} =$

$\beta' Y_{t-1}^{(0)}$, with $Y_{t-1}^{(0)} = Y_{t-1}$. This suggests to test the null hypothesis via a likelihood ratio test

$$LR^{tvc} = -2\left[\widehat{l}_T(r,0) - \widehat{l}_T(r,m)\right],$$

where $\widehat{l}_T(r,0)$ is the log-likelihood of the VECM(p) (3) in the case $m=0$, so that $Y_{t-1}^{(m)} = Y_{t-1}$, and $\widehat{l}_T(r,m)$ is the log-likelihood of the VECM(p) (3) in the case where $Y_{t-1}^{(m)}$ is given by (4), where in both cases r is the cointegration rank.

3 Testing TI Cointegration Against TV Cointegration

3.1 ML Estimation and the LR Test

Denote

$$\begin{aligned}
S_{00,T} &= \frac{1}{T}\sum_{t=1}^{T} \triangle Y_t \triangle Y_t' - \widehat{\Sigma}_{X\Delta Y}' \widehat{\Sigma}_{XX}^{-1} \widehat{\Sigma}_{X\Delta Y} \\
S_{11,T}^{(m)} &= \frac{1}{T}\sum_{t=1}^{T} Y_{t-1}^{(m)} Y_{t-1}^{(m)\prime} - \widehat{\Sigma}_{XY^{(m)}}' \widehat{\Sigma}_{XX}^{-1} \widehat{\Sigma}_{XY^{(m)}} \\
S_{01,T}^{(m)} &= \frac{1}{T}\sum_{t=1}^{T} \triangle Y_t Y_{t-1}^{(m)\prime} - \widehat{\Sigma}_{X\Delta Y}' \widehat{\Sigma}_{XX}^{-1} \widehat{\Sigma}_{XY^{(m)}} \\
S_{10,T}^{(m)} &= \left(S_{01,T}^{(m)}\right)',
\end{aligned}$$

where $\widehat{\Sigma}_{XX} = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$, $\widehat{\Sigma}_{X\Delta Y} = \frac{1}{T}\sum_{t=1}^{T} X_t \triangle Y_t'$, and $\widehat{\Sigma}_{XY^{(m)}} = \frac{1}{T}\sum_{t=1}^{T} X_t Y_{t-1}^{(m)\prime}$, and let $\widehat{\lambda}_{m,1} \geq \widehat{\lambda}_{m,2} \geq ... \geq \widehat{\lambda}_{m,r} \geq ... \geq \widehat{\lambda}_{m,(m+1)k}$ be the ordered solutions of the generalized eigenvalue problem

$$\det\left[\lambda S_{11,T}^{(m)} - S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)}\right] = 0. \tag{5}$$

Note that $\widehat{\lambda}_{m,k+1} = ... = \widehat{\lambda}_{m,(m+1)k} \equiv 0$, because the rank of $S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)}$ is k. Then similar to Johansen (1988) the log-likelihood $\widehat{l}_T(r,m)$, given r and

m, takes the form

$$\widehat{l}_T(r,m) = -0.5T.\sum_{j=1}^{r} \ln\left(1-\widehat{\lambda}_{m,j}\right) - 0.5T.\ln\left(\det\left(S_{00,T}\right)\right)$$

plus a constant. Therefore, given m and r, the LR test of the null hypothesis of standard (TI) cointegration against the alternative of TV cointegration takes the form

$$LR_T^{tvc} = -2\left[\widehat{l}_T(r,0) - \widehat{l}_T(r,m)\right] = T\sum_{j=1}^{r} \ln\left(\frac{1-\widehat{\lambda}_{0,j}}{1-\widehat{\lambda}_{m,j}}\right). \tag{6}$$

3.2 Data-Generating Process under the Null Hypothesis

For $m = 0$ we have the standard cointegration case:

Assumption 1. *$\triangle Y_t$ is a strictly stationary zero-mean k-variate Gaussian process with Wold decomposition $\triangle Y_t = C(L)U_t = \sum_{j=0}^{\infty} C_j U_{t-j}$, where $U_t \sim$ i.i.d. $N_k[0, I_k]$. The elements of the $k \times k$ matrices C_j decrease exponentially to zero as $j \to \infty$.*

We can write $\triangle Y_t$ as

$$\triangle Y_t = C(1)U_t + (1-L)D(L)U_t,$$

where

$$D(L) = \frac{C(L) - C(1)}{1-L}.$$

This is the well-known Beveridge-Nelson (1981) decomposition, which implies that

$$Y_t = C(1)\sum_{j=1}^{t} U_j + V_t + Y_0 - V_0, \tag{7}$$

where $V_t = D(L)U_t$ is a zero-mean stationary Gaussian process.

Assumption 2. *The matrix $C(1)$ is singular, with rank $1 \leq r < k$: There exists a $k \times r$ matrix β with rank r such that $\beta' C(1) = O_{r,k}$. Moreover, the $r \times k$ matrix $\beta' D(1)$ has rank r.*

For the time being we will also assume that

Assumption 3. $U_t = 0$ *for* $t < 1$,

so that $Y_0 = V_0 = 0$ in (7).

Admittedly, Assumption 3 is too restrictive, but is made to focus on the main issues. For the same reason we do not yet consider the more realistic case of drift in Y_t. Once we have completed the asymptotic analysis for the case under review, we will show what happens if there is drift in Y_t and Assumption 3 is dropped.

Under some further regularity conditions it follows from Assumptions 1 and 2 and the Granger representation theorem (see Engle and Granger 1987) that Y_t has a VECM representation. Rather than listing these standard regularity conditions we assume that the Granger representation theorem holds:

Assumption 4. *Under Assumptions 1-3*, Y_t *has the VECM*(p) *representation*

$$\triangle Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \triangle Y_{t-j} + \varepsilon_t,\ t \geq 1, \tag{8}$$

where $\varepsilon_t \sim$ *i.i.d.* $N_k[0, \Omega]$, *with* Ω *non-singular.*

Due to Assumption 3, there is no vector of constants in this model. Moreover, note that $\varepsilon_t = C_0 U_t$, so that $\Omega = C_0 C_0'$.

Similar to (3), model (8) can be written more conveniently as

$$\triangle Y_t = \alpha\beta' Y_{t-1} + \Gamma X_t + C_0 U_t,\ t \geq 1,$$

and replacing ε_t by $C_0 U_t$ in (3), the time-varying VECM(p) model becomes

$$\triangle Y_t = \alpha\xi' Y_{t-1}^{(m)} + \Gamma X_t + C_0 U_t,$$

where under the null hypothesis,

$$\xi = \begin{pmatrix} \beta \\ O_{m.k\times r} \end{pmatrix}. \tag{9}$$

Finally, to exclude the case that $\beta' Y_{t-1}$ and X_t are multicollinear we need to assume that

Assumption 5. $\mathrm{Var}\left[\left(Y_{t-1}'\beta, X_t'\right)'\right]$ *is nonsingular.*

3.3 Asymptotic Null Distribution

The asymptotic results in the standard cointegration case hinge on the following well-known convergence results. Under Assumptions 1-2,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} U_t Y_{t-1}' \xrightarrow{d} \int_0^1 (dW)\, W' C(1)'. \\
&\frac{1}{T}\sum_{t=1}^{T} (\triangle Y_{t-\ell})\, Y_{t-1}' \xrightarrow{d} C(1)\left(\int_0^1 (dW)\, W'\right) C(1)' + M_\ell,\ \ \ell \geq 0, \\
&\frac{1}{T^2}\sum_{t=1}^{T} Y_t Y_{t-1}' \xrightarrow{d} C(1)\left(\int_0^1 W(x)W'(x)dx\right) C(1)',
\end{aligned}$$

where W is a k-variate standard Wiener process, and the M_ℓ's are nonrandom $k \times k$ matrices. See Phillips and Durlauf (1986) and Phillips (1988).

We need to generalize these results to the case where Y_{t-1} is replaced by $Y_{t-1}^{(m)}$:

Lemma 2. *Under Assumptions 1-2,*

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} U_t \left(Y_{t-1}^{(m)}\right)' \xrightarrow{d} \int_0^1 (dW)\, \widetilde{W}_m' \left(C(1)' \otimes I_{m+1}\right), \\
&\frac{1}{T}\sum_{t=1}^{T} (\triangle Y_{t-\ell}) \left(Y_{t-1}^{(m)}\right)' \xrightarrow{d} C(1)\int_0^1 (dW)\, \widetilde{W}_m' \left(C(1)' \otimes I_{m+1}\right) \\
&\qquad + M_\ell^*,\ \ \ell \geq 0, \qquad\qquad (10) \\
&\frac{1}{T^2}\sum_{t=1}^{T} \left(Y_{t-1}^{(m)}\right)\left(Y_{t-1}^{(m)}\right)' \xrightarrow{d} \\
&\qquad \left(C(1) \otimes I_{m+1}\right)\int_0^1 \widetilde{W}_m(x)\widetilde{W}_m'(x)dx \left(C(1)' \otimes I_{m+1}\right),
\end{aligned}$$

where W is a k-variate standard Wiener process,

$$\widetilde{W}_m(x) = \left(W'(x), \sqrt{2}\cos(\pi x)W'(x), ..., \sqrt{2}\cos(m\pi x)W'(x)\right)',$$

and the M_ℓ^*'s are $k \times k(m+1)$ non-random matrices.

Proof. See the Appendix.

The result (10) implies that $(1/T)\sum_{t=1}^{T}(\triangle Y_{t-\ell})\left(Y_{t-1}^{(m)}\right)' = O_p(1)$. The latter is what is needed for our analysis. Therefore, the question of how the matrices M_ℓ^* look like is not relevant.

Note that

$$\int_0^1 (dW)\widetilde{W}_m' = \left(\int_0^1 (dW(x))W'(x), \sqrt{2}\int_0^1 \cos(1\pi x)\,dW(x)W'(x),\right.$$
$$\left.\sqrt{2}\int_0^1 \cos(2\pi x)\,dW(x)W'(x), ..., \sqrt{2}\int_0^1 \cos(m\pi x)\,dW(x)W'(x)\right).$$

In Bierens and Martins (2009) we define the proper meaning of the random matrices $\int_0^1 \cos(\ell\pi x)\,dW(x)W'(x)$ for $\ell = 1, 2, 3, ...$ In particular, if $W(x)$ is univariate then

$$\int_0^1 \cos(\ell\pi x)\,W(x)dW(x) = \frac{(-1)^\ell}{2}W^2(1) + \frac{\ell\pi}{2}\int_0^1 \sin(\ell\pi x)\,W^2(x)dx.$$

Using Lemma 2 (together with rather long list of auxiliary lemmas), the following results can be shown.

Lemma 3. *Under Assumptions 1-5 the r largest ordered solutions $\widehat{\lambda}_{m,1} \geq \widehat{\lambda}_{m,2} \geq ... \geq \widehat{\lambda}_{m,r}$ of the generalized eigenvalue problem (5) converge in probability to constants $1 > \overline{\lambda}_1 \geq ... \geq \overline{\lambda}_r > 0$, which do not depend on m. Thus, these probability limits are the same as in the standard TI cointegration case.*

Proof. See the Appendix.

As is well-known (see Johansen, 1988), in the standard TI cointegration case $m = 0$ and under Assumptions 1-5, $T\left(\widehat{\lambda}_{0,r+1}, \widehat{\lambda}_{0,r+2}, ..., \widehat{\lambda}_{0,k}\right)'$ converges in distribution to the vector of ordered solutions $\rho_{0,1} \geq \rho_{0,2} \geq .... \geq \rho_{0,k-r}$ of

$$\det\left[\rho\int_0^1 W_{k-r}(x)W_{k-r}'(x)dx - \int_0^1 W_{k-r}dW_{k-r}'\int_0^1 (dW_{k-r})\,W_{k-r}'\right] = 0, \tag{11}$$

where

$$W_{k-r}(x) = \left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' C_0 W(x)$$

is a $k-r$ variate standard Wiener process.[1] This result is based on the fact that one can choose an orthogonal complement $\beta_\perp$ of β such that

$$\frac{1}{T}\beta_\perp' S_{11,T}^{(0)}\beta_\perp \xrightarrow{d} \int_0^1 W_{k-r}(x)W_{k-r}'(x)dx,$$

$$\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' C_0 S_{01,T}^{(0)}\beta_\perp \xrightarrow{d} \int_0^1 (dW_{k-r})\,W_{k-r}'.$$

One would therefore expect that this result can be generalized to the TV cointegration case simply by replacing $W_{k-r}(x)$ in (11) with

$$\begin{aligned}\widetilde{W}_{k-r,m}(x) &= \left(W_{k-r}'(x), \sqrt{2}\cos(\pi x)W_{k-r}'(x), ..., \sqrt{2}\cos(m\pi x)W_{k-r}'(x)\right)' \\ &= \left(\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' C_0 \otimes I_{m+1}\right)\widetilde{W}_m(x) \end{aligned}\tag{12}$$

while leaving dW_{k-r} as is. However, that is not the case!

