A WALD TEST FOR THE COINTEGRATION RANK IN NONSTATIONARY FRACTIONAL SYSTEMS

MARCO AVARUCCI and CARLOS VELASCO

Dipartimento di Scienze economiche e aziendali, LUISS Guido Carli
and

Department of Economics, Universidad Carlos III de Madrid

January 13, 2008

Abstract

This paper develops new methods for determining the cointegration rank in a nonstationary fractionally integrated system, extending univariate optimal methods for testing the degree of integration. We propose a simple Wald test based on the singular value decomposition of the unrestricted estimate of the long run multiplier matrix. When the "strength" of the cointegrating relationship is less than $1/2$, the statistic has a standard asymptotic distribution, like Lagrange Multiplier tests exploiting local properties. We consider the behavior of our test under estimation of short run parameters and local alternatives. We compare our procedure with other cointegration tests based on different principles and find that the new method has better properties in a range of situations by using information on the alternative obtained through a preliminary estimate of the cointegration strength.

Keywords and phrases: Fractional integration, fractional error correction model, singular value decomposition, cointegration test.

1 Introduction

Fractional cointegration models are increasingly used as a flexible alternative for the modeling of long run relationships among economic time series. These models allow observed time series to be integrated of any arbitrary order, being even stationary as in many financial applications, and simultaneously permit any degree of persistence for the equilibrium relationship. Much effort has been dedicated in the last years to the estimation of the cointegrating relationship, including the asymptotic analysis of different variants of least squares (LS), such as narrow band and generalized LS (GLS) versions, see e.g. Robinson and Marinucci (2003) and Robinson and Hualde (2003). At the same time, a number of cointegration tests have been developed,

most of them built on the null hypothesis of no cointegration versus the alternative of fractional cointegration. If the cointegration vector is known, standard methods for testing the integration degree could be routinely applied, but if this vector, or the level of integration of the original series, has to be estimated, inference methods should adapt to these additional sources of uncertainty. Thus, testing for the cointegration rank in this framework poses further complications in systems with more than two series, even if certain restrictions on the definition of cointegration are imposed.

In a semiparametric frequency domain framework, Marinucci and Robinson (2001) suggested a Hausman-type cointegration test comparing different estimates of the integration orders of the observed series. Recently Robinson (2007) provided rigorous theoretical support to this idea. Marmol and Velasco (2004) proposed a Wald test of the null of spurious relationships against the alternative of a single cointegration relationship among the components of a nonstationary vector process. Their approach relies upon comparing OLS and narrow band GLS-type estimates of the cointegrating vector, with different properties under the competitive hypothesis. A similar idea was used by Hualde and Velasco (2007), employing the GLS estimates of Robinson and Hualde (2003). The chi-squared distribution of the GLS Wald statistic is inherited by a parallel cointegration test, hence avoiding the nonstandard asymptotic distribution of Marmol and Velasco's (2004) test and allowing for vector series with components of different integration orders.

In a parametric, time domain framework, Breitung and Hassler (2002) proposed a trace test for the cointegration rank based on a generalized eigenvalue problem of the type considered by Johansen (1988, 1991). The resulting limit distribution of the statistic was found to be chi squared, where the degrees of freedom depend only on the cointegration rank under the null hypothesis. Nielsen (2005) argued that the equivalence of this regression based test and the Lagrange Multiplier (LM) test for integration does not extend to the multivariate case and showed that the actual multivariate LM test is also implicitly a test of the null of no cointegration. Breitung and Hassler (2006) considered the case where the cointegrating vector has to be estimated allowing for only one cointegration relationship. They showed that the limit distribution of the statistic is standard under the null of no cointegration, when employing the residuals from a regression in differences. Gil-Alaña (2004) extended Engle and Granger (1987)'s procedure, testing for the equality of memory parameters of the original series and of regression residuals using Robinson (1994)'s univariate test, while Nielsen (2004a) proposed a residual based LM test of the null hypothesis of cointegration assuming that the integration orders were known a priori.

Despite this effort on cointegration testing, relatively few work has been dedicated to the analysis of cointegration matrices, subspaces, and rank in fractional

systems of dimension greater than two, allowing for multiple cointegrating relationships. For stationary series, Robinson and Yajima (2002) analyzed testing procedures based on the eigenvalues of the estimated and normalized spectral density matrix around frequency zero after a preliminary step to partition the vector series into subsets with identical differencing parameters. The restriction imposed by cointegration on the spectral density matrix at zero frequency was also investigated by Nielsen (2004b) and by Nielsen and Shimotsu (2007) using alternative semiparametric memory estimates. A different route was explored by Chen and Hurvich (2003), who proposed to estimate the cointegrating relationships by the eigenvectors corresponding to the smallest eigenvalues of an averaged periodogram matrix of tapered, differenced observations. Then, Chen and Hurvich (2006) developed and justified a test for fractional cointegration and a procedure for consistently determining the number and the dimension of the cointegrating subspaces.

A further line of work has focused on several fractional generalizations of Granger (1986)'s Error Correction Model (ECM), such as Davidson (2002), who applied parametric bootstrap to testing the existence of cointegrating relationships. Lasak (2007) considers Likelihood Ratio (LR) tests in a related framework, extending original Johansen's (1988, 1991) set up to allow explicitly for fractional cointegration alternatives with stationary residuals of unknown memory.

In this paper we focus on fractional cointegration methods inspired on a further test for the integration degree proposed by Lobato and Velasco (2007). They questioned the choice of the regressor of the fractional Dickey-Fuller test of Dolado, Gonzalo, and Mayoral (2002) for the null hypothesis of unit root against the alternative of fractional unit root, and proposed an efficient version based on a different regression model. In Lobato and Velasco's (2007) basic framework, x_t is a *Type II* $I(d)$ process,

$$\Delta^d x_t 1_{\{t>0\}} = \epsilon_t, \quad t = 1, 2, \dots, \quad (1)$$

where $\epsilon_t \sim \text{iid}(0, \sigma^2)$ and the fractional difference filter $\Delta^d = (1 - L)^d$ is given by its formal expansion,

$$(1 - z)^d = \sum_{j=0}^{\infty} z^j \psi_j(d),$$

with $\psi_j(d) = \Gamma(j - d) / (\Gamma(-d)\Gamma(j + 1))$, for any real $d \neq 1, 2, \dots$, where Γ is the gamma function and $\Gamma(0)/\Gamma(0) = 1$. Suppressing the truncation in the notation, for any d , equation (1) can be rewritten as

$$\Delta x_t = \phi(\Delta^{d-1} - 1)\Delta x_t + \epsilon_t, \quad t = 1, 2, \dots, \quad (2)$$

where $\phi = 0$ under the null $H_0 : d = 1$, and $\phi = -1$ under the alternative $H_a : d < 1$. The null hypothesis is tested by means a simple one-sided t -test for

$\phi = 0$ in a rescaled regression, since $(\Delta^{\delta-1} - 1)\Delta x_t$ is uncorrelated with Δx_t for any value of δ under H_0 .

In this paper we explore a multivariate extension of Lobato and Velasco's (2007) procedure for testing the cointegrating rank in a nonstationary fractionally integrated (NFI) system. The univariate regression model (2) is replaced by an appropriate ECM to be estimated by ordinary least squares. The idea is to test whether the smallest singular values of the long run multiplier matrix estimates are significantly different from zero, exploiting the approach recently proposed by Kleibergen and Paap (2006). We derive the limit distribution that is standard in the case of "weak cointegration". It is shown that the estimation of the memory of the residuals does not affect the asymptotic properties of the statistic, neither the estimation of other short run parameters.

The plan of the article is the following. In Section 2 we propose the Wald test to determine the cointegration rank in fractional systems, adapting Kleibergen and Paap's (2006) approach. In Section 3 we show the link between our test statistic and the canonical correlation test statistic and compare it with the trace test proposed by Breitung and Hassler (2002). Section 4 proposes a generalized model accounting for more complex dynamics. The behavior of the test under local alternatives is analyzed in Section 5. In Section 6 the finite sample properties of the considered test are investigated by means of a small Monte Carlo experiment. Section 7 concludes and proposes some further lines of research. The proofs are collected in the Appendix.

Throughout this paper we shall adopt the following notation: $(A : B)$ indicates the $p \times (q + s)$ matrix obtained by placing side by side the $p \times q$ matrix A and the $p \times s$ matrix B . For $p < q$ we define $A_{\perp}$ to be a $p \times (p - q)$ matrix of rank $p - q$, for which $A'A_{\perp} = 0$. $|A|$, $\text{rank}(A)$, $\text{tr}(A)$ denote respectively the determinant, the rank and the trace of the (square) matrix A , $\|\cdot\|$ the Euclidean norm of a matrix such that $\|A\| = \sqrt{\text{tr } A'A}$, $\text{vec}(A)$ the vec operator stacking the columns of a matrix one over the other, I_p the p -rowed identity matrix. We write $A > 0$ to indicate that A is positive definite, " $\otimes$ " indicates the Kronecker product and " $:=$ " stands for equality by definition.

2 Testing the cointegration rank

In our basic framework, X_t is a fractionally (co-) integrated system generated from the DGP

$$\begin{aligned} \beta' \Delta^{d-b} X_t &= U_{1t} 1_{\{t>0\}} \\ \gamma' \Delta^d X_t &= U_{2t} 1_{\{t>0\}}, \end{aligned} \quad t = 1, 2, \dots, \quad (3)$$

where $U_t = (U'_{1t}, U'_{2t})' \sim \text{iid}(0, \Sigma_u)$, $\Sigma_u > 0$. Let β and γ be $m \times r$ and $m \times (m - r)$ matrices respectively, $0 \leq r < m$, such that $(\beta' \gamma)$ has full rank m , and $b > 0$ when $r > 0$ to identify β and γ .

Using the identity

$$\gamma_{\perp}(\beta' \gamma_{\perp})^{-1} \beta' + \beta_{\perp}(\gamma' \beta_{\perp})^{-1} \gamma' = I_m,$$

the system (3) can then be rewritten when $r > 0$ as a fractional error correction model

$$\Delta^d X_t 1_{\{t > 0\}} = \alpha \beta' (\Delta^{d-b} - \Delta^d) X_t 1_{\{t > 0\}} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (4)$$

where $\varepsilon_t \sim \text{iid}(0, \Sigma_{\varepsilon})$ is a linear and invertible transformation of the vector U_t and $\alpha = -\gamma_{\perp}(\beta' \gamma_{\perp})^{-1}$ satisfies $\beta' \alpha = -I_r$. Applying Theorem 3 in Johansen (2007a) to (4) we also obtain a moving average representation for $d > b$,

$$X_t = C \Delta^{-d} \varepsilon_t + C^* \Delta^{b-d} \varepsilon_t,$$

where $C = \beta_{\perp}(\alpha'_{\perp} \beta_{\perp})^{-1} \alpha_{\perp}$ and C^* is defined in Lemma 2 in the Appendix. Then each element $x_{j,t}$ of the vector X_t , and each cointegrating residual $\beta'_i X_t$ are respectively *type II* $I(d)$ and $I(d - b)$ processes, for $j = 1, \dots, m$, $i = 1, \dots, r$, where β_i indicates the i -th column of the matrix β . Note that if $r = 0$ then there is no cointegration and b is not identified by (3), but in this case we set its true value to 0 because any non trivial linear combination of the elements of X_t is $I(d)$.