Lemma 4. *Under Assumptions 1-5,*

$$T\left(\widehat{\lambda}_{m,r+1}, \widehat{\lambda}_{m,r+2}, ..., \widehat{\lambda}_{m,k}\right)' \xrightarrow{d} \left(\rho_{m,1}, ..., \rho_{m,k-r}\right)',$$

where $\rho_{m,1} \geq \rho_{m,2} \geq .... \geq \rho_{m,k-r}$ *are the* $k-r$ *largest solutions of the generalized eigenvalue problem*

$$\begin{aligned}0 &= \det\left[\rho\begin{pmatrix}\int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}_{k-r,m}'(x)dx & O_{(k-r)(m+1),r.m} \\ O_{r.m,(k-r)(m+1)} & I_{r.m}\end{pmatrix}\right. \\ &\quad \left. - \begin{pmatrix}\int_0^1 \widetilde{W}_{k-r,m}dW_{k-r}' \\ V\end{pmatrix}\left(\int_0^1 (dW_{k-r})\,\widetilde{W}_{k-r,m}', V'\right)\right],\end{aligned}\tag{13}$$

[1]Because $\alpha_\perp' C_0 C_0' \alpha_\perp = \alpha_\perp'\Omega\alpha_\perp$.

with V an $r.m\times(k-r)$ random matrix with i.i.d. $N[0,1]$ elements. Moreover, V is independent of W_{k-r} and $\widetilde{W}_{k-r,m}$.

Proof. See the Appendix.

The reason for this unexpected result is the following. Under the null hypothesis (9), any orthogonal complement of the $(m+1)k\times r$ matrix ξ of TV cointegrating vectors is an $(m+1)k\times(k(m+1)-r)$ matrix of the form

$$\xi_{\perp}=\left(\beta_{\perp}\otimes I_{m+1},\left(\begin{array}{c}O_{k,m.r}\\ \beta\otimes I_m\end{array}\right)\right)\times R,$$

where the $k\times(k-r)$ matrix $\beta_{\perp}$ is an orthogonal complements of β and R is a nonsingular $(k(m+1)-r)\times(k(m+1)-r)$ matrix, possibly depending on T. We need to choose R such that $\frac{1}{T}\xi_{\perp}'S_{11,T}^{(m)}\xi_{\perp}$ converges in distribution to a nonsingular matrix.[2] A suitable version of $\xi_{\perp}$ that delivers this result is

$$\xi_{\perp,T}=\left(\beta_{\perp}\otimes I_{m+1},\left(\begin{array}{c}O_{k,m.r}\\ \sqrt{T}\left(\beta\Sigma_{\beta\beta}^{-1/2}\otimes I_m\right)\end{array}\right)\right),\tag{14}$$

where

$$\Sigma_{\beta\beta}=p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\beta'Y_{t-1}Y_{t-1}'\beta.$$

Then $\frac{1}{T}\xi_{\perp,T}'S_{11,T}^{(m)}\xi_{\perp,T}$ converges in distribution to the first matrix in (13). The matrix V involved is now due to

$$\left(\alpha_{\perp}'\Omega\alpha_{\perp}\right)^{-1/2}\alpha_{\perp}'C_0S_{01,T}^{(m)}\left(\begin{array}{c}O_{k,m.r}\\ \sqrt{T}\left(\beta\Sigma_{\beta\beta}^{-1/2}\otimes I_m\right)\end{array}\right)\overset{d}{\to}V'.$$

Under standard cointegration, the ML estimator $\widehat{\beta}$ of β, normalized as $\widetilde{\beta}=\widehat{\beta}\left(\beta'\widehat{\beta}\right)^{-1}\beta'\beta$, satisfies

$$T\left(\widetilde{\beta}-\beta\right)\overset{d}{\to}\beta_{\perp}\left(\int_0^1W_{k-r}W_{k-r}'\right)^{-1}\left(\int_0^1W_{k-r}d\underline{W}_{\alpha}'\right)\left(\alpha'\Omega^{-1}\alpha\right)^{-1/2},$$

[2] So that Lemma 2 in Andersson et al. (1983) can be applied.

where $\underline{W}_\alpha$ is an r-variate standard Wiener process which is independent of W_{k-r}. See Johansen (1988). In our case, however, the corresponding result is again quite different:

Lemma 5 *Let $\widehat{\xi}$ be the ML estimator of ξ,*[3] *and denote $\widetilde{\xi} = \widehat{\xi}\left(\xi'\widehat{\xi}\right)^{-1}\xi'\xi$. Let $\xi_{\perp,T}$ be the orthogonal complement of ξ defined by* (14). *We can always write $\widetilde{\xi} - \xi = \xi_{\perp,T}U_{m,T}$, where*

$$U_{m,T} = \left(\xi'_{\perp,T}\xi_{\perp,T}\right)^{-1}\left(\xi'_{\perp,T}\widehat{\xi}\right)\left(\xi'\widehat{\xi}\right)^{-1}(\xi'\xi).$$

Under Assumptions 1-5,

$$T.U_{m,T} \stackrel{d}{\to} \begin{pmatrix} \left(\int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx\right)^{-1}\int_0^1 \widetilde{W}_{k-r,m}d\underline{W}'_\alpha \\ \underline{V}_\alpha \end{pmatrix} \times \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}, \tag{15}$$

where $\underline{W}_\alpha$ is an r-variate standard Wiener process, $\underline{V}_\alpha$ is a $k.m \times r$ matrix with independent $N[0,1]$ distributed elements, and $\underline{V}_\alpha$, $\underline{W}_\alpha$ and $\widetilde{W}_{k-r,m}$ are independent. Consequently,

$$\begin{pmatrix} T.I_k & O_{k,k.m} \\ O_{k.m,k} & \sqrt{T}I_{k.m} \end{pmatrix}\left(\widetilde{\xi} - \xi\right)$$
$$\stackrel{d}{\to} \begin{pmatrix} (\beta_\perp, O_{k,k.m})\left(\int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx\right)^{-1}\int_0^1 \widetilde{W}_{k-r,m}d\underline{W}'_\alpha \\ \left(\beta\Sigma_{\beta\beta}^{-1/2} \otimes I_m\right)\underline{V}_\alpha \end{pmatrix} \times \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}.$$

Proof. See the Appendix.

The test for standard cointegration is based on a simple hypothesis, $\xi' = \left(\beta', O_{k.m,r}\right)$. The chi-square asymptotic distribution of the likelihood ratio statistic, derived in the Appendix, follows from the previous four lemmas and the Taylor expansion around the MLE of a function of the type

$$f_{m,T}(x) = T.\ln\left(\frac{\det\left(x'\left(S_{11,T}^{(m)} - S_{10,T}^{(m)}S_{00,T}^{-1}S_{01,T}^{(m)}\right)x\right)}{\det\left(x'S_{11,T}^{(m)}x\right)}\right).$$

[3] Recall that under the null hypothesis, $\xi' = (\beta', O_{r,k.m})$.

Then we simply apply the Taylor expansion, derived in Johansen (1988), to the decomposition of the LR statistic

$$f_{m,T}(\widetilde{\xi}) - f_{0,T}(\widetilde{\beta}) = \left(f_{m,T}(\widetilde{\xi}) - f_{0,T}(\beta)\right) - \left(f_{0,T}(\widetilde{\beta}) - f_{0,T}(\beta)\right),$$

where similar to $\widetilde{\xi}$ defined in Lemma 5, $\widetilde{\beta} = \widehat{\beta}\left(\beta'\widehat{\beta}\right)^{-1}\beta'\beta$. It follows then from Lemma 5 that under the null hypothesis,

$$\begin{aligned} & f_{m,T}(\widehat{\xi}) - f_{0,T}(\beta) \xrightarrow{d} \operatorname{trace}\left(\underline{V}'_\alpha \underline{V}_\alpha\right) \\ & + \operatorname{trace}\left[\left(\int_0^1 d\underline{W}_\alpha \widetilde{W}'_{k-r,m}\right)\left(\int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx\right)^{-1}\right. \\ & \qquad \left. \times \left(\int_0^1 \widetilde{W}_{k-r,m} d\underline{W}'_\alpha\right)\right] \sim \chi^2_{r.m.r} + \chi^2_{r(m+1)(k-r)}, \end{aligned}$$

where the two chi-square distributions are independent, whereas it has been shown by Johansen (1988) that

$$\begin{aligned} & T\left(\widehat{f}_0\left(\widetilde{\beta}\right) - \widehat{f}_0\left(\beta\right)\right) \\ & \xrightarrow{d} \operatorname{trace}\left[\left(\int_0^1 d\underline{W}_\alpha W'_{k-r}\right)\left(\int_0^1 W_{k-r}(x)W'_{k-r}(x)dx\right)^{-1}\right. \\ & \qquad \left. \times \left(\int_0^1 W_{k-r} d\underline{W}'_\alpha\right)\right] \sim \chi^2_{r(k-r)}. \end{aligned}$$

It follows now straightforwardly that:

Theorem 1 *Given $m \geq 1$ and $r \geq 1$, under the null hypothesis of standard cointegration the LR statistic LR_T^{tvc} defined in* (6) *is asymptotically χ^2_{mkr} distributed.*

3.4 Empirical Size

To check how close the asymptotic critical values based on the χ^2 distribution are to the ones based on the small sample null distribution, we have applied our test to 10,000 replications of the bivariate cointegrated vector time series process $Y_t = (Y_{1,t}, Y_{2,t})'$, where $Y_{1,t} = Y_{2,t} + U_{1,t}$, $Y_{2,t} = Y_{2,t-1} + U_{2,t}$ with $U_t = (U_{1,t}, U_{2,t})'$ drawn independently from the bivariate standard normal distribution, for various values of T and m. The numerical results are given in Bierens and Martins (2009).

For large T and small m the right tail of the distribution is very well approximated by the asymptotic one. For smaller T the test suffers from size distortion. For example, for $T = 100$ and 5% asymptotic size the nominal size is 3% for $m = 1$, 2% for $m = 3$ and 1.3% for $m = 5$. Thus, by using the asymptotic critical values the test tends to over-reject the correct null hypothesis of standard cointegration. As expected, for $T = 500$ the empirical and the asymptotic distributions almost coincide.

4 The LR Test under the Alternative of TV Cointegration

4.1 The Data Generating Process under TV Cointegration

A time-varying cointegrated data-generating process Y_t with VECM(p) representation (1) can be constructed, for example, as follows. Let

$$Y_t = AZ_t = \alpha Z_{1,t} + \gamma Z_{2,t},$$

where $A = (\alpha, \gamma)$ is a nonsingular $k \times k$ matrix, with α the matrix of the first r columns of A and γ the matrix of the remaining $k - r$ columns of A. In this expression $Z_{1,t} \in \mathbb{R}^r$ and $Z_{2,t} \in \mathbb{R}^{k-r}$ are $I(1)$ processes generated by

$$\begin{aligned} Z_{1,t} &= \sum_{j=1}^{p} D_j Z_{1,t-j} + B_2(t/T) Z_{2,t-1} + \sum_{j=1}^{p-1} C_{12,j} \Delta Z_{2,t-j} + U_{1,t}, \quad (16) \\ \Delta Z_{2,t} &= \sum_{j=1}^{p-1} C_{22,j} \Delta Z_{2,t-j} + U_{2,t}, \end{aligned}$$

respectively, where

Assumption 6. *$U_t = (U'_{1,t}, U'_{2,t})' \sim$ i.i.d. $N_k[0, V_u]$. The matrix valued lag polynomials $D(L) = I_r - \sum_{j=1}^{p} D_j L^j$ and $C_{22}(L) = I_{k-r} - \sum_{j=1}^{p-1} C_{22,j} L^j$ are invertible, with inverses $D(L)^{-1} = \sum_{j=0}^{\infty} \Pi_j L^j$ and $C_{22}(L)^{-1} = \sum_{j=0}^{\infty} \Gamma_j L^j$ satisfying $\Pi_j \to O$, $\Gamma_j \to O$ exponentially as $j \to \infty$. The elements of $B_2(\tau)$*

are continuously differentiable function on an open interval containing $[0,1]$ *with bounded derivatives, and*

$$B_2(\tau) = B_2(0) \ \textit{for}\ \tau < 0, B_2(\tau) = B_2(1) \ \textit{for}\ \tau > 1.$$

Note that the nonstationarity of $Z_{1,t}$ is due to the dependence of $Z_{1,t}$ on $B_2(t/T)\, Z_{2,t-1}$.

As is well-known, we can rewrite model (16) as

$$\begin{aligned}\Delta Z_{1,t} &= B_1 Z_{1,t-1} + B_2(t/T)\, Z_{2,t-1} \\ &\quad + \sum_{j=1}^{p-1} C_{11,j}\Delta Z_{1,t-j} + \sum_{j=1}^{p-1} C_{12,j}\Delta Z_{2,t-j} + U_{1,t}, \qquad (17)\end{aligned}$$

where $B_1 = \sum_{j=1}^{p} D_j - I_r$ is nonsingular, [4] hence

$$\Delta Z_t = \begin{pmatrix} B_1 & B_2(t/T) \\ O_{k-r,r} & O_{k-r,k-r} \end{pmatrix} Z_{t-1} + \sum_{j=1}^{p-1} C_j \Delta Z_{t-j} + U_t$$

where

$$C_j = \begin{pmatrix} C_{11,j} & C_{12,j} \\ O_{k-r,r} & C_{22,j} \end{pmatrix}.$$

Thus,

$$\begin{aligned}\Delta Y_t &= A\Delta Z_t \\ &= A \begin{pmatrix} B_1 & B_2(t/T) \\ O_{k-r,r} & O_{k-r,k-r} \end{pmatrix} A^{-1} Y_{t-1} + \sum_{j=1}^{p-1} A C_j A^{-1} \Delta Y_{t-j} + A U_t \\ &= \alpha \beta_t' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + A U_t, \qquad (18)\end{aligned}$$

for example, where

$$\beta_t' = (B_1, B_2(t/T))\, A^{-1},\ \Gamma_j = A C_j A^{-1}. \qquad (19)$$

[4] Because the invertibility of $D(L)$ implies that all the roots of the polynomial $\det\left(I_r - \sum_{j=1}^{p} D_j x^j\right)$ lie outside the complex unit circle.

A more general TV model can be formulated by allowing B_1 to be a function of t/T as well, and by including lagged $\Delta Z_{1,t}$'s in the equation for $\Delta Z_{2,t}$. However, that will make the power analysis too complicated.

Under Assumption 6 the process Y_t is TV cointegrated, in the sense that with β_t define in (19),

$$\beta_t' Y_{t-1} = R_t + O_p(1),$$

where R_t is a strictly stationary zero-mean Gaussian process, and the $O_p(1)$ term is uniform in $t = 1, 2, ..., T$. This follows from the result (21) in the following lemma.

Lemma 6. *Under Assumption 6 we can write*

$$\Delta Z_{1,t} = \sum_{j=0}^{t-1} \Pi_j \left(B_2\left((t-j)/T\right) - B_2(0)\right) \Delta Z_{2,t-1-j} + V_t + O_p\left(1/\sqrt{T}\right), \quad (20)$$

uniformly in $t = 1, ..., T$, *where* V_t *is a strictly stationary zero-mean Gaussian process. Moreover, denote* $\sum_{j=0}^{\infty} Q_j L^j = D(L)^{-1} C_{11}(L)$, *where* $C_{11}(L) = I_r - \sum_{j=1}^{p-1} C_{11,j} L^j$. *Then*

$$\begin{aligned} B_1 Z_{1,t-1} + B_2(t/T) Z_{2,t-1} &= \sum_{j=0}^{t-1} Q_j \left(B_2\left((t-j)/T\right) - B_2(0)\right) \Delta Z_{2,t-1-j} \\ &\quad + R_t + O_p\left(1/\sqrt{T}\right) \end{aligned}$$

uniformly in $t = 1, ..., T$, *where* R_t *is a strictly stationary zero-mean Gaussian process. Consequently,*

$$B_1 Z_{1,t-1} + B_2(t/T) Z_{2,t-1} = R_t + O_p(1) \quad (21)$$

uniformly in $t = 1, ..., T$.

Proof. See the Appendix.

4.2 Power of the LR test

To study the power of our test, we will adopt the VECM(p) model (18) with β_t defined by (2) as the data generating process. Moreover, to keep the power analysis tractable we will focus on the case $p=1,\ k=2,\ r=1,\ V_u = I_2,\ A = I_2$. Thus,

$$Y_t = Z_t = (Z_{1,t}, Z_{2,t})',$$

where $Z_{1,t} \in \mathbb{R}$ and $Z_{2,t} \in \mathbb{R}$ are assumed to be generated by

$$\begin{aligned} \Delta Z_{1,t} &= b_1 Z_{1,t-1} + b_2 (t/T) Z_{2,t-1} + U_{1,t}, \\ \Delta Z_{2,t} &= U_{2,t}, \\ U_t &= (U_{1,t}, U_{2,t})' \sim \text{ i.i.d. } N_2 [0, I_2]. \end{aligned}$$

Next, suppose that for some $m > 0$,

$$b_1^{-1} b_2 (t/T) = \sum_{j=0}^{m} \rho_j P_{j,T}(t),\ \rho' = (\rho_0, \rho_1, ..., \rho_m).$$

Then

$$(b_1, b_2 (t/T)) = b_1 \sum_{j=0}^{m} \varsigma_j' P_{j,T}(t),$$

where $\varsigma_0' = (1, \rho_0)$ and $\varsigma_j' = (0, \rho_j)$ for $j \geq 1$. Hence,

$$\begin{aligned} \Delta Z_{1,t} &= b_1 \left(Z_{1,t-1} + \sum_{j=0}^{m} \rho_j P_{j,T}(t) Z_{2,t-1} \right) + U_{1,t} \\ &= b_1 \sum_{j=0}^{m} \varsigma_j' P_{j,T}(t) Z_{t-1} + U_{1,t} = b_1 \varsigma' Z_{t-1}^{(m)} + U_{1,t}, \\ \Delta Z_{2,t} &= U_{2,t}, \end{aligned}$$

where

$$\varsigma' = (1, \rho_0, 0, \rho_1, 0, \rho_2, \ldots, 0, \rho_m) \tag{22}$$

and

$$Z_{t-1}^{(m)} = \begin{pmatrix} Z_{1,t-1}^{(m)} \\ Z_{2,t-1}^{(m)} \end{pmatrix} = Z_{t-1} \otimes \widehat{p}_m (t/T),$$

with

$$Z_{i,t-1}^{(m)} = \left(Z_{i,t-1}', P_{1,T}(t) Z_{i,t-1}', P_{2,T}(t) Z_{i,t-1}', ..., P_{m,T}(t) Z_{i,t-1} \right)',\ i = 1, 2,$$

and

$$\widehat{p}_m (t/T) = (1, P_{1,T}(t), ..., P_{m,T}(t))'.$$

We can now write the model in VECM(1) form as

$$\Delta Z_t = \delta\varsigma' Z_{t-1}^{(m)} + U_t$$

where

$$\delta = \begin{pmatrix} b_1 \\ 0 \end{pmatrix}.$$

In the sequel we will refer to this model, together with the applicable parts of Assumptions 1-2, as $H_1^{(m)}(p=1)$.

Under $H_1^{(m)}(p=1)$ the matrices $S_{00,T}$, $S_{11,T}^{(m)}$ and $S_{01,T}^{(m)}$ become

$$\begin{aligned} S_{00,T} &= \frac{1}{T}\sum_{t=1}^{T} \triangle Z_t \triangle Z_t' \\ S_{11,T}^{(m)} &= \frac{1}{T}\sum_{t=1}^{T} Z_{t-1}^{(m)} Z_{t-1}^{(m)\prime} \\ S_{01,T}^{(m)} &= \frac{1}{T}\sum_{t=1}^{T} \triangle Z_t Z_{t-1}^{(m)\prime} \end{aligned}$$

respectively. The maximum log-likelihood in the standard case with $r=1$ is

$$\begin{aligned} \widehat{l}_T(1,0) &= -\frac{1}{2}T.\ln\left(1-\max_{\beta}\frac{\beta' S_{10,T}^{(0)} S_{00,T}^{-1} S_{01,T}^{(0)}\beta}{\beta' S_{11,T}^{(0)}\beta}\right) \\ &\quad -\frac{1}{2}T.\ln\left(\det\left(S_{00,T}\right)\right) - T.k\ln\left(\sqrt{2\pi}\right) - \frac{1}{2}kT \end{aligned}$$

and in the TV case

$$\begin{aligned} \widehat{l}_T(1,m) &= -\frac{1}{2}T.\ln\left(1-\max_{\xi}\frac{\xi' S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)}\xi}{\xi' S_{11,T}^{(m)}\xi}\right) \\ &\quad -\frac{1}{2}T.\ln\left(\det\left(S_{00,T}\right)\right) - T.k\ln\left(\sqrt{2\pi}\right) - \frac{1}{2}kT. \end{aligned}$$

Thus,

$$p\lim_{T\to\infty} T^{-1}\left(\widehat{l}_T(1,m) - \widehat{l}_T(1,0)\right) > 0 \tag{23}$$

if $p\lim_{T\to\infty}\widehat{\lambda}_{\max}^{(0)} < p\lim_{T\to\infty}\widehat{\lambda}_{\max}^{(m)}$, where

$$\widehat{\lambda}_{\max}^{(0)} = \max_{\beta}\frac{\beta' S_{10,T}^{(0)} S_{00,T}^{-1} S_{01,T}^{(0)}\beta}{\beta' S_{11,T}^{(0)}\beta},\ \widehat{\lambda}_{\max}^{(m)} = \max_{\xi}\frac{\xi' S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)}\xi}{\xi' S_{11,T}^{(m)}\xi}.$$

Note that $\lambda_{\max}^{(m)}$ is the maximal solution of (5). Because

$$p \lim_{T\to\infty} \widehat{\lambda}_{\max}^{(m)} = p \lim_{T\to\infty} \frac{\varsigma' S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \varsigma}{\varsigma' S_{11,T}^{(m)} \varsigma}$$

where ς is defined by (22), the consistency of our test against the alternative $H_1^{(m)}(p = 1)$ follows from the following theorem.

Theorem 2. *Under* $H_1^{(m)}(p = 1)$,

$$p \lim_{T\to\infty} \frac{\varsigma' S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \varsigma}{\varsigma' S_{11,T}^{(m)} \varsigma} \in (0,1)\,,\ p \lim_{T\to\infty} \frac{\beta' S_{10,T}^{(0)} S_{00,T}^{-1} S_{01,T}^{(0)} \beta}{\beta' S_{11,T}^{(0)} \beta} = 0$$

for all nonzero vectors $\beta \in \mathbb{R}^2$, *hence* (23) *holds.*

The proof of Theorem 2 is not too difficult but tedious and lengthy. This proof is therefore given in Bierens and Martins (2009). It is our conjecture that Theorem 2 carries over to more general alternatives, but verifying this analytically proved to be too tedious an exercise. The same applies to the local power of the test. It is our conjecture that along the lines of the proof of Theorem 2 it can be shown that the test has nontrivial local power.

The power of our test depends on the choice of the Chebyshev polynomial order m. The optimal choice for m can be compared to the optimal choice of the order of an autoregressive process. As to the latter, researchers usually employ the Hannan-Quinn (1979) or Schwarz (1978) information criteria. The results in this section suggest that these information criteria can also be used to estimate m consistently if m is finite, but a formal proof is beyond the scope of this paper.