From now on, for notational convenience, we will suppress the truncation for nonpositive t in (4), assuming implicitly that $X_t = 0$, $t \leq 0$. Moreover, we focus on the case $d = 1$, as assumed in most economic applications. To simplify the exposition, we first assume that b is known (or equal to a fixed value if testing the null that $r = 0$). We later consider the case where we use a consistent estimate of b .

The basic idea of our procedure is testing the rank of the unrestricted OLS estimation of (4), i.e. a linear regression of ΔX_t on $(\Delta^{1-b} - \Delta) X_t = (\Delta^{-b} - 1) \Delta X_t$. We note that this regressor vanishes for $b = 0$. In order to make the regressor continuous at $b = 0$, following Lobato and Velasco (2007) we employ the rescaled regression model

$$\Delta X_t = \Pi Z_{t-1}^{(b)} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (5)$$

where

$$Z_{t-1}^{(b)} := \frac{(\Delta^{-b} - 1) \Delta X_t}{b}, \quad \Pi := -(\alpha \beta') b. \quad (6)$$

For $b \rightarrow 0$, the indetermination $0/0$ in the first equation of (6) is solved using the L'Hôpital's rule, since the ratio $\lim_{b \rightarrow 0} ((1 - z)^{-b} - 1)/b$ tends to the derivative

of the fractional filter $(1 - z)^{-\delta}$ evaluated at $\delta = 0$, that is, to the linear filter $J(z) = -\log(1 - z) = \sum_{j=1}^{\infty} j^{-1} z^j$. We therefore define $Z_{t-1}^{(0)} = J(L)\Delta X_t$. If $r = 0$ (so that $b = 0$), ΔX_t is an iid sequence and hence is uncorrelated with $Z_{t-1}^{(\delta)}$ for any value of δ . The hypothesis of no cointegration can be therefore easily nested in our framework and formulated as the null $\Pi = 0$ in the regression model $\Delta X_t = \Pi Z_{t-1}^{(0)} + \varepsilon_t$.

We stress that for $b \in [0, 0.5)$ the process $Z_{t-1}^{(b)}$ is asymptotically stationary and therefore standard asymptotics apply to the unrestricted ordinary least squares estimate of Π in (5),

$$\hat{\Pi} = S_{01} S_{11}^{-1}, \quad (7)$$

where we define the following sample moments

$$S_{01} = S_{01}(b) := \frac{1}{n-2} \sum_{t=2}^n \Delta X_t Z_{t-1}^{(b)'} \quad S_{11} = S_{11}(b) := \frac{1}{n-2} \sum_{t=2}^n Z_{t-1}^{(b)} Z_{t-1}^{(b)'}$$

Allowing for $b > 0.5$, the distribution of $n^b \text{vec}(\hat{\Pi} - \Pi)$ is expressed in terms of functionals of fractional Brownian motion, as can be shown by a multivariate generalization of Theorem 1 in Dolado, Gonzalo, and Mayoral (2002). This distribution would lead to a non-pivotal statistic depending on the memory parameter b . The above arguments justify the following assumption.

ASSUMPTION 1 We assume that the DGP is given by the model (4), d is known and equal to 1, $b \in [0, 0.5)$ and $\varepsilon_t \sim \text{iid}(0, \Sigma)$ with finite fourth order moment.

Assumption 1 allows only for “weak cointegration”, which maintains the economic meaning of long-run equilibrium, implying that deviations from the equilibrium are “mean reverting” but nonstationary. In this framework, the proposed test enjoys a standard distribution and could be more powerful with respect to standard procedures designed for the $I(1)/I(0)$ case.

The rank statistic rely on the following decomposition of a generic square matrix C of dimension m ,

$$C = A_r B_r + A_{r,\perp} \Lambda_r B_{r,\perp}, \quad 0 \leq r < m, \quad (8)$$

related to the singular value decomposition (hereafter SVD) $C = U S V'$. A_r, B_r' are $m \times r$ matrices, $A_{r,\perp}, B_{r,\perp}'$ $m \times (m-r)$ matrices, satisfying the relations $A_r' A_{r,\perp} = 0$, $B_{r,\perp} B_r' = 0$, $A_{r,\perp}' A_{r,\perp} = I_{m-r}$, $B_{r,\perp} B_{r,\perp}' = I_{m-r}$. Let S_{m-r} the $m-r$ square submatrix of S , containing in its diagonal the $m-r$ smallest singular values of the matrix C . The $(m-r) \times (m-r)$ matrix Λ_r is defined as a transformation of S_{m-r} . If $\Lambda_r = 0$ and both A_r and B_r have full column rank, the rank of C is equal to r . The exact relation between the decomposition (8) and the SVD

depends on the specification of A_r and B_r , obtained by imposing a normalization for $A_r B_r$. Every normalization leads to the same expressions of Λ_r , $A_{r,\perp}$ and $B_{r,\perp}$. For simplicity, we have followed here the notation of Kleibergen and Paap (2006) as close as possible, despite that it is not standard in the cointegrating literature.

If C is a consistent estimator of an unknown matrix C_0 , Λ_r reflects the distance to rank reduction, that is, a test for $\text{rank}(C_0) = r$ will be based on a test for $H_0 : \Lambda_r = 0$. Since $\text{vec}(\Lambda_r)$ is just a rotation of $\text{vec}(S_{m-r})$ around the origin, its elements are no longer restricted to be non-negative (as the singular values are), and can be asymptotically normally distributed. See Kleibergen and van Dijk (1998) and Kleibergen and Paap (2006) for details.

If we apply the SVD directly on $\hat{\Pi}$, the resulting testing procedure could be sensitive to scaling of $\hat{\Pi}$. Kleibergen and Paap (2006) suggest to normalize the estimator $\hat{\Pi}$ before we conduct the SVD of it, in order to improve the power properties of the test. Therefore, we decompose the matrix

$$\hat{\Theta} = S_{00}^{-\frac{1}{2}} \hat{\Pi} S_{11}^{\frac{1}{2}},$$

$S_{00} := n^{-1} \sum_{t=1}^n \Delta X_t \Delta X_t'$, as

$$\hat{\Theta} = \hat{A}_r \hat{B}_r + \hat{A}_{r,\perp} \hat{\Lambda}_r \hat{B}_{r,\perp}. \quad (9)$$

Under the model (4) and Assumption 1, the probability limits of the matrices S_{00} and S_{11} , Ω_{00} and $\Omega_{11} = \Omega_{11}(b)$ respectively, have full rank, see Lemma 2 in the Appendix, so

$$\Theta := \Omega_{00}^{-\frac{1}{2}} \Pi \Omega_{11}^{\frac{1}{2}} = A_r B_r + A_{r,\perp} \Lambda_r B_{r,\perp} \quad (10)$$

is well defined and it follows that Π and Θ have the same rank. We note that up to the constant b , $A_r = \Omega_{00}^{-\frac{1}{2}} \alpha$ and $B_r = \beta' \Omega_{11}^{\frac{1}{2}}$, so that A_r and B_r can still be interpreted respectively as adjustment coefficients and cointegrating matrices α and β' . We formulate the null hypothesis

$$H_0 : \text{rank}(\Theta) = r. \quad (11)$$

Testing (11) is equivalent to test for $\Lambda_r = 0$ if and only if A_r and B_r have full rank, as assumed when (4) was derived from (3). If the full rank condition on A_r, B_r is not satisfied, the hypothesis $\Lambda_r = 0$ should be interpreted as the null that the rank of the cointegrating space is at most r , against the alternative that it exceeds r ; see the discussion in Johansen (1996), Chapter 5. The alternative hypothesis $H_a : \text{rank}(\Theta) > r$ can be therefore formulated as $\Lambda_r \neq 0$. The hypothesis of no-cointegration ($r = 0$) corresponds to the case $A_r B_r = 0$ and $A_{r,\perp} \Lambda_r B_{r,\perp} = U S V' = \Theta$. In this case we test if all the singular values of $\hat{\Theta}$ are statistically different from zero.

In order to identify $\hat{A}_r \hat{B}_r$ we impose the normalization $\hat{B}_r = [I_r : \hat{B}_{r,2}]$ as suggested by Kleibergen and Paap (2006). The limiting behavior of the different elements of $\hat{\Theta}$ in (9) is stated in Lemma 1 and Theorem 1. The proof of Lemma 1 is similar to the proof of Theorem 1 in Kleibergen and Paap (2006) and it is reported in the Appendix for completeness.

LEMMA 1. If $\hat{\Theta} \xrightarrow{p} \Theta$, then, under $H_0(r)$, $\hat{A}_r, \hat{B}_r, \hat{A}_{r,\perp}, \hat{B}_{r,\perp}$ converge in probability, respectively to $A_r, B_r, A_{r,\perp}, B_{r,\perp}$ and $\hat{\Lambda}_r \xrightarrow{p} 0$.

THEOREM 1 (*b* KNOWN). Assume that *b* is known. Under $H_0(r)$ and Assumption 1,

$$\sqrt{n} \hat{\lambda}_r \xrightarrow{d} \mathcal{N}(0, I_{(m-r)^2})$$

where $\hat{\lambda}_r = \text{vec}(\hat{\Lambda}_r)$ and $\hat{\Lambda}_r = \hat{A}'_{r,\perp} \hat{\Theta} \hat{B}'_{r,\perp}$.

The proof is in the Appendix. The main difference with the proof of Theorem 1 in Kleibergen and Paap (2006) lies in derivation of the intermediate results discussed in Remark 1 below. We use $\hat{\lambda}_r$ to define the statistic to test (11) in next corollary, whose proof is straightforward and hence omitted.

COROLLARY 1. Under Assumption 1 and $H_0(r)$ the statistic

$$\text{rk}(r) := n \hat{\lambda}'_r \hat{\lambda}_r$$

converges in distribution to a χ^2 random variable with $(m-r)^2$ degrees of freedom. Under the alternative H_a the statistic $\text{rk}(r)$ diverges to infinity at rate n .

Remark 1. Kleibergen and Paap (2006) argue in the proof of their Theorem 1 that, under the null hypothesis $H_0(r)$ and provided that $\hat{\Lambda}_r \xrightarrow{p} \Lambda_r = 0$, the normality of the (rescaled) LS estimator $\hat{\Theta}$ of Θ ,

$$\sqrt{n} \text{vec}(\hat{\Theta} - A_r B_r) = \sqrt{n} \left[\text{vec}(\hat{A}_r \hat{B}_r - A_r B_r) - \text{vec}(\hat{A}_{r,\perp} \hat{\Lambda}_r \hat{B}_{r,\perp}) \right] \xrightarrow{d} \mathcal{N}(0, \mathcal{V}),$$

implies that $\hat{A}_r \hat{B}_r$ is a root- n consistent estimator of $A_r B_r$. However this needs not to be the case because of possible cancelation between some non negligible components of the two vec terms inside the square brackets. In our proofs we use an alternative argument that only requires consistency as stated in Lemma 1.

In the remainder of this section we consider the model (5) allowing b to be unknown, whereas d remains fixed and equal to one. Assumption 1 is still valid for the “true” b . To perform inference on the rank of Π in the equation (5) we need

a consistent estimator $\hat{b}$ of b . We label $\tilde{\Pi}$ the least square estimator of Π obtained plugging in $\hat{b}$ in $Z_{t-1}^{(b)}$,

$$\tilde{\Pi} = S_{02}S_{22}^{-1},$$

where S_{02} and S_{22} are defined as S_{01} and S_{11} , respectively, with $Z_{t-1}^{(b)}$ replaced by $Z_{t-1}^{(\hat{b})}$. As done before, the SVD is implemented on the scaled matrix

$$\tilde{\Theta} = S_{00}^{-\frac{1}{2}}\tilde{\Pi}S_{22}^{\frac{1}{2}} = \tilde{A}_r\tilde{B}_r + \tilde{A}_{r,\perp}\tilde{\Lambda}_r\tilde{B}_{r,\perp}. \quad (12)$$

If $\tilde{\Pi} \xrightarrow{p} \Pi$, then $\tilde{\Lambda}_r \xrightarrow{p} \Lambda_r$ by Lemma 1, where Λ_r is given by (10) and it is equal to zero under $H_0(r)$. The following theorem shows that the first order asymptotic properties of the proposed test are not affected by the pre-estimation of the memory parameter b .

THEOREM 2 (b UNKNOWN). Let Assumption 1 hold and let the input $\hat{b}$ of $Z_{t-1}^{(\hat{b})}$ satisfy

$$\hat{b} - b = O_p(n^{-\tau}), \quad \text{with } \tau > 0, \quad \text{and } \hat{b} \in [0, 0.5). \quad (13)$$

Under $H_0(r)$

$$\sqrt{n}\tilde{\lambda}_r \xrightarrow{d} \mathcal{N}(0, I_{(m-r)^2}),$$

where $\tilde{\lambda}_r = \text{vec}(\tilde{\Lambda}_r)$ and $\tilde{\Lambda}_r = \tilde{A}_{r,\perp}'\tilde{\Theta}\tilde{B}_{r,\perp}'$. The rank statistic

$$\widetilde{\text{rk}(r)} := n\tilde{\lambda}_r'\tilde{\lambda}_r \quad (14)$$

converges in distribution to a χ^2 random variable with $(m-r)^2$ degrees of freedom. Under the alternative H_a the statistic $\widetilde{\text{rk}(r)}$ diverges to infinity at rate n .

Remark 2. If also the fractional difference parameter of the observed series d is unknown, a test could be performed by replacing it by some consistent estimator $\hat{d}$. However, using similar arguments to Breitung and Hassler (2006), it can be shown that the estimation of d may affect the limiting distribution of the test, so Theorems 1 and 2 would be no longer valid.

Remark 3. We propose to estimate b from the residuals of the univariate regression in levels

$$x_{i,t} = \Gamma'X_t^{[i]} + e_{i,t},$$

where $X_t^{[i]}$ is the $(m-1)$ -vector resulting from the deletion of the i -th component from X_t . If $r > 1$ and the observables and cointegrating residuals are purely non-stationary, Marmol and Velasco (2005) showed that, in contrast to the standard $C(1,1)$ case (see Wooldrige (1991), Johansen (2002)), the OLS estimate of the cointegrating vector $(1, \Gamma')'$ does not provide a consistent estimate of a suitable linear

combination of the cointegrating relations, though remains bounded in probability. Despite of that, in our setting of common error memory of cointegrating residuals, it was shown that the OLS residuals $\hat{e}_t$ still approximate an $I(1 - b)$ process as in the single equation set-up. See also Marmol and Velasco (2004) for a discussion.

If d is taken to be known and equal to 1, the condition $\hat{b} \in [0, 0.5)$ can be imposed naturally for implicitly defined memory estimators, such as the Gaussian semiparametric procedure of Robinson (1995) and the related Exact local Whittle procedure of Shimotsu and Phillips (2005). The upper bound implies that $Z_{t-1}^{(\hat{b})}$ is always asymptotically stationary. The lower bound is justified by the fact that order of integration of a linear combination of $I(1)$ systems cannot be greater than one, and in case of no cointegration it should be $I(1)$, being this the reason to fix the true b to 0 in this case. We do not discuss formally in this paper which estimation procedures satisfy the condition of power-rate convergence of the estimator $\hat{b}$ stated in (13). In the case of “weak cointegration”, the estimator of the cointegration relationships is not superconsistent, affecting the rate of consistency of $\hat{b}$ (see Velasco (2003)). Bias reduction techniques like higher-order kernels suggested by Hualde and Robinson (2006), should be useful to augment the speed of convergence of $\hat{b}$. Further improvements could be obtained employing spectral regression methods for the estimation of $\hat{\Gamma}$. These issues deserve further investigation.

3 A comparison with a related approach

Kleibergen and Paap (2006), Proposition 1, show that, if the covariance matrix of $\tilde{\lambda}_r$ has a Kronecker structure, $\tilde{\Pi}$ is the least square estimator and the normalization matrices are appropriately specified, the rank statistic $\widetilde{\text{rk}}(r)$ can be computed multiplying by “ n ” the smallest $m - r$ eigenvalues of $\tilde{\Theta}'\tilde{\Theta}$. It follows that $\tilde{\lambda}'\tilde{\lambda} = \sum_{j=r+1}^m \mu_j$, where $\mu_1 \geq \mu_2 \geq \dots \geq \mu_m$ are the ordered eigenvalues solving the generalized eigenvalue problem, see also Johansen (1991), p. 94,

$$|\mu S_{00} - S_{02} S_{22}^{-1} S_{20}| = 0, \quad (15)$$

where $S_{20} = S'_{02}$.

We note that our approach is not equivalent to maximum likelihood inference and therefore it cannot be considered an extension to fractional set up of Johansen (1988)’s analysis of nonstationary systems. However, an alternative proof of Theorems 1-3 could be provided adapting the proof of Theorem 11.1 and 14.4 in Johansen (1996), where likelihood ratio statistics were considered. The rank statistic (14) results directly from the decomposition (9), making clear that it can be viewed as a multivariate extension of the Wald statistic proposed by Lobato and Velasco (2007).

The approach based on the solution of (15) is a useful computational device and allows us to compare our method with the trace statistic proposed by Breitung and Hassler (2002). They test the null hypothesis that the cointegration rank is equal to r , checking if the $m - r$ smallest eigenvalues solving

$$|\mu^* M_{00} - M_{10} M_{11}^{-1} M_{10}| = 0 \quad (16)$$

are equal to zero, where

$$M_{10} = \frac{1}{n-1} \sum_{t=2}^n Y_{t-1}^* Y_t', \quad M_{11} = \frac{1}{n-1} \sum_{t=2}^n Y_{t-1}^* Y_{t-1}^{*'}, \quad M_{00} = \frac{1}{n} \sum_{t=1}^n Y_t Y_t',$$

with $Y_t := (1 - L)^d X_t$, $Y_{t-1}^* := \sum_{j=1}^{t-1} j^{-1} Y_{t-j} = J(L) Y_t$ and $M_{10} = M_{01}'$. The proposed trace statistic has the form $\mu_r(d) = n \sum_{j=r+1}^m \mu_j^*$ and it was proven that under the null it is asymptotically distributed as a χ^2 random variable with $(m-r)^2$ degrees of freedom. Nielsen (2005) showed that $\mu_r(d)$ is not a regression variant of the multivariate score statistic, as demonstrated by Breitung and Hassler (2002) in the univariate case.

The eigenvectors corresponding to the eigenvalues μ_j^* , $j = 1, \dots, m$, solving the generalized eigenvalue problem (16), can always be chosen to be orthogonal respect to M_{00} . Stated differently, given two generalized eigenvectors v_i, v_j , $i, j = 1, \dots, m$, $v_i' M_{00} v_j = 1_{\{i=j\}}$. It turns out that

$$\mu_j^* = \frac{v_j' M_{01} M_{11}^{-1} M_{10} v_j}{v_j' M_{00} v_j}, \quad j = 1, \dots, m,$$

is a test-statistic for $\phi_j = 0$ in the auxiliary regression

$$(v_j' \Delta^d X_t) = \phi_j' J(L) \Delta^d X_t + e_t, \quad t = 1, 2, \dots, n. \quad (17)$$

Therefore, $\mu_r(d)$ can be regarded as a Wald statistic rather than as a Score statistic. The test based on (17) should not reject the null hypothesis that $\phi_j = 0$ if $v_j \in \text{span}(\gamma)$, see (3), and reject it when $v_j \in \text{span}(\beta)$.

Rewriting the DGP as

$$\begin{aligned} \beta' \Delta^d X_t &= \beta' (\Delta^d - \Delta^{d-b}) X_t + U_{1t} \\ \gamma' \Delta^d X_t &= U_{2t} \end{aligned} \quad (18)$$

we note that the model (17) is misspecified because it does not include the DGP defined by (18) as a particular case under the alternative hypothesis, as pointed out by Lobato and Velasco (2007) in the univariate case. This misspecification can affect the efficiency of the resulting Wald test compared to a statistic based on the regression model

$$(v_j' \Delta^d X_t) = \varphi_j' Z_{t-1}^{(b)} + e_t, \quad t = 1, 2, \dots, n. \quad (19)$$

As noted in Section 2, $\lim_{b \rightarrow 0} Z_{t-1}^{(b)} = J(L)\Delta^d X_t$ so that models (17) and (19) are identical in this case, when the normalization S_{00} is adopted. From the proof of Theorem 1, is easy to show that, under the null, $A_{r,\perp} S_{00.1}^{\frac{1}{2}} \Delta X_t$ has a smaller variance than $A_{r,\perp} S_{00}^{\frac{1}{2}} \Delta X_t$, where $S_{00.1} = S_{00} - S_{01} S_{11}^{-1} S_{10}$. It follows that the normalization S_{00} leads to better empirical size of the test. Employing the normalization $S_{00.1}$, the asymptotic variance of the vector $\hat{\lambda}_r$ is the identity matrix both under the null and under the alternative, improving the power.

4 Short run dynamics

The model (4) was adequate to illustrate the idea behind the test procedure, but it is undoubtedly very restrictive for empirical applications. In order to allow for a richer dynamics we propose the process

$$\Phi(L)\Delta X_t = \alpha\beta'(\Delta^{-b} - 1)\Phi(L)\Delta X_t + \varepsilon_t, \quad (20)$$

where the roots of the polynomial matrix $\Phi(z) = I_m - \Phi_1 z - \dots - \Phi_p z^p$ are all outside the unit circle.

From Lemma 2 it follows that

$$X_t = \xi(1)C\Delta^{-1}\varepsilon_t + \xi(L)C^*\Delta^{-b+1}\varepsilon_t + \xi^*(L)C\varepsilon_t,$$

where $\xi(z) = \Phi(z)^{-1}$ and we exploit the well-known decomposition $\xi(z) = \xi(1) + (1-z)\xi^*(z)$. The cointegrating matrix of X_t defined by (20) is given by $\beta^* = \xi(1)^{-1'}\beta = \Phi(1)'\beta$, so the cointegration rank is preserved and depends only on $\Pi = -b(\alpha\beta')$. Now the cointegrating residuals are given by the sum on an $I(1-b)$ process plus an $I(0)$ component. For the general case $d > 1$, $\beta'\Phi(1)X_t$ is a fractionally integrated process of order $\delta^* := \max\{d-1, d-b\}$.