4.3 Empirical Power

The assumption that the time varying cointegrating vector can be exactly represented by a fixed number of Chebyshev polynomials is quite restrictive. Therefore, in this subsection we check via a limited Monte Carlo study how the test performs if this assumption is not true.

The data generating process we have used is

$$\begin{aligned} Z_{1,t} &= 0.5\left(Z_{1,t-1} - \left(1 - \omega + \omega f\left(t/T\right)\right) Z_{2,t-1}\right) + 0.25 \triangle Z_{1,t-1} + U_{1,t}, \\ \Delta Z_{2,t} &= U_{2,t}, t = 1, ..., T, \end{aligned}$$

where $Z_{1,t} \in \mathbb{R}$, $Z_{2,t} \in \mathbb{R}$, the error vectors $(U_{1,t}, U_{2,t})'$ are independently $N_2[0, I_2]$ distributed, and $\omega \in [0,1]$. The number of replications is $10,000$. For the function f we have chosen the following S-shaped function on $[0,1]$:

$$f(x) = 12\int_0^x y(1-y)\,dy - 1 = 6x^2 - 4x^3 - 1.$$

Note that $f(t/T)$ cannot be represented by a fixed number m of Chebyshev polynomials.

The results are presented in Table 1, for $m = 1, 5$ and $T = 100, 200$. The number of replications is $10,000$. In order to check the size and to mimic local alternatives we have conducted the power simulations for $\omega \in \{0, 0.01, 0.05, 0.1, 0.2, 0.5, 1\}$. The case $\omega = 0$ corresponds to TI cointegration, with cointegrating vector $\beta = (1, -1)'$, whereas for $\omega > 0$ we have time-varying cointegration with cointegrating vector β_t moving smoothly from $\beta_0 = (1, 2\omega - 1)'$ to $\beta_T = (1, -1)'$.

In Table 1, α_{asy} indicates the asymptotic size, so that the rejection rates involved are with respect to the asymptotic critical values, whereas α_{real} is the empirical size, so that rejection rates involved are with respect to the empirical critical values. See Bierens and Martins (2009) for the latter.

As expected, our test suffers from size distortion in small samples if the asymptotic critical values are used. This size distortion increases with m. On the other hand, the size distortion is modest if the empirical critical values are used. In view of the results for $\omega = 0.01$ and $\omega = 0.05$ our test seems to have non-trivial local power. Moreover, note that in general the power is not affected much by the choice of m, despite the fact that $f(t/T)$ cannot be represented by a fixed number of Chebyshev polynomials.

Finally, we have also analyzed the size and power properties of the two tests proposed by Park and Hahn (1999), for the same cases as in Table 1. The results are presented in Bierens and Martins (2009). Surprisingly, both Park-Hahn tests suffer from extreme size distortion. Therefore, it is difficult to compare the actual power of these tests with the power of our test.

Table 1: Power of the LR test

$T=100, m=1$	$\alpha_{asy}=0.10$	$\alpha_{asy}=0.05$	$\alpha_{asy}=0.01$	$\alpha_{\text{real}}=0.05$
$\omega=0$	0.163	0.093	0.026	0.063
$\omega=0.01$	0.164	0.095	0.027	0.065
$\omega=0.05$	0.236	0.152	0.052	0.110
$\omega=0.1$	0.396	0.290	0.141	0.227
$\omega=0.2$	0.692	0.600	0.402	0.532
$\omega=0.5$	0.976	0.958	0.896	0.941
$\omega=1$	0.999	0.999	0.997	0.998
$T=200, m=1$	$\alpha_{asy}=0.10$	$\alpha_{asy}=0.05$	$\alpha_{asy}=0.01$	$\alpha_{\text{real}}=0.05$
$\omega=0$	0.127	0.071	0.015	0.057
$\omega=0.01$	0.140	0.079	0.019	0.067
$\omega=0.05$	0.374	0.278	0.133	0.253
$\omega=0.1$	0.693	0.603	0.430	0.577
$\omega=0.2$	0.947	0.920	0.829	0.909
$\omega=0.5$	0.999	0.999	0.998	0.999
$\omega=1$	1.000	1.000	1.000	1.000

$T=100, m=5$	$\alpha_{asy}=0.10$	$\alpha_{asy}=0.05$	$\alpha_{asy}=0.01$	$\alpha_{\text{real}}=0.05$
$\omega=0$	0.366	0.251	0.092	0.105
$\omega=0.01$	0.369	0.252	0.093	0.106
$\omega=0.05$	0.405	0.284	0.112	0.127
$\omega=0.1$	0.491	0.373	0.178	0.196
$\omega=0.2$	0.683	0.574	0.360	0.382
$\omega=0.5$	0.950	0.914	0.800	0.818
$\omega=1$	0.998	0.997	0.985	0.988
$T=200, m=5$	$\alpha_{asy}=0.10$	$\alpha_{asy}=0.05$	$\alpha_{asy}=0.01$	$\alpha_{\text{real}}=0.05$
$\omega=0$	0.202	0.126	0.036	0.076
$\omega=0.01$	0.208	0.130	0.038	0.079
$\omega=0.05$	0.347	0.237	0.096	0.162
$\omega=0.1$	0.594	0.487	0.299	0.403
$\omega=0.2$	0.884	0.825	0.688	0.772
$\omega=0.5$	0.998	0.996	0.991	0.995
$\omega=1$	1.000	1.000	1.000	1.000

5 The Drift Case

Assumptions 1-2 imply that ΔY_t and $\beta' Y_t$ are zero-mean stationary processes. However, for most cointegrated macroeconomic time series, ΔY_t and $\beta' Y_t$ are nonzero-mean stationary processes, which correspond to the following modification of Assumption 1:

Assumption 1*. *Assume* $\triangle Y_t = C(L)(U_t + \mu) = \sum_{j=0}^{\infty} C_j (U_{t-j} + \mu)$, *where* $\mu \neq 0$ *is a vector of imbedded drift parameters, and* U_t *and* $C(L)$ *are the same as in Assumption 1.*

Then similar to (7) we can write

$$Y_t = C(1)\sum_{j=1}^{t} U_j + C(1)\mu.t + V_t + Y_0 - V_0.$$

Under Assumption 2,

$$\beta' Y_t = \beta' V_t + \beta' (Y_0 - V_0).$$

Thus, Assumption 2 can be adopted without modifications, but Assumption 3 needs to be dropped as otherwise $\beta' (Y_0 - V_0) = 0$. However, due to the drift we now need to include a vector of intercepts in VECM (8), as in Johansen (1991):

Assumption 4*. *Assume* ΔY_t *has the VECM*(p) *representation*

$$\triangle Y_t = \gamma_0 + \alpha\beta' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \triangle Y_{t-j} + C_0 U_t.$$

Moreover, with $X_t = (\triangle Y_{t-1}, ..., \triangle Y_{t-p+1})'$, Assumption 5 still applies. These modified Assumptions 1-5 will be referred to as "the drift case".

The corresponding time-varying VECM(p) is now

$$\triangle Y_t = \gamma_0 + \alpha\xi' Y_{t-1}^{(m)} + \sum_{j=1}^{p-1} \Gamma_j \triangle Y_{t-j} + C_0 U_t.$$

To re-derive our previous results for this drift case, we need some additional notation. First, let

$$\overline{\mu} = \left(\mu' C_0 \alpha_\perp \left(\alpha_\perp' \Omega \alpha_\perp\right)^{-1} \alpha_\perp' C_0' \mu\right)^{-1/2} \left(\alpha_\perp' \Omega \alpha_\perp\right)^{-1/2} \alpha_\perp' C_0' \mu,$$

which is a vector in $\mathbb{R}^{k-r}$. Note that $\overline{\mu}'\overline{\mu} = 1$ by normalization. Let $\overline{\mu}_\perp$ be an orthogonal complement of $\overline{\mu}$, normalized such that $\overline{\mu}_\perp'\overline{\mu}_\perp = I_{k-r-1}$. Then

Lemma 7. *In the drift case,*

$$\begin{aligned} \left(\overline{\mu}_\perp' \otimes I_{m+1}\right) \left(\beta_\perp' \otimes I_{m+1}\right) \frac{1}{\sqrt{T}} Y_{[xT]}^{(m)} &\Rightarrow p(x) \otimes \underline{W}_{k-r-1}(x) \\ \left(\overline{\mu}' \otimes I_{m+1}\right) \left(\beta_\perp' \otimes I_{m+1}\right) \frac{1}{T} Y_{[x.T]}^{(m)} &\Rightarrow p(x) \otimes x \end{aligned}$$

for $x \in [0,1]$*, where* $p(x) = \left(1, \sqrt{2}\cos(\pi x), ..., \sqrt{2}\cos(m\pi x)\right)'$ *and*

$$\underline{W}_{k-r-1} = \overline{\mu}_\perp' \left(\alpha_\perp' \Omega \alpha_\perp\right)^{-1/2} \alpha_\perp' C_0' W \tag{24}$$

is a $(k-r-1)$*-variate standard Wiener process.*

Next, let

$$M_T = \left(T^{-1/2}\overline{\mu}, \overline{\mu}_\perp\right).$$

Redefine the orthogonal complement $\xi_{\perp,T}$ of ξ in (14) as

$$\xi_{\perp,T} = \left(\left(\beta_\perp \otimes I_{m+1}\right)\left(M_T \otimes I_{m+1}\right), \begin{pmatrix} O_{k,m.r} \\ \sqrt{T}\left(\beta \Sigma_{\beta\beta}^{-1/2} \otimes I_m\right) \end{pmatrix}\right) \tag{25}$$

and redefine $\widetilde{W}_{k-r,m}$ in (12) as

$$\begin{aligned} \widetilde{W}_{k-r,m}(x) &= p(x) \otimes \begin{pmatrix} \underline{W}_{k-r-1}(x) \\ x \end{pmatrix} \\ &\quad - \int_0^1 p(y) \otimes \begin{pmatrix} \underline{W}_{k-r-1}(y) \\ y \end{pmatrix} dy, \end{aligned} \tag{26}$$

where $\underline{W}_{k-r-1}$ is defined by (24). Then

Theorem 3. *With* $\xi_{\perp,T}$ *in* (14) *replaced by* (25) *and* $\widetilde{W}_{k-r,m}$ *in* (12) *replaced by* (26) *the results of Lemmas 3-5 and Theorem 1 carry over.*

The proofs of Lemma 7 and Theorem 3 are not too difficult but rather lengthy. These proofs are therefore given in Bierens and Martins (2009).

6 An Empirical Application

The validity of the purchasing power parity (PPP) hypothesis has generated a great deal of controversy, intimately related to the type of method applied. Recently, Falk and Wang (2003) found that the PPP hypothesis holds for some economies but not for all. Their work is based on Caner's (1998) concept of cointegration where the VECM errors follow a stable distribution.

A reason why linear VECM models may be unable to detect long run PPP is the presence of transaction costs in equilibrium models of real exchange rate determination, which imply a nonlinear adjustment process in the PPP relationship. Michael et al. (1997) successfully fit an exponential smooth transition autoregressive model, thus capturing the implied nonlinearities.

We propose an alternative framework where the cointegrating vectors fluctuate over time. We test the TI cointegration hypothesis against TV cointegration for

$$Y_t = \left(\ln S_t^f, \ln P_t^n, \ln P_t^f\right)',$$

where P_t^n and P_t^f are the price indices in the domestic and foreign economies, respectively, and S_t^f is the nominal exchange rate in home currency per unit of the foreign currency. Since the log-prices are unit root with drift processes the tests will be conducted under the "drift-case" assumptions. The time-varying cointegrating relation is $\beta_t' Y_t = e_t$, where the process e_t represents the short run deviations from the PPP due to disturbances in the economic system (real or monetary shocks), and β_t is an unknown vector-valued function of time. Using Chebyshev time polynomials $P_{i,T}(t)$, β_t will be approximated by $\beta_t(m) = \sum_{i=0}^{m} \xi_i P_{i,T}(t)$, where the ξ_i's are the Fourier coefficients.