Equation (20) motivates a nonlinear regression model but we propose to consider the rescaled linear regression

$$\Delta X_t = \Pi^* Z_{t-1}^{(b)} + \sum_{j=1}^p \Psi_j \Delta Z_{t-j}^{(b)} + \sum_{j=1}^p \Phi_j \Delta X_{t-j} + \varepsilon_t, \quad (21)$$

where $\Phi(z) = \Phi(1) + (1-z)\Phi^*(z)$, so that $\Psi_j = \Pi\Phi_j^*$ and $\Pi^* = \Pi\Phi(1) = b(\alpha\beta^*)$. We build the rank test on the unrestricted LS estimate of Π^* in model (21). Our approach ignores the multiplicative structure of the Ψ_j , therefore it can be inefficient but keeps the test procedure simple. Rewrite the model (21) as

$$\Delta X_t = \Pi^* Z_{t-1}^{(b)} + \Gamma W_{t-1}^{(b)} + \varepsilon_t$$

with

$$\begin{aligned}\Gamma &= \begin{bmatrix} \Psi_1 \vdots \dots \vdots \Psi_p \vdots \Phi_1 \vdots \dots \vdots \Phi_p \end{bmatrix} \\ W_{t-1}^{(b)} &= \left[\Delta Z_{t-1}^{(b)'} , \dots , \Delta Z_{t-p}^{(b)'} , \Delta X_{t-1}' , \dots , \Delta X_{t-p}' \right]'.\end{aligned}$$

Consider the partitioned linear regression model

$$\Delta X_t = \Pi^* Z_{t-1}^{(\hat{b})} + \Gamma W_{t-1}^{(\hat{b})} + e_t. \quad (22)$$

Applying the *Frish-Waugh Theorem* it follows that the LS estimation of Π^* satisfies the equation

$$\Delta \bar{X}_t - \tilde{\Pi}^* \bar{Z}_{t-1}^{(\hat{b})} = \tilde{e}_t$$

where

$$\begin{aligned}\Delta \bar{X}_t &= \Delta X_t - \sum_{t=2}^n \Delta X_t W_{t-1}^{(\hat{b})'} \left(\sum_{t=2}^n W_{t-1}^{(\hat{b})} W_{t-1}^{(\hat{b})'} \right)^{-1} W_{t-1}^{(\hat{b})} \\ \bar{Z}_{t-1}^{(\hat{b})} &= Z_{t-1}^{(\hat{b})} - \sum_{t=2}^n Z_{t-1}^{(\hat{b})} W_{t-1}^{(\hat{b})'} \left(\sum_{t=2}^n W_{t-1}^{(\hat{b})} W_{t-1}^{(\hat{b})'} \right)^{-1} W_{t-1}^{(\hat{b})}.\end{aligned}$$

Define the matrices $\bar{S}_{00}$, and $\bar{S}_{22}$ as in Section 2, substituting $\Delta X_t, Z_{t-1}^{(\hat{b})}$ with $\Delta \bar{X}_t, \bar{Z}_{t-1}^{(\hat{b})}$. Then we obtain the following generalization of Theorem 2. Its proof is very similar to Theorem 2 and hence omitted.

THEOREM 3 Let Assumption 1 hold and $\hat{b}$ satisfy (13). Assume that the DGP is given by (20) and define $\tilde{\Theta} = \bar{S}_{00}^{-\frac{1}{2}} \tilde{\Pi} \bar{S}_{22}^{\frac{1}{2}}$. Then, under $H_0(r)$

$$\sqrt{n} \tilde{\lambda}_r \xrightarrow{d} \mathcal{N}(0, I_{(m-r)^2}),$$

and the rank statistic

$$\widetilde{\text{rk}(r)} := n \tilde{\lambda}_r' \tilde{\lambda}_r$$

converges in distribution to a χ^2 random variable with $(m-r)^2$ degrees of freedom. Under the alternative H_a the statistic $\widetilde{\text{rk}(r)}$ diverges to infinity at rate n .

5 Local alternatives

In this section we investigate the power properties of our test against local alternatives. We consider first a general class of alternatives local to the null of $\text{rank}(\Theta) = r \geq 0$, for a fixed value of b , and later we also reinterpret these hypotheses under the null of no cointegration, $r = 0$, as a weakly cointegrated system with $b \rightarrow 0$ with n .

Consider the general model (21), and let $s_i, i = 1, \dots, m$, be the singular values of the matrix $\Theta := \bar{\Omega}_{00}^{-\frac{1}{2}} \Pi^* \bar{\Omega}_{11}^{\frac{1}{2}}$, i.e. the limit of the coefficient $\bar{S}_{00}^{-\frac{1}{2}} \hat{\Pi} \bar{S}_{22}^{\frac{1}{2}}$ corresponding to the normalized regression (22). We define the class of local alternatives for $r = 0, 1, \dots, m-2$ and $p = 1, \dots, m-r-1$,

$$H_{1n}(r, p) : \text{diag}\{s_{r+1}, \dots, s_{r+p}\} = n^{-\frac{1}{2}} K,$$

where $K = \text{diag}\{k_1, \dots, k_p\}$, $k_i > 0$ for $i = 1, \dots, p$. The $s_1, \dots, s_r$ are positive constants and the remaining $m - (r + p)$ singular values are equal to zero. The interpretation of this parameterization is that under the local alternative $H_{1n}(r, p)$ the process has p extra cointegrating vectors, $0 < p < m - r$, that are difficult to detect, apart from the r fixed cointegrating relationships included in $H_0(r)$. We label as $\Lambda_r^{(1,n)}$ the $(m-r) \times (m-r)$ matrix Λ_r under the local alternatives $H_{1n}(r, p)$. In the next theorem we provide the asymptotic behavior of Wald tests under H_{1n} .

THEOREM 4 (LOCAL ALTERNATIVES). Let Assumption 1 hold and let the input $\hat{b}$ of $Z_{t-1}^{(\hat{b})}$ satisfy (13). Under $H_{1n}(r, p)$,

$$\sqrt{n} \tilde{\lambda}_r \xrightarrow{d} N(\xi_r, I_{m-r} \otimes I_{m-r}),$$

where $\xi_r = \lim_{n \rightarrow \infty} \sqrt{n} \lambda_r^{(1,n)}$ and $\lambda_r^{(1,n)} = \text{vec}(\Lambda_r^{(1,n)})$. The rank statistic $\widetilde{\text{rk}}(r) := n \tilde{\lambda}_r' \tilde{\lambda}_r$ converges to a noncentral $\chi_{(m-r)^2}^2$, with noncentrality parameter $\sum_{j=1}^p k_j^2$.

Remark 4. Theorem 4 shows that our Wald tests have nontrivial power under local alternatives converging to the null at a parametric rate. This result is in line with the standard asymptotics of the tests under the null. The drift depends naturally on the strength of the extra local cointegrating relationships. However this drift does not depend on the presence of extra short term lags in the model, since the local hypothesis is established on the singular values of Θ , which is already normalized by sample moments of sequences projected out on these lags. If the local hypothesis were written in terms of Π , then the probability limits of $\bar{S}_{00}$ and $\bar{S}_{22}$ will appear in the asymptotics, so that the larger (smaller) the limit of $\bar{S}_{22}$ ($\bar{S}_{00}$) more local power of the test along the usual signal-to-noise ratio tradeoff. Similarly, the larger p , i.e. the more extra cointegrating relationships in $H_{1,n}(r, p)$, more local power.

Remark 5. The effect of the parameter b on the local power is however twofold. First, we note that $\Pi^* = O(b)$, for given cointegrating and adjustment matrices. On the other hand, the variance on $Z_{t-1}^{(b)}$ depends also on b , both because normalization and fractional integration, so the overall impact factor of b on the noncentrality parameter is given by $\sum_{j=1}^{\infty} \psi_j(-b)^2$, which is increasing with b , and zero for $b = 0$.

It is also important to note that in this case the pre-estimation of b does not play a role in asymptotics, but that when $r > 0$ in $H_{1n}(r, p)$, b has to be strictly positive to give a full meaning to the (fixed) r cointegration relationships, and $\hat{b}$ estimates their cointegration degree consistently.

Alternatively, when $r = 0$, we can set up explicitly local alternatives in the cointegrating degree b , for fixed Π with n . Under the local alternatives $H_{1n}(0, p)$, $S = \text{diag}\{n^{-\frac{1}{2}}K, \mathbf{0}_{m-p}\}$, where $\mathbf{0}_{m-p}$ is a $m - p$ vector of zeros. Then, instead of considering the singular values of Θ to establish the local alternatives, we could consider the SVD $\bar{\Omega}_{00}^{-\frac{1}{2}}\alpha\beta'\bar{\Omega}_{11}^{\frac{1}{2}} = USV'$, where S and $\mathcal{S}$ are related by the following relation

$$S = \frac{b_0}{\sqrt{n}}\mathcal{S}, \quad \text{with} \quad \mathcal{S} = \text{diag}\left\{\frac{K}{b_0}, \mathbf{0}\right\},$$

and now $b = b_0/\sqrt{n} \rightarrow 0$ as $n \rightarrow \infty$, with b_0 and $\mathcal{S}$ fixed with n .

Therefore, if s_i^* , $i = 1, \dots, m$, are the singular values of the matrix $\bar{\Omega}_{00}^{-\frac{1}{2}}\alpha\beta'\bar{\Omega}_{11}^{\frac{1}{2}}$, we define the equivalent class of local alternatives

$$H_{1n}^*(0, p) : b = b_0/\sqrt{n},$$

with $s_1^*, \dots, s_p^*$ being positive constants, $s_j^* = k_j/b_0$, the remaining $m - p$ singular values in $\mathcal{S}$ being equal to zero.

Then, under $H_{1n}^*(0, p)$, $\sqrt{n}\tilde{\lambda}_0$ and the rank statistic $\widehat{\text{rk}}(0) := n\tilde{\lambda}'_0\tilde{\lambda}_0$ have the same asymptotic behavior as under $H_{1n}(0, p)$, where the noncentrality parameter can be written as $\sum_{j=1}^p k_j^2 = b_0^2 \sum_{j=1}^p s_j^{*2}$. In this case the relevant regressor is $Z_{t-1}^{(b_0/\sqrt{n})}$ which in the limit, as well as $Z_{t-1}^{(b)}$ under (13), lead to the local regressor $J(L)X_t$, the natural one to test for local alternatives around $b = 0$, and the proportionally factor $\sum_{j=0}^{\infty} j^{-2} = \pi^2/6$ in the variance matrix of $Z_{t-1}^{(0)}$.

6 Simulations

In this section we examine the finite sample performance of the proposed test by means a small Monte Carlo experiment. The data are generated according to the triangular model (see Nielsen and Shimotsu (2007))

$$\begin{cases} (1 - L)^{1-b}(X_{1t} - \Phi'X_{2t}) = \varepsilon_{1t}1_{\{t>0\}} \\ (1 - L)X_{2t} = \varepsilon_{2t}1_{\{t>0\}}, \end{cases}$$

$t = 1, \dots, n$. The dimension of the system m is set to 4 and $r_0 = 2$ cointegration relationships are imposed. The cointegrating matrix is given by $\beta = (I_2 : -\Phi)'$, with $\Phi = ((1, 0.5)' : (0.5, 1)')$ and $(\varepsilon'_{1t}, \varepsilon'_{2t})' \sim \text{iid } \mathcal{N}(0, I_4)$. Let $X_t = (X'_{1t} : X'_{2t})'$ and

x_{it} the i -th component of X_t , $i = 1, \dots, 4$. The memory parameter b is estimated from the OLS residuals of the auxiliary regression

$$x_{1t} = \Gamma' \mathbf{z}_t + e_t, \quad \text{with} \quad \mathbf{z}_t = (x_{2t}, x_{3t}, x_{4t})'$$

by means of the Exact Local Whittle (hereafter ELW) estimator by Shimotsu and Phillips (2005). More precisely we apply the ELW to the series $\hat{e}_t = x_{1t} - \hat{\Gamma}' \mathbf{z}_t$, maximizing over the compact set $\hat{\delta} \in [0.500001, 1]$, with $\delta = 1 - b$, so that $\hat{b} := 1 - \hat{\delta}$ satisfies condition (13) in Theorem 2.

Let $H_0(r)$ be the shorthand notation for null hypothesis $H_0 : \text{rank}(\Theta) = r$. Our inference procedure consist in testing $H(r) : \hat{\Lambda}_r = 0$. If $r > r_0$, then $\text{rank}(A_r B_r) < r$ and $H(r)$ is not correctly specified, that is, it is not equivalent to $H_0(r)$. To avoid this case, we suggest to perform the test $H(r)$, for $r = 0, 1, \dots, \hat{r}$, where $\hat{r} < m$ is the first value assumed by the index r for which we can not reject $H(r)$. In other words, the test rejects $H(r)$ for $r = 0, 1, \dots, \hat{r} - 1$ but not $H(\hat{r})$. $\hat{r}$ is a consistent estimator of r_0 . In our experiment we test $H(r)$, $r = 0, 1, 2$. The tables report the percentage of rejection of the null hypothesis $H(r)$ at the nominal level of 5%. The number of replications is 50,000, the sample lengths n are 100, 200, 500. The parameter b takes the values

$$b = \{0.1, 0.2, 0.3, 0.4, 0.499999\}.$$

When we test for no cointegration ($r = 0$), we also include $b = 0$. The first row of the three tables indicate the memory of the cointegrating residuals $1 - b$, approximated to the first decimal number. $\widehat{\text{BH}}$ indicate the Breitung and Hassler (2002)'s trace statistic. The rank statistic $\widehat{\text{rk}}(r)$ proposed in this paper has been computed employing different bandwidths M for the ELW estimation of b in the auxiliary univariate regression; $[x]$ denotes the largest integer smaller or equal to x . The (unfeasible) statistic $\text{rk}(r)$ is computed using the "true" b .

The test is implemented by solving the generalized eigenvalue problem (15). An alternative scaling of the matrix $\tilde{\Pi}$ could be obtained pre-multiplying this matrix by $S_{00.2}^{-\frac{1}{2}}$, with $S_{00.2} = S_{00} - S_{02} S_{22}^{-1} S_{20}$ (see Johansen (1996), p. 94). The two normalizations are asymptotically equivalent under the null, affecting the performance of the test in small samples. From simulations unreported here, it emerged that building test statistics on the matrix $S_{00.2}^{-\frac{1}{2}} \tilde{\Pi} S_{22}^{\frac{1}{2}}$ the size is upper-biased especially for $T = 100$ and large b , but power increases.

Table 1 reports the result when the hypothesis $H(0)$ of no cointegration is tested. As $b \rightarrow 0$ all the eigenvalues of Π approach zero. In this case, corresponding to the last column of the table ($1 - b = 1$), X_t can be considered as equivalent to a multivariate random walk with full rank covariance matrix and therefore the null is true. The empirical rejection frequencies of $\widehat{\text{rk}}(0)$ are above the nominal level,

because the estimation of b leads to an increase in the sampling variation of the test statistic. The distortion is higher for small sample sizes and more narrow bandwidths.

In Table 2 we consider the rejection frequencies for the hypothesis that the cointegration rank is one. The test shows good power also in small samples for $b \geq 0.3$. For $b = 1$ the rank of $A_r B_r$ is near to be zero and the power of the test is very poor for $n \leq 200$. This indicates that large samples are needed for detecting very weak cointegration relationships. Similar arguments can be used to explain the empirical size of the test, examined in Table 3. Simulations unreported here show that for $n = 1000$ the size for $1 - b = 0.9$ is still around 2%, showing that is very complicated to estimate precisely very weak cointegration relationships with small b . Finally we note that in the simple framework considered here the behavior of the proposed test is not greatly affected from the first step estimation of $\hat{b}$, at least for $n \geq 200$. Moreover its power is superior to the trace statistic by Breitung and Hassler (2002) while the size is comparable.

7 Conclusion and directions for further research

In this paper we have introduced a simple Wald test for determining the cointegration rank of a nonstationary system, allowing to the cointegrating residuals to be fractionally integrated of unknown order. The test is regression based but can be easily implemented solving a generalized eigenvalue problem of the type proposed by Johansen (1988). Many directions for further investigation could be suggested. First, we only allow for weak cointegration leading to standard asymptotics. If $b > 1/2$ then the limit distribution of the test is not standard and bootstrap techniques could be employed, following Davidson (2002, 2006). Second, we recall that we used the ELW estimator by Shimotsu and Phillips (2005) without formally proving that the assumption of the power rate consistency of $\hat{b}$ was satisfied. This issue deserve further investigation. The finite sample performance of these estimate could be improved applying spectral regression techniques to the univariate auxiliary regression. Third, in the presence of short run correlation in the ECM we propose to use a linear unrestricted estimation. However efficiency improvements could be achieve by using nonlinear estimates or two-step procedures that use the multiplicative structure of the matrix coefficients.

8 Appendix

Lemma 2

Assume that the DGP $\{X_t\}$ is given by (3). Then under Assumption 1 and $r > 0$ (so $b > 0$),

(a) $X_t = C\Delta^{-d}\varepsilon_t + C^*\Delta^{b-d}\varepsilon_t$ with

$$C = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp \quad (23)$$

$$C^* = -\bar{\beta}\bar{\alpha}' + C\bar{\beta}\bar{\alpha}' + \bar{\beta}\bar{\alpha}'C - C\bar{\beta}\bar{\alpha}'C \quad (24)$$

and $\bar{V} = V(V'V)^{-1}$ for any matrix V of full column rank.

(b) The covariance matrices

$$\Omega_{00} := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\Delta X_t \Delta X_t'), \quad \Omega_{11} = \Omega_{11}(b) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=2}^n \mathbb{E}(Z_{t-1}^{(b)} Z_{t-1}^{(b)'})$$

have full rank.

Proof of Lemma 2

We first prove (a). Following Johansen (2007b) we define the characteristic polynomial associated to the model (4)

$$\Pi^*(z) = (1-z)^d - \alpha\beta'(1-(1-z)^b)(1-z)^{d-b}$$

and the polynomial

$$\Pi(u) = (1-u)I_p - \alpha\beta'u, \quad u := 1 - (1-z)^b \quad (25)$$

so that $\Pi^*(z) = (1-z)^{d-b}\Pi(1-(1-z)^b)$.

If $r > 0$, it is easy to check that the assumptions in Theorem 8 Johansen (2007b), basically requiring that $u = 1$ is the unique unstable solution of $|\Pi(u)| = 0$ and $|\alpha_\perp\beta'_\perp| \neq 0$, are satisfied by (3).

Adapting the proof of Theorem 3 in Johansen (2007a), we rewrite (25) as $\Pi(u) = -\alpha\beta' + \dot{\Pi}(u-1)$, where $\dot{\Pi} = -(I_p + \alpha\beta')$ is the first derivative of $\Pi(u)$. Let $A = (\bar{\alpha}, \alpha_\perp)$, $B = (\bar{\beta}, \beta_\perp)$ and define

$$A'\Pi(u)B = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \dot{\Pi}_{00} & \dot{\Pi}_{01} \\ \dot{\Pi}_{10} & \dot{\Pi}_{11} \end{bmatrix} (u-1)$$

with

$$\begin{aligned} \dot{\Pi}_{00} &:= \bar{\alpha}'\dot{\Pi}\bar{\beta} = -\bar{\alpha}'\bar{\beta} - I_r & \dot{\Pi}_{01} &:= \bar{\alpha}'\dot{\Pi}\beta_\perp = -\bar{\alpha}'\beta_\perp \\ \dot{\Pi}_{10} &:= \alpha'_\perp\dot{\Pi}\bar{\beta} = -\alpha'_\perp\bar{\beta} & \dot{\Pi}_{11} &:= \alpha'_\perp\dot{\Pi}\beta_\perp = -\alpha'_\perp\beta_\perp. \end{aligned}$$

We can factorize $(u - 1)$ from the last $p - r$ columns by multiplying from the right by

$$F(u) = \begin{bmatrix} I_r & 0 \\ 0 & (u - 1)^{-1} I_{p-r} \end{bmatrix}$$

and we find that

$$K(u) := A' \Pi(z) B F(u) = \begin{bmatrix} -I_r + \dot{\Pi}_{00}(u - 1) & \dot{\Pi}_{01} \\ \dot{\Pi}_{10}(u - 1) & \dot{\Pi}_{11} \end{bmatrix}.$$

Recalling the formula for the partitioned matrix inverse we obtain that

$$K^{-1}(u) = \begin{bmatrix} -I_r & \dot{\Pi}_{01} \dot{\Pi}_{11}^{-1} \\ \dot{\Pi}_{11}^{-1} \dot{\Pi}_{10}(u - 1) & \dot{\Pi}_{11}^{-1} - \dot{\Pi}_{11}^{-1} \dot{\Pi}_{10} \dot{\Pi}_{01} \dot{\Pi}_{11}^{-1}(u - 1) \end{bmatrix}$$

because $\alpha(\beta' \alpha)^{-1} \beta' + \beta'_{\perp} (\beta'_{\perp} \alpha_{\perp})^{-1} \alpha'_{\perp} = \beta' \alpha = -I_r$. Then we compute

$$\begin{aligned} F(u) K^{-1}(u) &= \begin{bmatrix} 0 & 0 \\ 0 & (\alpha'_{\perp} \beta_{\perp})^{-1} \end{bmatrix} \frac{1}{1 - u} \\ &+ \begin{bmatrix} -I_r & \bar{\alpha}' \beta_{\perp} (\alpha'_{\perp} \beta_{\perp})^{-1} \\ (\alpha'_{\perp} \beta_{\perp})^{-1} \alpha'_{\perp} \bar{\beta} & -(\alpha'_{\perp} \beta_{\perp})^{-1} \alpha'_{\perp} \bar{\beta} \bar{\alpha}' \beta_{\perp} (\alpha'_{\perp} \beta_{\perp})^{-1} \end{bmatrix} \end{aligned}$$

so that $\Pi^{-1}(u) = B F(u) (A' \Pi(u) B F(u))^{-1} A' = B F(u) K^{-1}(u) A'$ and we obtain with some further simple manipulations the desired result for the coefficients in (23)-(24).

We prove now only the full rank of Ω_{00} in (b), the same arguments can be used to show the invertibility of Ω_{11} . Consider the representation

$$\Delta^d X_t = C \varepsilon_t + C^* \Delta^b \varepsilon_t \quad (26)$$

and the linear combinations $\beta' \Delta^d X_t$ and $\beta'_{\perp} \Delta^d X_t$, provided that $|(\beta' \beta_{\perp})| \neq 0$. First we note that $\beta'_{\perp} C \Sigma C' \beta_{\perp} > 0$ because $\beta' C \Sigma C' \beta = 0$, so that, the variance of $\beta'_{\perp} \Delta^d X_t$ is dominated by the first term in the right hand side of (26). Define $W = \Sigma^{-1/2} \alpha$, $W_{\perp} = \Sigma^{1/2} \alpha_{\perp}$. From the identity $P_W + P_{W_{\perp}} = I_m$, with $P_V = \bar{V} V'$, it follows that

$$\Sigma - \alpha(\alpha' \Sigma^{-1} \alpha)^{-1} \alpha' = \Sigma \alpha_{\perp} (\alpha'_{\perp} \Sigma \alpha_{\perp})^{-1} \alpha'_{\perp} \Sigma \geq 0$$

showing that, up to a constant given by $\sum_{j=0}^t \psi_j^2(-b)$, the dominating term of the variance of $\beta' \Delta X_t$ is given by

$$\beta' C_0^* \Sigma C_0^{*'} \beta' \geq \beta' C_0^* \alpha (\alpha' \Sigma \alpha)^{-1} \alpha' C_0^{*'} \beta = (\alpha' \Sigma \alpha) > 0$$

because $\beta' C_0^* \alpha = -I_r$.