We use the same data as Falk and Wang (2003), downloaded from the *Journal of Applied Econometrics* data archives web site. The domestic country is the US and the bilateral relationship of study is with Canada, France, Germany, Italy, Japan, and the UK. The data are monthly and cover the period from January 1973 to December 1999, so that the time series involved have length 324. Falk and Wang (2003) find support for the presence of unit roots in all series. By means of the standard Johansen's approach, they find support of the PPP hypothesis at the 5% level in eight of the 12 cases. With one cointegrating vector, Belgium, Denmark, France, Japan, Netherlands, Norway, Spain, and UK are found to have PPP with the US. At the 10% level, Italy and Sweden were added to the list. Therefore, Canada and Germany are the only countries for which the US has not had price parity according to the standard approach.

The asymptotic p-values of our test, for different combinations of the order m of the Chebyshev polynomial expansion and the lag order p, are presented in Bierens and Martins (2009). We find that, regardless of the lag order, the p-values are zero for any m larger than four. Hence, for all cases there is strong evidence of a time varying type of cointegration between international prices and nominal exchange rates. Thus, our results refute Falk and Wang's (2003) findings of standard PPP for all countries except Canada and Germany.

The plots of the time-varying coefficients β_{1t}, β_{2t} and β_{3t} in the cointegrating PPP relation $\beta_t^{\prime} Y_t = \beta_{1t} \ln S_t^f + \beta_{2t} \ln P_t^n + \beta_{3t} \ln P_t^f$ are also presented in Bierens and Martins (2009). The patterns of these parameters suggest that, approximately, $\beta_{2t} + \beta_{3t} = \delta$ for some constant δ. This is related to the symmetry assumption in the standard PPP theory, where $\beta_3 + \beta_2 = 0$. However, δ seems to be positive for Canada and the UK, and negative for the other countries. Moreover, the variation of β_{1t} is minor compared with the variation of β_{2t} and β_{3t}, suggesting that β_{1t} may be constant.

It is unclear from these plots why Falk and Wang (2003) find standard PPP for all countries except Canada and Germany because the patterns of β_{2t} and β_{3t} for Canada and Germany do not look distinct from those of the other countries. On the other hand, Caner's (1998) concept of cointegration employed by Falk and Wang (2003) is fundamentally different from our time-varying cointegration concept, so that one should expect differences in findings as well.

7 Conclusion

In Johansen's standard approach it is assumed that the cointegrating vector is constant over time. This assumption may be restrictive in practice due to changes in taste, technology, or economic policies. We propose a generalization of the standard approach by allowing the cointegrating vectors to be time-varying and we approximate them by using orthogonal Chebyshev time polynomials.

We propose a cointegration model that captures smooth time transitions of the cointegrating vectors - the time varying error correction model - and estimate it by maximum likelihood. To distinguish our model from the time invariant Johansen's specification, we construct a likelihood ratio test for the null hypothesis of standard cointegration. The limiting law appears to be chi-square. To illustrate the practical significance of our approach we apply our test to international prices and nominal exchange rates. We find evidence of time-varying cointegration between these series.

There are issues that merit further research. In particular, the analytical study of the power of the test against local alternatives deserves attention. Moreover, a natural extension of our approach is to include deterministic components such time trends and/or seasonal dummy variables, and to allow for other time varying parameters.

References

Andersson, S.A., H.K. Brons and S.T. Jensen (1983), Distribution of Eigenvalues in Multivariate Statistical Analysis. *Annals of Statistics* **11**, 392-415.

Andrade, P., C. Bruneau and S. Gregoir (2005), Testing for the Cointegration Rank when some Cointegrating Directions are Changing. *Journal of Econometrics* **124**, 269-310.

Andrews, D.W.K., I. Lee and W. Ploberger (1996), Optimal Changepoint Tests for Normal Linear Regression. *Journal of Econometrics* **70**, 9-38.

Bai, J., R.L. Lumsdaine and J.H. Stock (1998), Testing for and Dating Breaks in Multivariate Time Series. *Review of Economic Studies* **65**, 395-432.

Beveridge, S., and C.R. Nelson (1981), A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle. *Journal of Monetary Economics* **7**, 151-174.

Bierens, H.J. (1994), *Topics in Advanced Econometrics: Estimation, Testing and Specification of Cross-Section and Time Series Models.* Cambridge University Press.

Bierens, H.J. (1997), Testing the Unit Root with Drift Hypothesis Against Nonlinear Trend Stationarity, with an Application to the US Price Level and Interest Rate. *Journal of Econometrics* **81**, 29-64.

Bierens, H.J., and L. Martins (2009), "Appendix: Time Varying Cointegration". http://econ.la.psu.edu/~hbierens/TVCOINT_APPENDIX.PDF

Blake, N.S., and T.B. Fomby (1997), Threshold Cointegration. *International Economic Review* **38**, 627-645.

Caner, M. (1998), Tests for Cointegration with Infinite Variance Errors. *Journal of Econometrics* **86**, 155-175.

De Jong, R.M. (2001), Nonlinear Estimation Using Estimated Cointegrating Relations. *Journal of Econometrics* **101**, 109-122.

Engle, R.F., and C.W.J. Granger (1987), Cointegration and Error Correction: Representations, Estimation and Testing. *Econometrica* **55**, 251-276.

Falk, B., and C. Wang (2003), Testing Long Run PPP with Infinite Variance Returns. *Journal of Applied Econometrics* **18**, 471-484.

Granger, C.W.J. (1987), Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics* **48**, 213-218.

Granger, C.W.J., and G. Yoon (2002), Hidden Cointegration. Working Paper, Department of Economics, UCSD.

Hall, S.G., Z. Psaradakis and M. Sola (1997), Cointegration and Changes in Regime: the Japanese Consumption Function. *Journal of Applied Econometrics* **12**, 151-168.

Hamming, R.W. (1973), *Numerical Methods for Scientists and Engineers.* Dover.

Hannan, E.J., and B.G. Quinn (1979), The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society B* **41**, 190-195.

Hansen, B.E. (1992), Tests for Parameter Instability in Regressions with $I(1)$ Processes. *Journal of Business and Economic Statistics* **10**, 321-335.

Hansen, H., and S. Johansen (1999), Some Tests for Parameter Constancy in Cointegrated VAR-Models. *Econometrics Journal* **2**, 306-333.

Hansen, P.R. (2003), Structural Changes in the Cointegrated Vector Autoregressive Model. *Journal of Econometrics* **114**, 261-295.

Hao, K. (1996), Testing for Structural Change in Cointegrated Regression Models: Some Comparisons and Generalizations. *Econometric Reviews* **15**, 401-429.

Harris, D., B. McCabe and S. Leybourne (2002), Stochastic Cointegration: Estimation and Inference. *Journal of Econometrics* **111**, 363-384.

Inoue, A. (1999), Tests of Cointegration Rank with a Trend Break. *Journal of Econometrics* **90**, 215-237.

Johansen, S. (1988), Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* **12**, 231-254.

Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* **59**, 1551-1580.

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

Johansen, S., R. Mosconi and B. Nielsen (2000), Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend. *Econometrics Journal* **3**, 216-249.

Juhl, T., and Z. Xiao (2005), Testing for Cointegration using Partially Linear Models. *Journal of Econometrics* **124**, 363-394.

Kuo, B.S. (1998), Test for Partial Parameter Instability in Regressions with I(1) Processes. *Journal of Econometrics* **86**, 337-368.

Lütkepohl, H., P. Saikkonen and C. Trenkler (2003), Comparison of Tests for the Cointegrating Rank of a VAR Process with a Structural Shift. *Journal of Econometrics* **113**, 201-229.

Lütkepohl, H., T. Terasvirta and J. Wolters (1999), Investigating Stability and Linearity of a German M1 Money Demand Function. *Journal of Applied Econometrics* **14**, 511-525.

Maddala, G.S., and I-M. Kim (1998), *Unit Roots, Cointegration and Structural Change.* Cambridge University Press.

Michael, P., A.R. Nobay and D.A. Peel (1997), Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. *Journal of Political Economy* **105**, 862-879.

Park, J.Y., and S.B. Hahn (1999), Cointegrating Regressions with Time Varying Coefficients. *Econometric Theory* **15**, 664–703.

Phillips, P.C.B. (1988), Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$. *Journal of Multivariate Time Series Analysis* **24**, 252-264.

Phillips, P.C.B., and S.N. Durlauf (1986), Multiple Time Series Regression with Integrated Processes. *Review of Economic Studies* **53**, 473-496.

Quintos, C.E. (1997), Stability Tests in Error Correction Models. *Journal of Econometrics* **82**, 289-315.

Quintos, C.E., and P.C.B. Phillips (1993), Parameter Constancy in Cointegrating Regressions. *Empirical Economics* **18**, 675-706.

Saikkonen, P., and I. Choi (2004), Cointegrating Smooth Transition Regressions. *Econometric Theory* **20**, 301–340.

Schwarz, G. (1978), Estimating the Dimension of a Model. *Annals of Statistic*s **6**, 461-464.

Seo, B. (1998), Tests for Structural Change in Cointegrated Systems. *Econometric Theory* **14**, 222-259.

Terasvirta, T., and A.C. Eliasson (2001), Non-Linear Error Correction and the UK Demand for Broad Money, 1878-1993. *Journal of Applied Econometrics* **16**, 277-288.

Appendix: Proofs

The proofs of Lemmas 2-4 employ the following auxiliary results.

Lemma A.1. *Under Assumptions 1-2,*

$$\frac{1}{T}\sum_{t=1}^{T}(\triangle Y_{t-\ell})\,Y_{t-1}' \xrightarrow{d} C(1)\left(\int_0^1 (dW)\,W'\right)C(1)' + M_\ell,\ \ \ell \geq 0,$$

$$\frac{1}{T}\sum_{t=1}^{[xT]}(\triangle Y_t)\,Y_{t-1}' \xrightarrow{d} C(1)\left(\int_0^x (dW)\,W'\right)C(1)' + xM_0,$$

$$\frac{1}{T}\sum_{t=1}^{T}U_tY_{t-1}' \xrightarrow{d} \int_0^1 (dW)\,W'C(1)',$$

where W is a k-variate standard Wiener process, and the M_ℓ's are non-random $k \times k$ matrices.

Proof. See Phillips and Durlauf (1986) and Phillips (1988).

Lemma A.2. *Let η_t be an arbitrary sequence in $\mathbb{R}^n$, and let $F(x)$ be an arbitrary differentiable function on $[0,1]$, with derivative $f(x)$. Then*

$$\sum_{t=1}^{T}\eta_t F(t/T) = \sum_{t=1}^{T}\eta_t F(1) - \int_0^1 f(x)\left(\sum_{t=1}^{[xT]}\eta_t\right)dx.$$

Proof. See Bierens (1994, Lemma 9.6.3, p. 200).

Lemma A.3. *Under Assumptions 1-2 the following probability limits exist:*

$$\Sigma_{\beta\beta} \ \ = \ \ p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\beta' Y_{t-1}Y_{t-1}'\beta,$$

$$\begin{aligned}\Sigma_{X\beta} &= p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}X_tY'_{t-1}\beta, \\ \Sigma_{XX} &= p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}X_tX'_t.\end{aligned}$$

Moreover, under the additional Assumption 5, Σ_{XX} *is nonsingular and the matrix*

$$\Sigma^*_{\beta\beta} = \Sigma_{\beta\beta} - \Sigma'_{X\beta}\Sigma^{-1}_{XX}\Sigma_{X\beta}$$

is nonsingular. Furthermore, under Assumptions 1,2 and 5,

$$\begin{aligned}\Sigma_{\beta\otimes I_{m+1},\beta\otimes I_{m+1}} &= p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left(\beta'\otimes I_{m+1}\right)Y^{(m)}_{t-1}Y^{(m)\prime}_{t-1}\left(\beta\otimes I_{m+1}\right) \\ &= \Sigma_{\beta\beta}\otimes I_{m+1}, \\ \Sigma_{\beta,\beta\otimes I_{m+1}} &= p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\beta' Y_{t-1}Y^{(m)\prime}_{t-1}\left(\beta\otimes I_{m+1}\right) \\ &= \left(\Sigma_{\beta\beta}, O_{r,r.m}\right), \\ \Sigma_{X,\beta\otimes I_{m+1}} &= p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}X_tY^{(m)\prime}_{t-1}\left(\beta\otimes I_{m+1}\right) \\ &= \left(\Sigma_{X\beta}, O_{k(p-1),r.m}\right).\end{aligned}$$

Consequently,

$$\Sigma_{\beta,\beta\otimes I_{m+1}} - \Sigma_{\beta X}\Sigma^{-1}_{XX}\Sigma_{X,\beta\otimes I_{m+1}} = \left(\Sigma^*_{\beta\beta}, O_{r,r.m}\right)$$

and

$$\begin{aligned}\Sigma^*_{\beta\otimes I_{m+1},\beta\otimes I_{m+1}} &= \Sigma_{\beta\otimes I_{m+1},\beta\otimes I_{m+1}} - \Sigma'_{X,\beta\otimes I_{m+1}}\Sigma^{-1}_{XX}\Sigma_{\beta\otimes I_{m+1},X} \\ &= \begin{pmatrix}\Sigma^*_{\beta\beta} & O_{r,r.m} \\ O_{r.m,r} & \Sigma_{\beta\beta}\otimes I_m\end{pmatrix}.\end{aligned}$$

The latter is a nonsingular matrix.