8.1 Proof of Lemma 1

If $\hat{\Theta}$ is a consistent estimator of Θ and null hypothesis $H_0 : \text{rank}(\Theta) = r$ holds true, then

$$\|\hat{\Theta} - A_r B_r\| \xrightarrow{p} 0, \quad (27)$$

where A_r and B_r are of full rank r . Therefore, pre- and post-multiplying $\hat{\Theta} - \Theta$ by the matrices $(\hat{A}_r' : \hat{A}_{r,\perp}')'$ and $(\hat{B}_r' : \hat{B}_{r,\perp}')$ we obtain

$$\mathbf{R} = \begin{bmatrix} \hat{A}_r'(\hat{A}_r \hat{B}_r - A_r B_r) \hat{B}_r' & -\hat{A}_r'(A_r B_r) \hat{B}_{r,\perp}' \\ -\hat{A}_{r,\perp}'(A_r B_r) \hat{B}_r' & \hat{A}_r - \hat{A}_{r,\perp}'(A_r B_r) \hat{B}_{r,\perp}' \end{bmatrix} \xrightarrow{p} 0. \quad (28)$$

We label $\mathbf{R}_{ij}$, $i, j = 1, 2$ the sub-matrices obtained partitioning the matrix $\mathbf{R}$ as in (28). The singular values of $\hat{\Theta}$ are equal to $\sqrt{\text{eig}(\hat{\Theta}'\hat{\Theta})}$ and the eigenvalues are continuous function of $\hat{\Theta}$. Therefore Slutsky's theorem implies that the smallest $m - r$ singular values of $\hat{\Theta}$ converge in probability to the $m - r$ null singular values of $\Theta = A_r B_r$, that is $\hat{A}_r \xrightarrow{p} 0$. It follows that also the second component of the submatrix $\mathbf{R}_{22}$, i.e. $\hat{A}_{r,\perp}'(A_r B_r) \hat{B}_{r,\perp}'$, converges in probability to the null matrix. Considering jointly $\mathbf{R}_{12}$ and $\mathbf{R}_{22}$ we have

$$\hat{A}_r'(A_r B_r) \hat{B}_{r,\perp}' \xrightarrow{p} 0, \quad \hat{A}_{r,\perp}'(A_r B_r) \hat{B}_{r,\perp}' \xrightarrow{p} 0$$

allowing us to conclude that $B_r \hat{B}_{r,\perp}' \xrightarrow{p} 0$, since it is not possible that both $\hat{A}_r' A_r \xrightarrow{p} 0$ and $\hat{A}_{r,\perp}' A_r \xrightarrow{p} 0$ hold true. Then $\hat{B}_{r,\perp} \xrightarrow{p} B_{r,\perp}$, and this implies, directly from the identity

$$\hat{B}_{r,\perp} \hat{B}_r' \equiv 0 \Leftrightarrow (\hat{B}_{r,\perp} - B_{r,\perp}) B_r' \equiv \hat{B}_{r,\perp} (B_r - \hat{B}_r)'$$

so that also $\hat{B}_r \xrightarrow{p} B_r$.

The consistency of $\hat{A}_r$ and $\hat{A}_{r,\perp}$ can be derived by the same arguments, considering the blocks $\mathbf{R}_{21}$ and $\mathbf{R}_{22}$ and the identity $\hat{A}_r' \hat{A}_{r,\perp} \equiv 0$.

8.2 Proof of Theorem 1

As discussed in Section 3, the proof of Theorem 1 we provide is not a simple application to fractional cointegration of Kleibergen and Paap's (2006) Theorem 1, which relies on root- n consistency of $\hat{A}_r$ and $\hat{B}_r$. We instead only use consistency of $\hat{A}_r$ and $\hat{B}_r$.

Under Assumption 1, ΔX_t and $Z_{t-1}^{(b)}$ are asymptotically stationary and using a standard central limit theorem for martingale difference sequences it follows that

$$\sqrt{n} \text{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} \mathcal{N}(0, \Omega_{11}^{-1} \otimes \Sigma_\varepsilon).$$

Using the decomposition $\hat{\Theta} - \Theta = \hat{A}_r \hat{B}_r - A_r B_r + \hat{A}_{r,\perp} \hat{\Lambda}_r \hat{B}_{r,\perp}$, and recalling that $\Theta = \Omega_{00}^{-\frac{1}{2}} \Pi \Omega_{11}^{\frac{1}{2}}$, we have that under $H_0(r)$

$$\sqrt{n} \text{vec} \left(\hat{A}_r \hat{B}_r - A_r B_r \right) + \sqrt{n} \text{vec} \left(\hat{A}_{r,\perp} \hat{\Lambda}_r \hat{B}_{r,\perp} \right) \xrightarrow{d} N(0, \mathcal{V}) \quad (29)$$

where $\mathcal{V} = (\Omega_{11}^{\frac{1}{2}} \otimes \Omega_{00}^{-\frac{1}{2}})(\Omega_{11}^{-1} \otimes \Sigma_\varepsilon)(\Omega_{11}^{\frac{1}{2}} \otimes \Omega_{00}^{-\frac{1}{2}})' = I_m \otimes \Omega_{00}^{-\frac{1}{2}} \Sigma_\varepsilon \Omega_{00}^{-\frac{1}{2}}$, with

$$\begin{aligned} \Omega_{00}^{-\frac{1}{2}} \Sigma_\varepsilon \Omega_{00}^{-\frac{1}{2}} &= \Omega_{00}^{-\frac{1}{2}} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E(Y_t - \Pi Z_{t-1}^{(b)})(Y_t - \Pi Z_{t-1}^{(b)})' \Omega_{00}^{-\frac{1}{2}} \\ &= I_m - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E \left(\Omega_{00}^{-\frac{1}{2}} Y_t Z_{t-1}^{(b)'} \Omega_{11}^{-\frac{1}{2}} \Theta' - \Theta \Omega_{11}^{-\frac{1}{2}} Z_{t-1}^{(b)} Y_t' \Omega_{00}^{-\frac{1}{2}} \right) + \Theta \Theta', \end{aligned}$$

so that $A_{r,\perp}' \Omega_{00}^{-\frac{1}{2}} \Sigma_\varepsilon \Omega_{00}^{-\frac{1}{2}} A_{r,\perp} = A_{r,\perp}' A_{r,\perp} = I_m$ because $A_{r,\perp}' \Theta = 0$ under $H_0(r)$. Using that $A_r' A_{r,\perp} \equiv 0$, $\hat{B}_{r,\perp} \hat{B}_{r,\perp}' = 0$, $A_{r,\perp}' A_{r,\perp} = I_{m-r}$, $\hat{\Theta} - \Theta = O_p(n^{-1/2})$ and the consistency of $\hat{A}_{r,\perp}$ from (29), we get

$$\sqrt{n} \text{vec} \left\{ A_{r,\perp}' (\hat{\Theta} - \Theta) \hat{B}_{r,\perp}' \right\} = (1 + o_p(1)) \sqrt{n} \hat{\lambda} \xrightarrow{d} \mathcal{N}(0, I_{m-r} \otimes I_{m-r}),$$

proving the theorem. $\square$

Proof of Theorem 2

We first show the consistency of the estimator. Consider the regression model

$$\Delta X_t = \Pi Z_{t-1}^{(b)} + \varepsilon_t = \Pi Z_{t-1}^{(\hat{b})} + \Pi(Z_{t-1}^{(b)} - Z_{t-1}^{(\hat{b})}) + \varepsilon_t. \quad (30)$$

In order to show that $\tilde{\Pi}$ is consistent, we rewrite $\tilde{\Pi} - \Pi$ as

$$= \Pi \sum_{t=2}^n \left\{ \frac{(\Delta^{-b} - \Delta^{-\hat{b}})}{b} \Delta X_t \right\} Z_{t-1}^{(\hat{b})'} \left(\sum_{t=2}^n Z_{t-1}^{(\hat{b})} Z_{t-1}^{(\hat{b})'} \right)^{-1} + O_p(n^{-\tau}) \quad (31)$$

$$+ \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})} \left(\sum_{t=2}^n Z_{t-1}^{(\hat{b})} Z_{t-1}^{(\hat{b})'} \right)^{-1}. \quad (32)$$

We consider the single terms distinctly, and firstly (31). We first note that

$$b Z_{t-1}^{(b)} - \hat{b} Z_{t-1}^{(\hat{b})} = (\Delta^{-b} - \Delta^{-\hat{b}}) \Delta X_t = \sum_{j=1}^{t-1} \left\{ \psi_j(b) - \psi_j(\hat{b}) \right\} \Delta X_{t-j}. \quad (33)$$

Following Lobato and Velasco (2007) and Robinson and Hualde (2003), Proposition 9, we note that, for $j = 1, 2, \dots, n$, the expression $\left\{ \psi_j(b) - \psi_j(\hat{b}) \right\}$ in (33) equals to

$$\sum_{q=1}^{Q-1} \frac{1}{q!} (b - \hat{b})^q \psi_j^{(q)}(b) + \frac{1}{Q} (b - \hat{b})^Q \psi_j^{(Q)}(\bar{b}) \quad (34)$$

where, $\bar{b}$ is an intermediate stochastic point between b and $\hat{b}$ and $\psi_j^{(q)}(b) = d^q \psi_j(b)/db^q$. It follows that

$$\begin{aligned} & \frac{1}{n} \sum_{t=2}^n (\Delta^{-b} - \Delta^{-\hat{b}}) \Delta X_t Z_{t-1}^{(\hat{b})'} = \frac{1}{n} \sum_{t=2}^n \left(\sum_{j=1}^{t-1} (\psi_j(b) - \psi_j(\hat{b})) \Delta X_{t-j} \right) Z_{t-1}^{(\hat{b})'} \\ &= \sum_{q=1}^{Q-1} \frac{1}{q!} (\hat{b} - b)^q \left[\frac{1}{n} \sum_{t=2}^n \left(\sum_{j=1}^{t-1} \psi_j^{(q)}(b) \Delta X_{t-j} \right) Z_{t-1}^{(\hat{b})'} \right] \end{aligned} \quad (35)$$

$$+ \frac{1}{Q!} (\hat{b} - b)^Q \frac{1}{n} \left(\sum_{t=2}^n \sum_{j=1}^{t-1} \psi_j^{(Q)}(\bar{b}) \Delta X_{t-j} Z_{t-1}^{(\hat{b})'} \right) \quad (36)$$

Since $|\psi^{(q)}(b)| \leq C(1+j)^{b-1}(\log(1+j))^q$, $j = 1, 2, \dots, n$, $1 \leq q \leq Q$ (see Robinson and Hualde (2003), Lemma D3) the sequence $\psi_j^{(q)}(b)$ is square summable when $b < 1/2$, then the term in square brackets in (35) is $O_p(1)$. Moreover, for $\hat{b} < 1/2$ $n^{-1} \sum_{t=2}^n Z_{t-1}^{(\hat{b})} Z_{t-1}^{(\hat{b})'} = O_p(1)$, then if $\hat{b} - b = O_p(n^{-\tau})$, $\tau > 0$, (35) is $O_p(n^{-\tau})$.