Proof. See Bierens and Martins (2009).

Lemma A.4. *Let $\alpha_\perp$ be an orthogonal complement of α. Then under Assumptions 1-5,*

$$S_{00,T}^{-1} = \begin{pmatrix} \left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' \\ \left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1} \end{pmatrix}' \begin{pmatrix} I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & \left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma_{\beta\beta}^*\right)^{-1} \end{pmatrix} \times \begin{pmatrix} \left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' \\ \left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1} \end{pmatrix} + o_p(1).$$

Proof. This is a standard result. See for example Johansen (1995).

Lemma A.5. *Let* ξ *be given by* (9). *Under Assumptions 1-5,*

$$\begin{aligned} N_T &= S_{00,T}^{-1} - S_{00,T}^{-1}S_{01,T}^{(m)}\xi\left(\xi'S_{10,T}^{(m)}S_{00,T}^{-1}S_{01,T}^{(m)}\xi\right)^{-1}\xi'S_{10,T}^{(m)}S_{00,T}^{-1} \\ &= S_{00,T}^{-1} - S_{00,T}^{-1}S_{01,T}^{(0)}\beta\left(\beta'S_{10,T}^{(0)}S_{00,T}^{-1}S_{01,T}^{(0)}\beta\right)^{-1}\beta'S_{10,T}^{(0)}S_{00,T}^{-1} \\ &= \alpha_\perp\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1}\alpha_\perp' + o_p(1). \end{aligned}$$

Proof. See Johansen (1995, Lemma 10.1).

Lemma A.6. *There exists an orthogonal complement $\beta_\perp'$ of β such that*

$$\beta_\perp'C(1) = \left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp'C_0'.$$

Proof. This is a standard result. See Johansen (1995).

Lemma A.7. *Let $\beta_\perp$ be the orthogonal complement of β defined in Lemma A.6. Let Assumptions 1-5 hold. Then*

$$\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp'S_{01,T}^{(m)}\left(\beta_\perp'\otimes I_{m+1}\right) \xrightarrow{d} \int_0^1 \left(dW_{k-r}\right)\widetilde{W}_{k-r,m}'$$

and

$$\sqrt{T}\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp'S_{01,T}^{(m)}\left(\beta\otimes I_{m+1}\right) \xrightarrow{d} Z \qquad (27)$$

jointly, where Z is a $(k-r) \times r(m+1)$ random matrix. In particular, the $k-r$ columns of Z' are independent

$$N_{r(m+1)}\left[0, \begin{pmatrix} \Sigma^*_{\beta\beta} & O_{r,r.m} \\ O_{r.m,r} & \Sigma_{\beta\beta} \otimes I_m \end{pmatrix}\right] \tag{28}$$

distributed. Moreover,

$$\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}S^{(m)}_{01,T}\left(\beta'_\perp \otimes I_{m+1}\right) \xrightarrow{d} M,$$

where M is a $r \times (k-r)(m+1)$ random matrix, and

$$\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}S^{(m)}_{01,T}\left(\beta \otimes I_{m+1}\right) = \left(\Sigma_{\beta\beta}, O_{r,r.m}\right) + o_p(1).$$

Proof. See Bierens and Martins (2009).

Lemma A.8. *Under Assumptions 1-5,*

$$\begin{aligned}&\left(\beta' \otimes I_{m+1}, T^{-1/2}\beta'_\perp \otimes I_{m+1}\right) S^{(m)}_{10,T} S^{-1}_{00,T} S^{(m)}_{01,T}\left(\beta \otimes I_{m+1}, T^{-1/2}\beta'_\perp \otimes I_{m+1}\right)\\ &\xrightarrow{d} \begin{pmatrix} \Sigma^*_{\beta\beta}\left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma^*_{\beta\beta}\right)^{-1}\Sigma^*_{\beta\beta} & O_{r,k-r+k.m} \\ O_{k-r+k.m,r} & O_{k-r+k.m,k-r+k.m} \end{pmatrix}.\end{aligned}$$

Proof. This result follows straightforwardly from Lemmas A.4 and A.7.

Lemma A.9. *Under Assumptions 1-5,*

$$\begin{aligned}&\left(T^{-1/2}\beta'_\perp \otimes I_{m+1}, \beta' \otimes I_{m+1}\right) S^{(m)}_{11,T}\left(T^{-1/2}\beta'_\perp \otimes I_{m+1}, \beta \otimes I_{m+1}\right)\\ &\xrightarrow{d} \begin{pmatrix} \int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx & O_{(k-r)(m+1),r(m+1)} \\ O_{r(m+1),(k-r)(m+1)} & \begin{pmatrix} \Sigma^*_{\beta\beta} & O_{r,r.m} \\ O_{r.m,r} & \Sigma_{\beta\beta} \otimes I_m \end{pmatrix} \end{pmatrix}\end{aligned}$$

Proof. See Bierens and Martins (2009).

Combining the results of Lemmas A.8 and A.9 it follows from Lemma 2 in Andersson et al. (1983) that

Lemma A.10. *Under Assumptions 1-5 the ordered solutions* $\lambda_{1,T} \geq \lambda_{2,T} \geq \ldots \geq \lambda_{(m+1)k,T}$ *of the generalized eigenvalue problem* (5) *converge in distribution to the ordered solutions* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{(m+1)k}$ *of*

$$\det\left[\lambda\begin{pmatrix}\begin{pmatrix}\Sigma^*_{\beta\beta} & O_{r,r.m}\\ O_{r.m,r} & \Sigma_{\beta\beta}\otimes I_m\end{pmatrix} & O_{r(m+1),(k-r)(m+1)}\\ O_{(k-r)(m-1),r(m+1)} & \int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx\end{pmatrix}\right.$$
$$\left.-\begin{pmatrix}\Sigma^*_{\beta\beta}\left((\alpha'\Omega^{-1}\alpha)^{-1}+\Sigma^*_{\beta\beta}\right)^{-1}\Sigma^*_{\beta\beta} & O_{r,k(m+1)-r}\\ O_{k(m+1)-r,r} & O_{k(m+1)-r,k(m+1)-r}\end{pmatrix}\right]=0.$$

Obviously, all but r solutions are zero, and the non-zero solutions are the solutions of eigenvalue problem

$$\det\left(\lambda\Sigma^*_{\beta\beta}-\Sigma^*_{\beta\beta}\left((\alpha'\Omega^{-1}\alpha)^{-1}+\Sigma^*_{\beta\beta}\right)^{-1}\Sigma^*_{\beta\beta}\right)=0$$

This is the same result as in the standard TI cointegration case!

With these results at hand we are now able to prove our main results.

Proof of Lemma 2: As shown in Bierens and Martins (2009), Lemma 2 follows straightforwardly from Lemmas A.1 and A.2.

Proof of Lemma 3: Lemma 3 follows from Lemma A.10.

Proof of Lemma 4: To derive the limiting distribution of

$$T\left(\lambda_{r+1,T},\lambda_{r+2,T},\ldots,\lambda_{k,T}\right)',$$

we follow a similar procedure as in Johansen (1995, p.159). Let

$$\begin{aligned}S(\lambda) &= \lambda S^{(m)}_{11,T}-S^{(m)}_{10,T}S^{-1}_{00,T}S^{(m)}_{01,T}\\ \xi_{\perp,T} &= \left(\beta_\perp\otimes I_{m+1},\begin{pmatrix}O_{k,m.r}\\ \sqrt{T}\left(\beta\Sigma^{-1/2}_{\beta\beta}\otimes I_m\right)\end{pmatrix}\right), \qquad (29)\\ \rho &= T.\lambda = O_p(1).\end{aligned}$$

The reason for the factor $\sqrt{T}$ in (29) is to prevent $T^{-1}\xi'_{\perp,T}S^{(m)}_{11,T}\xi_{\perp,T}$ from converging to a singular matrix, because otherwise we cannot apply Lemma

2 in Andersson et al. (1983). The reason for the normalization of β by $\Sigma_{\beta\beta}^{-1/2}$ will become clear below. Then

$$\det\left(\left(\begin{array}{c}\xi'\\ \xi'_{\perp,T}\end{array}\right)S(\lambda)(\xi,\xi_{\perp,T})\right)=\det\left(\begin{array}{cc}\xi'S(\lambda)\xi & \xi'S(\lambda)\xi_{\perp,T}\\ \xi'_{\perp,T}S(\lambda)\xi & \xi'_{\perp,T}S(\lambda)\xi_{\perp,T}\end{array}\right)$$
$$=\det\left(\xi'S(\lambda)\xi\right)\det\left(\xi'_{\perp,T}\left(S(\lambda)-S(\lambda)\xi\left(\xi'S(\lambda)\xi\right)^{-1}\xi'S(\lambda)\right)\xi_{\perp,T}\right),$$

where ξ is defined by (9).

It follows from Lemma A.9 that $\beta' S_{11,T}^{(0)}\beta = O_p(1)$ and $\xi'_{\perp,T}S_{11,T}^{(m)}\xi = O_p(1)$, whereas by assumption, $\rho = O_p(1)$. Therefore, similar to Johansen (1995, p.159),

$$\xi'S(\lambda)\xi = -\beta' S_{10,T}^{(0)}S_{00,T}^{-1}S_{01,T}^{(0)}\beta + o_p(1)$$

and

$$\xi'_{\perp,T}S(\lambda)\xi = -\xi'_{\perp,T}S_{10,T}^{(m)}S_{00,T}^{-1}S_{01,T}^{(0)}\beta + o_p(1).$$

Combining these results it follows that

$$\xi'_{\perp,T}\left(S(\lambda)-S(\lambda)\xi\left(\xi'S(\lambda)\xi\right)^{-1}\xi'S(\lambda)\right)\xi_{\perp,T}$$
$$=\rho\frac{1}{T}\xi'_{\perp,T}S_{11,T}^{(m)}\xi_{\perp,T}-\xi'_{\perp,T}S_{10,T}^{(m)}N_T S_{01,T}^{(m)}\xi_{\perp,T}+o_p(1),$$

where N_T is defined in Lemma A.5. Since by Lemma A.7, $S_{01,T}^{(m)}\xi_{\perp,T} = O_p(1)$, it follows now from Lemmas A.5-A.7 that

$$\xi'_{\perp,T}\left(S(\lambda)-S(\lambda)\xi\left(\xi'S(\lambda)\xi\right)^{-1}\xi'S(\lambda)\right)\xi_{\perp,T}$$
$$=\rho\frac{1}{T}\xi'_{\perp,T}S_{11,T}^{(m)}\xi_{\perp,T}$$
$$-\xi'_{\perp,T}S_{10,T}^{(m)}\alpha_\perp\left(\alpha'_\perp\Omega\alpha_\perp\right)^{-1}\alpha'_\perp S_{01,T}^{(m)}\xi_{\perp,T}+o_p(1).$$

Next, observe from Lemma A.9 that

$$\frac{1}{T}\left(\begin{array}{cc}I_{(m+1)(k-r)} & O_{(m+1)(k-r),r.m}\\ O_{r.m,(m+1)(k-r)} & \Sigma_{\beta\beta}^{1/2}\otimes I_m\end{array}\right)\xi'_{\perp,T}S_{11,T}^{(m)}\xi_{\perp,T}$$
$$\times\left(\begin{array}{cc}I_{(m+1)(k-r)} & O_{(m+1)(k-r),r.m}\\ O_{r.m,(m+1)(k-r)} & \Sigma_{\beta\beta}^{1/2}\otimes I_m\end{array}\right)$$
$$\stackrel{d}{\to}\left(\begin{array}{cc}\int_0^1\widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx & O_{(k-r)(m+1),r.m}\\ O_{r.m,(k-r)(m+1)} & \overline{\Psi}\end{array}\right),$$

where by Lemma A.9,

$$\begin{aligned}
\overline{\Psi} &= p\lim_{T\to\infty}\begin{pmatrix} O_{k,m.r} \\ \beta\otimes I_m \end{pmatrix}' S_{11,T}^{(m)} \begin{pmatrix} O_{k,m.r} \\ \beta\otimes I_m \end{pmatrix} \\
&= \begin{pmatrix} O_{r,r.m} \\ I_{r.m} \end{pmatrix}' p\lim_{T\to\infty}\left(\beta'\otimes I_{m+1}\right) S_{11,T}^{(m)} \left(\beta\otimes I_{m+1}\right) \begin{pmatrix} O_{r,r.m} \\ I_{r.m} \end{pmatrix} \\
&= \Sigma_{\beta\beta}\otimes I_m.
\end{aligned}$$

Hence

$$\frac{1}{T}\xi_{\perp,T}' S_{11,T}^{(m)} \xi_{\perp,T} \xrightarrow{d} \begin{pmatrix} \int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}_{k-r,m}'(x)dx & O_{(k-1)(m+1),r.m} \\ O_{r.m,(k-1)(m+1)} & I_{r.m} \end{pmatrix}. \tag{30}$$

Moreover, it follows from Lemma A.7 that

$$\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' S_{01,T}^{(m)}\left(\beta_\perp\otimes I_{m+1}\right) \xrightarrow{d} \int_0^1 \left(dW_{k-r}\right)\widetilde{W}_{k-r,m}' \tag{31}$$

and

$$\begin{aligned}
&\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' S_{01,T}^{(m)} \begin{pmatrix} O_{k,m.r} \\ T^{1/2}\beta\otimes I_m \end{pmatrix} \\
&\quad = \left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1/2}\alpha_\perp' S_{01,T}^{(m)} \left(T^{1/2}\beta\otimes I_{m+1}\right) \begin{pmatrix} O_{r,r.m} \\ I_{r.m} \end{pmatrix} \\
&\quad \xrightarrow{d} Z_2,
\end{aligned} \tag{32}$$

where

$$Z_2 = Z\begin{pmatrix} O_{r,r.m} \\ I_{r.m} \end{pmatrix} \tag{33}$$

with Z defined in Lemma A.7. Hence, the columns of Z_2' are independently $N_{r.m}\left[0,\Sigma_{\beta\beta}\otimes I_m\right]$ distributed. Consequently, the columns of

$$V = \left(\Sigma_{\beta\beta}^{-1/2}\otimes I_m\right) Z_2' \tag{34}$$

are independent $N_{r.m}\left[0, I_{r.m}\right]$ distributed, and it follows from (31) and (32) that

$$\begin{aligned}
&\xi_{\perp,T}' S_{10,T}^{(m)}\alpha_\perp\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1}\alpha_\perp' S_{01,T}^{(m)}\xi_{\perp,T} \\
&\quad \xrightarrow{d} \begin{pmatrix} \int_0^1 \widetilde{W}_{k-r,m} dW_{k-r}' \\ V \end{pmatrix} \left(\int_0^1 \left(dW_{k-r}\right)\widetilde{W}_{k-r,m}', V'\right).
\end{aligned} \tag{35}$$

Lemma 4 now follows from (30), (35), Lemma 2 in Anderson et al. (1983), and the next lemma:

Lemma A.11. *Under Assumptions 1-5, V is independent of W_{k-r} and $\widetilde{W}_{k-r,m}$.*

Proof. See Bierens and Martins (2009).

Proof of Lemma 5: Let $\widehat{\xi} = \left(\widehat{\xi}_1, ..., \widehat{\xi}_r\right)$ be the ML estimator of ξ, where $\widehat{\xi}_i$, $i = 1, ..., r$, are the eigenvectors associated with the r largest eigenvalues $\widehat{\lambda}_{m,i}$,

$$S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \widehat{\xi}_i = \widehat{\lambda}_{m,i} S_{11,T}^{(m)} \widehat{\xi}_i, \; i = 1, .., r.$$

If we normalize $\widehat{\xi}$ as

$$\widetilde{\xi} = \widehat{\xi} \left(\xi' \widehat{\xi}\right)^{-1} \xi' \xi$$

then similar to Johansen (1988) we can write

$$\widetilde{\xi} - \xi = \xi_{\perp,T} U_{m,T} \tag{36}$$

where $\xi_{\perp,T}$ is defined by (29), and

$$U_{m,T} = \left(\xi_{\perp,T}' \xi_{\perp,T}\right)^{-1} \left(\xi_{\perp,T}' \widehat{\xi}\right) \left(\xi' \widehat{\xi}\right)^{-1} \left(\xi' \xi\right).$$

Similar to Johansen (1988) we can expand $T.U_T$ as

$$T.U_{m,T} = \left(T^{-1} \xi_{\perp,T}' \widehat{S}_{11,T}^{(m)} \xi_{\perp,T}\right)^{-1} \xi_{\perp,T}' \left(\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)} \xi \alpha'\right) \times \Omega^{-1} \alpha \left(\alpha' \Omega^{-1} \alpha\right)^{-1} + o_p(1).$$

Moreover, similar to Johansen (1988) it can be shown that

$$\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)} \xi \alpha' = \frac{1}{T} \sum_{t=1}^{T} Y_{t-1}^{(m)} U_t' C_0' - \left(\frac{1}{T} \sum_{t=1}^{T} Y_{t-1}^{(m)} X_t'\right) \left(\frac{1}{T} \sum_{t=1}^{T} X_t X_t'\right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} X_t U_t' C_0'\right).$$

Thus,

$$\left(\beta_\perp' \otimes I_{m+1}\right)\left(\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)}\xi\alpha'\right) = \left(\frac{1}{T}\sum_{t=1}^{T}\left(\beta_\perp' \otimes I_{m+1}\right)Y_{t-1}^{(m)}U_t'C_0'\right) + o_p(1),$$

and by Lemma A.3,

$$\sqrt{T}\left(\beta' \otimes I_{m+1}\right)\left(\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)}\xi\alpha'\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\beta' \otimes I_{m+1}\right)Y_{t-1}^{(m)}U_t'C_0' - \begin{pmatrix} \Sigma_{X\beta}'\Sigma_{XX}^{-1} \\ O_{r.m,k(p-1)} \end{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_tU_t'C_0' + o_p(1).$$

Hence,

$$\begin{aligned}
&\xi_{\perp,T}'\left(\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)}\xi\alpha'\right) \\
&= \begin{pmatrix} \left(\beta_\perp' \otimes I_{m+1}\right)\left(\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)}\xi\alpha'\right) \\ \left(O_{m.r,k}, \sqrt{T}\left(\Sigma_{\beta\beta}^{-1/2}\beta' \otimes I_m\right)\right)\left(\widehat{S}_{10,T}^{(m)} - \widehat{S}_{11,T}^{(m)}\xi\alpha'\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\left(\beta_\perp' \otimes I_{m+1}\right)Y_{t-1}^{(m)}U_t'C_0' \\ \left(O_{m.r,k}, I_{r.m}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\Sigma_{\beta\beta}^{-1/2}\beta' \otimes I_{m+1}\right)Y_{t-1}^{(m)}U_t'C_0' \end{pmatrix} \\
&\quad + o_p(1).
\end{aligned}$$

Similar to Johansen (1988) it follows now that

$$\frac{1}{T}\sum_{t=1}^{T}\left(\beta_\perp' \otimes I_{m+1}\right)Y_{t-1}^{(m)}U_t'C_0'\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \xrightarrow{d} \int_0^1 \widetilde{W}_{k-r,m}d\underline{W}_\alpha',$$

where

$$\underline{W}_\alpha = \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}\alpha'\Omega^{-1}C_0W$$

is an r-variate standard Wiener process, which is independent of $\widetilde{W}_{k-r,m}$. Moreover, similar to parts (27) and (28) of Lemma A.7 it follows that

$$\left(O_{m.r,k}, I_{r.m}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\Sigma_{\beta\beta}^{-1/2}\beta' \otimes I_{m+1}\right)Y_{t-1}^{(m)}U_t'C_0'\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \xrightarrow{d} \underline{V}_\alpha,$$

where $\underline{V}_\alpha$ is an $r.m \times r$ matrix with independent $N[0,1]$ distributed elements, which is also independent of $\widetilde{W}_{k-r,m}$. However, similar to Lemma A.11 it can be shown that $\underline{V}_\alpha$, $\underline{W}_\alpha$ and $\widetilde{W}_{k-r,m}$ are independent. Therefore, it follows from Lemma A.9 that (15) holds.

Denoting $\widetilde{\xi} = \left(\widetilde{\xi}_0', \widetilde{\xi}_m'\right)$, where $\widetilde{\xi}_0$ is a $k \times r$ matrix and $\widetilde{\xi}_m$ a $k.m \times r$ matrix, it follows now from (15), (29) and (36) that jointly,

$$\begin{aligned} T\left(\widehat{\xi}_0 - \beta\right) &\xrightarrow{d} (\beta_\perp, O_{k,k.m}) \left(\int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}_{k-r,m}'(x)dx\right)^{-1} \\ &\quad \times \int_0^1 \widetilde{W}_{k-r,m} d\underline{W}_\alpha' \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}, \\ \sqrt{T}\widetilde{\xi}_m &\xrightarrow{d} \left(\beta\Sigma_{\beta\beta}^{-1/2} \otimes I_m\right) \underline{V}_\alpha \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}. \end{aligned}$$

Proof of Theorem 1: Consider the likelihood-ratio statistic $f_{m,T}\left(\widetilde{\xi}\right) - f_{0,T}\left(\widetilde{\beta}\right)$, where

$$\widetilde{\xi} = \widehat{\xi}\left(\xi'\widehat{\xi}\right)^{-1}\xi'\xi, \ \widetilde{\beta} = \widehat{\beta}\left(\beta'\widehat{\beta}\right)^{-1}\beta'\beta$$

and

$$\begin{aligned} f_{0,T}(\beta) &= T.\ln\left(\frac{\det\left(\beta'\left(S_{11,T}^{(0)} - S_{10,T}^{(0)}S_{00,T}^{-1}S_{01,T}^{(0)}\right)\beta\right)}{\det\left(\beta' S_{11,T}^{(0)}\beta\right)}\right), \\ f_{m,T}(\xi) &= T.\ln\left(\frac{\det\left(\xi'\left(S_{11,T}^{(m)} - S_{10,T}^{(m)}S_{00,T}^{-1}S_{01,T}^{(m)}\right)\xi\right)}{\det\left(\xi' S_{11,T}^{(m)}\xi\right)}\right). \end{aligned}$$

Recall from Lemma 5 that $\widetilde{\xi} = \xi + \xi_{\perp,T}U_{m,T}$, where $U_{m,T} = O_p\left(T^{-1}\right)$. It follows from Johansen (1988, Lemma 7, p. 249) [5] that under the null hypothesis $\xi = (\beta', O_{r,k.m})'$,

$$f_{m,T}\left(\widetilde{\xi}\right) = f_{m,T}\left(\xi + \xi_{\perp,T}U_{m,T}\right)$$

[5] See also Johansen (1995), equation A.11 on page 224.