In order to analyze (36), we focus on the term:

$$\frac{1}{n} \sum_{t=2}^n \sum_{j=1}^{t-1} \psi_j^{(Q)}(\bar{b}) \Delta X_{t-j}^{(\kappa)} Z_{t-1}^{(\hat{b}, \ell)} \leq \frac{1}{n} \sqrt{\sum_{t=2}^n \left| \sum_{j=1}^{t-1} \psi_j^{(Q)}(\bar{b}) \Delta X_{t-j}^{(\kappa)} \right|^2} \sqrt{\sum_{t=2}^n Z_{t-1}^{(\hat{b}, \ell)^2}}$$

where $\Delta X_{t-j}^{(\kappa)}, Z_{t-1}^{(\hat{b}, \ell)}$ are, respectively, the κ -th and ℓ -th elements of the vectors ΔX_{t-j} and $Z_{t-1}^{(\hat{b})}$. The first term of the product in the right hand side of the above equation can be bounded by

$$\begin{aligned} & \sup_t \left| \sum_{j=1}^{t-1} \psi_j^{(Q)}(\bar{b}) \Delta X_{t-j}^{(\kappa)} \right| \leq \sum_{j=1}^n \left| \psi_j^{(Q)}(\bar{b}) \right| \left| \Delta X_{t-j}^{(\kappa)} \right| \\ & \leq \left(\sum_{j=1}^n \left| \psi_j^{(Q)}(\bar{b}) \right|^2 \sum_{j=1}^n \left| \Delta X_{t-j}^{(\kappa)} \right|^2 \right)^{1/2}. \end{aligned} \quad (37)$$

By Lemma D.5 of Robinson and Hualde (2003), for any $\epsilon > 0$,

$$\left| \psi_j^{(Q)}(\bar{b}) \right| = O_p \left((j+1)^{b+\epsilon-1} (\log(j+1))^Q \right), \quad \text{as } n \rightarrow \infty \quad (38)$$

so that

$$\sum_{j=1}^n \left| \psi_j^{(Q)}(\bar{b}) \right|^2 \leq C n^{2b+2\epsilon-1} (\log n)^Q.$$

Noting that $\sum_{j=1}^n (\Delta X_{t-j}^{(\kappa)})^2 = O_p(n)$, the term (37) is $O_p(n^{b+\epsilon} (\log n)^Q)$ and taking Q large enough we can make (36) $o_p(n^{-\tau})$, since $n^{-1} \sum_{t=2}^n |Z_{t-1}^{(\hat{b}, \ell)}|^2$ is $O_p(1)$.

It remains to analyze the asymptotic behavior of the term (32), and, in particular, the term $\sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'}$

$$\sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} = \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(b)'} + \left(\sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} - \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(b)'} \right). \quad (39)$$

Scaling by $\sqrt{n}$ the term in brackets in the above expression, we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \left(\sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} - \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(b)'} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{t=2}^n \varepsilon_t \sum_{j=1}^{t-1} \left\{ \sum_{q=1}^{Q-1} \frac{1}{q!} (b - \hat{b})^q \psi_j^{(q)}(b) \right\} \Delta X_{t-j} \end{aligned} \quad (40)$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=2}^n \varepsilon_t \sum_{j=1}^{t-1} \left\{ \frac{1}{Q!} (b - \hat{b})^Q \psi_j^{(Q)}(\bar{b}) \right\} \Delta X_{t-j}, \quad (41)$$

where

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=2}^n \varepsilon_t \sum_{j=1}^{t-1} \frac{1}{q!} \psi_j^{(q)}(b) \Delta X'_{t-j} &= O_p(1), \quad \text{for } q = 1, 2, \dots, Q-1 \\ \frac{1}{\sqrt{n}} \sum_{t=2}^n \varepsilon_t \sum_{j=1}^{t-1} \left\{ \frac{1}{Q!} \psi_j^{(Q)}(\bar{b}) \right\} \Delta X'_{t-j} &= O_p(n^{b+\epsilon+1/2} (\log n)^Q). \end{aligned}$$

For Q large enough, (41) is $O_p(1)$ and the higher order terms in (40) are $O_p(n^{-\tau})$. Therefore (32) is $o_p(1)$, proving that $\tilde{\Theta} \xrightarrow{p} \Theta$, so that Lemma 2 applies.

From the first order condition for the test least squares estimator $\tilde{\Pi}$, we obtain

$$\begin{aligned} 0 &= \sum_{t=2}^n (\Delta X_t - \tilde{\Pi} Z_{t-1}^{(\hat{b})'}) Z_{t-1}^{(\hat{b})'} \Leftrightarrow \\ 0 &= \sum_{t=2}^n (S_{00}^{-\frac{1}{2}} \Delta X_t - \tilde{\Theta} S_{22}^{-\frac{1}{2}} Z_{t-1}^{(\hat{b})}) Z_{t-1}^{(\hat{b})'} S_{22}^{-\frac{1}{2}}. \end{aligned} \quad (42)$$

Given the regression model $S_{00}^{-\frac{1}{2}} \Delta X_t = S_{00}^{-\frac{1}{2}} \Pi S_{22}^{\frac{1}{2}} S_{22}^{-\frac{1}{2}} Z_{t-1}^{(b)} + S_{00}^{-\frac{1}{2}} \varepsilon_t$, (42) can therefore be written as

$$0 = \sum_{t=2}^n (\Theta^{(n)} S_{22}^{-\frac{1}{2}} Z_{t-1}^{(b)} + S_{00}^{-\frac{1}{2}} \varepsilon_t - \tilde{\Theta} S_{22}^{-\frac{1}{2}} Z_{t-1}^{(\hat{b})}) Z_{t-1}^{(\hat{b})'} S_{22}^{-\frac{1}{2}}$$

with $\Theta^{(n)} = S_{00}^{-\frac{1}{2}} \Pi S_{22}^{\frac{1}{2}}$ and $\lim_{n \rightarrow \infty} \Theta^{(n)} = \Theta$. If S_{00} and S_{22} has full rank, considering Θ or $\Theta^{(n)}$ makes no difference to our aim, because under the null both have

exactly rank r . With a slight abuse of notation, we will use the notation Θ for both, without making further distinctions. Under $H_0(r) : \Lambda = 0$, $\Theta = A_r B_r$; using the decomposition (12) we get

$$\begin{aligned} \tilde{A}_{r,\perp} \tilde{\Lambda}_r \tilde{B}_{r,\perp} S_{22}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(\hat{b})} Z_{t-1}^{(\hat{b})'} \right) S_{22}^{-\frac{1}{2}} &= A_r B_r S_{22}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(b)} Z_{t-1}^{(b)'} \right) S_{22}^{-\frac{1}{2}} \\ &- \tilde{A}_r \tilde{B}_r S_{22}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(\hat{b})} Z_{t-1}^{(\hat{b})'} \right) S_{22}^{-\frac{1}{2}} + S_{00}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} \right) S_{22}^{-\frac{1}{2}} \end{aligned}$$

that simplifies to

$$\tilde{A}_{r,\perp} \tilde{\Lambda}_r \tilde{B}_{r,\perp} = A_r B_r S_{22}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(b)} Z_{t-1}^{(b)'} \right) S_{22}^{-\frac{1}{2}} - \tilde{A}_r \tilde{B}_r + S_{00} \left(\frac{1}{n} \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} \right) S_{22}^{-\frac{1}{2}}. \quad (43)$$

Pre-multiplying the above equation by $A'_{r,\perp}$, post-multiplying by $\tilde{B}'_{r,\perp}$ and recalling that $A_{r,\perp} = \tilde{A}_{r,\perp} + O_p(n^{-\tau})$, we obtain

$$\sqrt{n} \tilde{\lambda} = (1 + O_p(n^{-\tau}))^{-1} \left(\tilde{B}_{r,\perp} S_{22}^{-\frac{1}{2}} \otimes A'_{r,\perp} S_{00}^{-\frac{1}{2}} \right) \left(\frac{1}{\sqrt{n}} \text{vec} \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} \right).$$

Using (40)-(41) it follows that

$$\frac{1}{\sqrt{n}} \text{vec} \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} \xrightarrow{d} N(0, \Omega_{11} \otimes \Sigma_\varepsilon),$$

whereas, the proof that $\lim_{n \rightarrow \infty} S_{22} = \Omega_{11}$ is based on the result (34). Proceeding as in the proof of Theorem 1, it is easy to check that, as $n \rightarrow \infty$

$$\sqrt{n} \tilde{\lambda} \xrightarrow{d} N(0, I_{m-r} \otimes I_{m-r}),$$

proving the theorem. $\square$

8.3 Proof of Theorem 4

We consider first the case without lags. Under local alternative H_{1n} , equation (43) should be substituted by

$$\begin{aligned} \tilde{A}_{r,\perp} \tilde{\Lambda}_r \tilde{B}_{r,\perp} &= \left(A_r B_r + A_{r,\perp} \Lambda_p^{(1)} B_{r,\perp} \right) S_{22}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(b)} Z_{t-1}^{(b)'} \right) S_{22}^{-\frac{1}{2}} \\ &- \tilde{A}_r \tilde{B}_r + S_{00}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} \right) S_{22}^{-\frac{1}{2}}. \end{aligned}$$

Pre- and post- multiplying the above equation, respectively by $A'_{r,\perp}$, by $\tilde{B}'_{r,\perp}$ and taking the vec we obtain

$$\begin{aligned} (1 + o_p(1))\sqrt{n}\tilde{\lambda}_r - \left[\tilde{B}_{r,\perp} S_{22}^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(\hat{b})} Z_{t-1}^{(b)'} \right) S_{22}^{-\frac{1}{2}} B'_{r,\perp} \otimes I \right] \sqrt{n}\lambda_r^{(1,n)} \quad (44) \\ = \left(\tilde{B}_{r,\perp} S_{22}^{-\frac{1}{2}} \otimes A'_{r,\perp} S_{00}^{-\frac{1}{2}} \right) \left(\frac{1}{\sqrt{n}} \text{vec} \sum_{t=2}^n \varepsilon_t Z_{t-1}^{(\hat{b})'} \right). \end{aligned}$$

Rewriting $Z_{t-1}^{(b)} = Z_{t-1}^{(\hat{b})} - (Z_{t-1}^{(\hat{b})} - Z_{t-1}^{(b)})$, equation (44) can be stated as

$$(1 + o_p(1))\sqrt{n}\tilde{\lambda} - \left[\tilde{B}_{r,\perp} B'_{r,\perp} \otimes I \right] \sqrt{n}\lambda_r^{(1,n)} \quad (45)$$

$$+ \left\{ \tilde{B}_{r,\perp} S_{22}^{-\frac{1}{2}} \left[\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(\hat{b})} (Z_{t-1}^{(\hat{b})} - Z_{t-1}^{(b)})' \right] S_{22}^{-\frac{1}{2}} B'_{r,\perp} \otimes I \right\} \sqrt{n}\lambda_r^{(1,n)}. \quad (46)$$

Provided that

$$\frac{1}{n} \sum_{t=2}^n Z_{t-1}^{(\hat{b})} (Z_{t-1}^{(\hat{b})} - Z_{t-1}^{(b)})' = o_p(1), \quad \text{and} \quad \sqrt{n}\lambda_r^{(1,n)} = O(1) \quad (47)$$

equation (46) is $o_p(1)$ and the theorem follows because

$$\begin{aligned} \lambda_r^{(1,n)'} \lambda_r^{(1,n)} &= \text{vec}(\Lambda_r^{(1,n)})' \text{vec}(\Lambda_r^{(1,n)}) = \text{tr}(\Lambda_r^{(1,n)} \Lambda_r^{(1,n)'}) \\ &= \text{tr} \left[(U_{22} U_{22}')^{-1/2} U_{22} S_2 U_{22}' (U_{22} U_{22}')^{-1/2} \right] \\ &= \text{tr}(S_2^2) = n^{-1} \sum_{j=1}^p k_j^2. \end{aligned}$$