$$
\begin{aligned}
= \quad & f_{0,T}(\beta) + T.\text{trace}\left\{ \left(\beta' \left(S_{11,T}^{(0)} - S_{10,T}^{(0)} S_{00,T}^{-1} S_{01,T}^{(0)} \right) \beta \right)^{-1} \right. \\
& \times \left(U_{m,T}' \xi_{\perp,T}' \left(S_{11,T}^{(m)} - S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \right) \xi_{\perp,T} U_T \right. \\
& - U_{m,T}' \xi_{\perp,T}' \left(S_{11,T}^{(m)} - S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \right) \xi \\
& \times \left(\beta' \left(S_{11,T}^{(0)} - S_{10,T}^{(0)} S_{00,T}^{-1} S_{01,T}^{(0)} \right) \beta \right)^{-1} \\
& \left. \left. \times \xi' \left(S_{11,T}^{(m)} - S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \right) \xi_{\perp,T} U_{m,T} \right) \right\} \\
& - T.\text{trace}\left\{ \left(\beta' S_{11,T}^{(0)} \beta \right)^{-1} \left(U_T' \xi_{\perp,T}' S_{11,T}^{(m)} \xi_{\perp,T} U_{m,T} \right. \right. \\
& \left. - U_{m,T}' \xi_{\perp,T}' S_{11,T}^{(m)} \xi \left(\beta' S_{11,T}^{(0)} \beta \right)^{-1} \xi' S_{11,T}^{(m)} \xi_{\perp,T} U_{m,T} \right\} \\
& + O\left(T. \left\| \xi_{\perp,T} U_{m,T} \right\|^3 \right),
\end{aligned}
$$

where for a matrix the norm $\|.\|$ denotes the maximum absolute value of its elements. Since

$$
\begin{aligned}
& U_{m,T} = O_p\left(T^{-1}\right), \; \xi_{\perp,T} U_{m,T} = O_p\left(T^{-1/2}\right), \\
& \xi_{\perp,T}' \left(S_{11,T}^{(m)} - S_{10,T}^{(m)} S_{00,T}^{-1} S_{01,T}^{(m)} \right) \xi = O_p(1), \\
& \xi_{\perp,T}' S_{11,T}^{(m)} \xi = O_p(1),
\end{aligned}
$$

and by Johansen (1995, Lemma 10.1),

$$
\left(\beta' S_{11,T}^{(0)} \beta \right)^{-1} - \left(\beta' \left(S_{11,T}^{(0)} - S_{10,T}^{(0)} S_{00,T}^{-1} S_{01,T}^{(0)} \right) \beta \right)^{-1} = -\alpha' \Omega^{-1} \alpha + o_p(1),
$$

it follows now from (30) and Lemma 5 that

$$
\begin{aligned}
& f_{m,T}\left(\widetilde{\xi}\right) - f_{0,T}(\beta) \\
& = \text{trace}\left[\left(\alpha' \Omega^{-1} \alpha \right) \left(T.U_{m,T}' \right) \xi_{\perp,T}' \left(\frac{1}{T} S_{11,T}^{(m)} \right) \xi_{\perp,T} \left(T.U_{m,T} \right) \right] + o_p(1) \\
& \xrightarrow{d} \text{trace}\left(\underline{V}_\alpha' \underline{V}_\alpha \right) \\
& \qquad + \text{trace}\left[\left(\int_0^1 d\underline{W}_\alpha \widetilde{W}_{k-r,m}' \right) \left(\int_0^1 \widetilde{W}_{k-r,m}(x) \widetilde{W}_{k-r,m}'(x) dx \right)^{-1} \right. \\
& \qquad\qquad \left. \times \left(\int_0^1 \widetilde{W}_{k-r,m} d\underline{W}_\alpha' \right) \right]
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&f_{0,T}\left(\widetilde{\beta}\right) - f_{0,T}\left(\beta\right) \\
&\xrightarrow{d} \text{trace}\left[\left(\int_0^1 d\underline{W}_\alpha W'_{k-r}\right)\left(\int_0^1 W_{k-r}(x)W'_{k-r}(x)dx\right)^{-1}\right. \\
&\qquad\qquad \left.\times\left(\int_0^1 W_{k-r}d\underline{W}'_\alpha\right)\right].
\end{aligned}
$$

Johansen (1995, page 192) has shown that, with $V_\alpha = \left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1/2}W$,

$$
\begin{aligned}
&\text{trace}\left[\left(\alpha'\Omega^{-1}\alpha\right)\left(\int_0^1 dV_\alpha W'_{k-r}\right)\left(\int_0^1 W_{k-r}(x)W'_{k-r}(x)dx\right)^{-1}\right. \\
&\qquad\qquad \left.\times\left(\int_0^1 W_{k-r}dV'_\alpha\right)\right] \sim \chi^2_{r(k-r)}
\end{aligned}
$$

In our notation, $\underline{W}_\alpha = \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}\alpha'\Omega^{-1}C_0W$ is a r-variate standard Wiener process, which is distributed as $\left(\alpha'\Omega^{-1}\alpha\right)^{1/2}V_\alpha$ with V_α as in Johansen(1995). Thus,

$$
\begin{aligned}
&\text{trace}\left[\left(\int_0^1 d\underline{W}_\alpha W'_{k-r}\right)\left(\int_0^1 W_{k-r}(x)W'_{k-r}(x)dx\right)^{-1}\right. \qquad (37) \\
&\qquad\qquad \left.\times\left(\int_0^1 W_{k-r}d\underline{W}'_\alpha\right)\right] \sim \chi^2_{r(k-r)}.
\end{aligned}
$$

Similarly, it follows that

$$
\begin{aligned}
&\text{trace}\left[\left(\int_0^1 d\underline{W}_\alpha \widetilde{W}'_{k-r,m}\right)\left(\int_0^1 \widetilde{W}_{k-r,m}(x)\widetilde{W}'_{k-r,m}(x)dx\right)^{-1}\right. \qquad (38) \\
&\qquad\qquad \left.\times\left(\int_0^1 \widetilde{W}_{k-r,m}d\underline{W}'_\alpha\right)\right] \sim \chi^2_{r(m+1)(k-r)},
\end{aligned}
$$

because $\underline{W}_\alpha$ and $\widetilde{W}_{k-r,m}$ are independent. Then the difference of (38) and (37) is $\chi^2_{r.m.(k-r)}$ distributed, which follows from the following easy result:

If $Z = \begin{pmatrix} Y \\ X \end{pmatrix} \sim N_{p+q}\left[0, \Sigma\right]$, where $Y \in \mathbb{R}^p$, $X \in \mathbb{R}^q$ and $\Sigma = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}$, $\det(\Sigma) > 0$, then $Z'\Sigma^{-1}Z - X'\Sigma^{-1}_{XX}X \sim \chi^2_p$.

Moreover, since $\underline{V}_\alpha$ is a $r.m \times r$ matrix with independent $N[0,1]$ distributed elements, it follows that

$$\text{trace}\left(\underline{V}'_\alpha \underline{V}_\alpha\right) \sim \chi^2_{r.m.r}. \tag{39}$$

Furthermore, since $\underline{V}_\alpha$ and $\underline{W}_\alpha$ are independent, (39) is independent of (37) and (38), conditional on $\widetilde{W}_{k-r,m}$. Hence, the likelihood-ratio statistic $T\left(\widehat{f}_1\left(\widetilde{\xi}\right) - \widehat{f}_0\left(\beta\right)\right) - T\left(\widehat{f}_0\left(\widetilde{\beta}\right) - \widehat{f}_0\left(\beta\right)\right)$ converges in distribution to (39) plus (38) minus (37), resulting in a χ^2_{mkr} distribution.

Proof of Lemma 6: To prove (20), let $b_{2,i,j}(\tau)$ be element (i,j) of $B_2\left(t/T\right)$ with derivative $b'_{2,i,j}(\tau)$ and let $Z_{2,j,t-1}$ be component j of $Z_{2,t-1}$. Then by the mean value theorem,

$$\begin{aligned}
&\Delta\left(b_{2,i,j}(t/T)Z_{2,j,t-1}\right) \\
&\quad = \left(b_{2,i,j}(t/T) - b_{2,i,j}((t-1)/T)\right) Z_{2,j,t-1} + b_{2,i,j}((t-1)/T)\Delta Z_{2,j,t-1} \\
&\quad = b'_{2,i,j}((t-\lambda_{t,i,j,T})/T)Z_{2,j,t-1}/T + b_{2,i,j}((t-1)/T)\Delta Z_{2,j,t-1}
\end{aligned}$$

for some $\lambda_{t,i,j,T} \in [0,1]$. Denote by $\Psi_{t,T}$ be the matrix with elements $\Psi_{i,j,t,T} = b'_{2,i,j}((t-\lambda_{t,i,j,T})/T)$. Then

$$\begin{aligned}
\Delta\left(B_2(t/T)Z_{2,t-1}\right) &= \Psi_{t,T}Z_{2,t-1}/T + B_2(t/T)\Delta Z_{2,t-1} \\
&= B_2(t/T)\Delta Z_{2,t-1} + O_p\left(1/\sqrt{T}\right) \qquad (40)
\end{aligned}$$

where the latter follows from the fact that $\Psi_{t,T}$ is uniformly bounded and that $Z_{2,t-1}/\sqrt{T} = O_p(1)$.

Next, observe from (16) and (40) that

$$\begin{aligned}
\Delta Z_{1,t} &= \sum_{j=1}^{p} D_j \Delta Z_{1,t-j} + \Delta\left(B_2\left(t/T\right)Z_{2,t-1}\right) + \sum_{j=1}^{p-1} C_{12,j}\Delta^2 Z_{2,t-j} + \Delta U_{1,t} \\
&= \sum_{j=0}^{t-1} \Pi_j \Delta\left(B_2\left((t-j)/T\right)Z_{2,t-1-j}\right) + \sum_{j=t}^{\infty} \Pi_j B_2\left(0\right)\Delta Z_{2,t-1-j} \\
&\quad + \sum_{i=1}^{p-1} C_{12,i} \sum_{j=0}^{\infty} \Pi_j \Delta^2 Z_{2,t-i-j} + \sum_{j=0}^{\infty} \Pi_j \Delta U_{1,t-j} \\
&= \sum_{j=0}^{t-1} \Pi_j \left(B_2\left((t-j)/T\right) - B_2\left(0\right)\right)\Delta Z_{2,t-1-j}
\end{aligned}$$

$$+V_t + O_p\left(1/\sqrt{T}\right), \tag{41}$$

where

$$V_t = \sum_{j=0}^{\infty} \Pi_j B_2(0)\, \Delta Z_{2,t-1-j} + \sum_{i=1}^{p-1} C_{12,i} \sum_{j=0}^{\infty} \Pi_j \Delta^2 Z_{2,t-i-j} + \sum_{j=0}^{\infty} \Pi_j \Delta U_{1,t-j}.$$

This proves (20).

Finally, it follows from (17) and (41) that

$$\begin{aligned}
& B_1 Z_{1,t-1} + B_2(t/T)\, Z_{2,t-1} \\
&\quad = \Delta Z_{1,t} - \sum_{j=1}^{p-1} C_{11,j} \Delta Z_{1,t-j} - \sum_{j=1}^{p-1} C_{12,j} \Delta Z_{2,t-j} - U_{1,t} \\
&\quad = \sum_{j=0}^{t-1} \Pi_j \left(B_2\left((t-j)/T\right) - B_2(0)\right) \Delta Z_{2,t-1-j} \\
&\quad - \sum_{i=1}^{p-1} C_{11,i} \sum_{j=0}^{t-1-i} \Pi_j \left(B_2\left((t-j-i)/T\right) - B_2(0)\right) \Delta Z_{2,t-i-j-1} \\
&\quad + V_t - \sum_{i=1}^{p-1} C_{11,i} V_{t-i} - \sum_{j=1}^{p-1} C_{12,j} \Delta Z_{2,t-j} - U_{1,t} + O_p\left(1/\sqrt{T}\right) \\
&\quad = \sum_{j=0}^{t-1} Q_j \left(B_2\left((t-j)/T\right) - B_2(0)\right) \Delta Z_{2,t-1-j} + R_t + O_p\left(1/\sqrt{T}\right),
\end{aligned}$$

where

$$R_t = V_t - \sum_{i=1}^{p-1} C_{11,i} V_{t-i} - \sum_{j=1}^{p-1} C_{12,j} \Delta Z_{2,t-j} - U_{1,t}.$$