When we have lags, the first order condition is given by

$$\begin{aligned} \sum_{t=2}^n \left(\bar{S}_{00}^{-\frac{1}{2}} \Delta X_t - \tilde{\Theta} \bar{S}_{22}^{-\frac{1}{2}} \bar{Z}_{t-1}^{(\hat{b})} \right) \bar{Z}_{t-1}^{(\hat{b})'} \bar{S}_{22}^{-\frac{1}{2}} = 0 \iff \\ \sum_{t=2}^n \left(\Theta \bar{S}_{22}^{-\frac{1}{2}} \Phi(L) Z_{t-1}^{(b)} + \sum_j \bar{S}_{00}^{-\frac{1}{2}} \Phi_j \Delta X_{t-j} + \bar{S}_{00}^{-\frac{1}{2}} \varepsilon_t - \tilde{\Theta} \bar{S}_{22}^{-\frac{1}{2}} \bar{Z}_{t-1}^{(\hat{b})} \right) \bar{Z}_{t-1}^{(\hat{b})'} \bar{S}_{22}^{-\frac{1}{2}} = 0, \end{aligned}$$

where the terms ΔX_{t-j} and $\bar{Z}_{t-1}^{(\hat{b})}$ are orthogonal. Rewriting

$$\Theta \bar{S}_{22}^{-\frac{1}{2}} \Phi(L) Z_{t-1}^{(b)} = \Theta \bar{S}_{22}^{-\frac{1}{2}} Z_{t-1}^{(b)} + \sum_{j=1}^p \bar{S}_{22}^{-\frac{1}{2}} \Phi_j \Delta Z_{t-j}^{(b)},$$

$\bar{S}_{22}^{-\frac{1}{2}} \Phi_j \sum_t \left(\Delta Z_{t-j}^{(b)} \bar{Z}_{t-1}^{(\hat{b})'} \right) \bar{S}_{22}^{-\frac{1}{2}}$ should be $o_p(1)$ using arguments similar to (44). Finally we have

$$\sum_t \left(\Theta \bar{S}_{22}^{-\frac{1}{2}} Z_{t-1}^{(b)} \bar{Z}_{t-1}^{(\hat{b})'} \bar{S}_{22}^{-\frac{1}{2}} \right) = \sum_t \left(\Theta \bar{S}_{22}^{-\frac{1}{2}} Z_{t-1}^{(\hat{b})} \bar{Z}_{t-1}^{(\hat{b})'} \bar{S} \right) + o_p(1)$$

where $\sum_t \left(Z_{t-1}^{(\hat{b})} \bar{Z}_{t-1}^{(\hat{b})'} \right) = \sum_t \left(\bar{Z}_{t-1}^{(\hat{b})} \bar{Z}_{t-1}^{(\hat{b})'} \right)$, and this shows that the introduction of the short run dynamics does not change the drift.

1-b		0.5	0.6	0.7	0.8	0.9	1
n=100							
BH		99.83	96.91	76.61	35.96	10.59	4.83
$\widetilde{\text{rk}}(0)$	$M = [n^{0.55}]$	100	99.28	86.43	46.56	16.20	7.94
	$M = [n^{0.65}]$	100.	99.20	85.90	45.13	14.96	7.01
	$M = [n^{0.75}]$	99.99	98.96	84.27	42.68	13.34	6.01
rk(0)	(true b)	100	99.30	84.30	40.57	11.43	4.83
n=200							
BH		100	100	99.63	80.57	21.49	4.84
$\widetilde{\text{rk}}(0)$	$M = [n^{0.55}]$	100	100	99.92	86.14	26.97	6.79
	$M = [n^{0.65}]$	100	100	99.90	85.39	25.52	5.96
	$M = [n^{0.75}]$	100	100	99.88	84.31	24.07	5.43
rk(0)	(true b)	100	100	99.92	84.16	22.65	4.84
n=500							
BH		100	100	100	99.98	62.73	4.77
$\widetilde{\text{rk}}(0)$	$M = [n^{0.55}]$	100	100	100	99.99	66.40	5.53
	$M = [n^{0.65}]$	100	100	100	99.99	65.79	5.23
	$M = [n^{0.75}]$	100	100	100	99.99	64.97	5.04
rk(0)	(true b)	100	100	100	99.99	64.19	4.77

Table 1: Power of the test. Rejection frequencies (in %) of the null hypothesis of no-cointegration ($H_0 : \text{rank}(\Theta) = 0$). $1 - b$ is the memory of the cointegrating residuals, so $1 - b = 0$ corresponds to the null hypothesis because $r = 0$, and the displayed frequencies in this case refer to size.

1-b		0.5	0.6	0.7	0.8	0.9
n=100						
BH		79.10	49.81	19.48	4.24	0.61
$\widetilde{\text{rk}}(1)$	$M = [n^{0.55}]$	93.51	68.20	29.60	7.02	1.25
	$M = [n^{0.65}]$	93.52	68.03	29.09	6.60	1.17
	$M = [n^{0.75}]$	92.99	66.49	27.54	5.86	0.88
rk(1)	(true b)	94.29	66.93	26.44	5.18	0.71
n=200						
BH		99.96	98.07	75.60	22.40	1.67
$\widetilde{\text{rk}}(1)$	$M = [n^{0.55}]$	100	99.71	85.49	28.67	2.50
	$M = [n^{0.65}]$	100	99.73	85.13	27.91	2.22
	$M = [n^{0.75}]$	100	99.65	84.25	26.74	2.02
rk(1)	(true b)	100	99.75	84.70	26.13	1.80
n=500						
BH		100	100	100	91.40	11.38
$\widetilde{\text{rk}}(1)$	$M = [n^{0.55}]$	100	100	100	94.08	12.90
	$M = [n^{0.65}]$	100	100	100	94.13	12.67
	$M = [n^{0.75}]$	100	100	100	93.87	12.42
rk(1)	(true b)	100	100	100	94.04	12.11

Table 2: Power of the test. Rejection frequencies (in %) of the null hypothesis of one cointegration relationship ($H_0 : \text{rank}(\Theta) = 1$). $1 - b$ is the memory of the cointegrating residuals.

1-b		0.5	0.6	0.7	0.8	0.9
n=100						
BH		3.16	2.24	1.14	0.26	0.05
$\widetilde{\text{rk}}(2)$	$M = [n^{0.55}]$	4.79	3.56	1.86	0.52	0.08
	$M = [n^{0.65}]$	4.78	3.52	1.81	0.49	0.06
	$M = [n^{0.75}]$	4.72	3.37	1.66	0.43	0.05
rk(2)	(true b)	4.98	3.35	1.56	0.36	0.05
n=200						
BH		4.24	3.89	3.01	1.18	0.12
$\widetilde{\text{rk}}(2)$	$M = [n^{0.55}]$	5.38	5.14	3.89	1.63	0.17
	$M = [n^{0.65}]$	5.44	5.06	3.79	1.56	0.14
	$M = [n^{0.75}]$	5.39	4.90	3.60	1.47	0.13
rk(2)	(true b)	5.57	4.95	3.52	1.41	0.13
n=500						
BH		4.74	4.70	4.16	3.35	0.63
$\widetilde{\text{rk}}(2)$	$M = [n^{0.55}]$	5.43	5.26	4.68	3.75	0.69
	$M = [n^{0.65}]$	5.47	5.25	4.54	3.69	0.67
	$M = [n^{0.75}]$	5.48	5.11	4.46	3.62	0.63
rk(2)	(true b)	5.57	5.11	4.44	3.57	0.65

Table 3: Size of the test. Rejection frequencies (in %) of the null hypothesis of two cointegration relationship ($H_0 : \text{rank}(\Theta) = 2$). $1 - b$ is the memory of the cointegrating residuals.

References

- BREITUNG, J., AND U. HASSLER (2002): “Inference on the cointegration rank in fractionally integrated process,” *Journal of Econometrics*, 110, 167–185.
- (2006): “A residual-based LM type test against fractional cointegration,” *Econometric Theory*, 22(6), 1091–1111.
- CHEN, W. W., AND C. M. HURVICH (2003): “Semiparametric estimation of multivariate fractional cointegration,” *Journal of American Statistical Association*, 98(463), 629–642.
- (2006): “Semiparametric estimation of fractional cointegrating subspaces,” *Annals of Statistic*, 34(6), 2939–2979.

- DAVIDSON, J. (2002): “A model of fractional cointegration, and test for cointegration using the bootstrap,” *Journal of Econometrics*, 110(2), 187–212.
- (2006): “Alternative bootstrap procedures for testing cointegration in fractionally integrated processes,” *Journal of Econometrics*, 133(2), 741–777.
- DOLADO, J. J., J. GONZALO, AND L. MAYORAL (2002): “A fractional Dickey-Fuller test for unit roots,” *Econometrica*, 70(5), 1963–2006.
- ENGLE, R., AND C. W. J. GRANGER (1987): “Cointegration and error correction: representation, estimation and testing,” *Econometrica*, 55(2), 251–276.
- GIL-ALANÍA, L. (2004): “Testing of fractional cointegration in macroeconomic time series,” *Oxford Bulletin of Economics and Statistics*, 65, 517–529.
- GRANGER, C. W. J. (1986): “Developments in the study of cointegrated economic variables,” *Oxford Bulletin of Economics and Statistics*, (48), 213–228.
- HUALDE, J., AND P. M. ROBINSON (2006): “Semiparametric estimation of fractional cointegration,” Working Paper 07/06, Universidad de Navarra.
- HUALDE, J., AND C. VELASCO (2007): “Distribution-Free test of fractional cointegration,” *Econometric Theory*, 23, forthcoming.
- JOHANSEN, S. (1988): “Statistical analysis of cointegration vectors,” *Journal of Economic Dynamics and Control*, 12, 231–254.
- (1991): “Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models,” *Econometrica*, 59, 1551–1580.
- (1996): *Likelihood based inference on cointegrated vector autoregressive model*. Oxford University Press.
- (2002): “Cointegration in partial systems and efficiency of single equation analysis,” *Journal of Econometrics*, 52, 389–452.
- (2007a): “Representation of cointegrated autoregressive processes with application to fractional cointegration,” *Econometric Reviews*, forthcoming.
- (2007b): “A representation theory for a class of vector autoregressive models for fractional processes,” *Econometric Theory*, forthcoming.
- KLEIBERGEN, F., AND R. PAAP (2006): “Generalized reduced rank tests using singular value decomposition,” *Journal of Econometrics*, 133, 97–126.

- KLEIBERGEN, F., AND H. K. VAN DIJK (1998): “Bayesian simultaneous equations analysis using reduced rank structures,” *Econometric Theory*, 14, 701–743.
- LASAK, K. (2007): “Likelihood based testing for fractional cointegration,” UAB.
- LOBATO, I., AND C. VELASCO (2007): “Efficient Wald test for fractional unit roots,” *Econometrica*, 75(2), 575–589.
- MARINUCCI, D., AND P. M. . ROBINSON (2001): “Semiparametric fractional cointegration analysis,” *Journal of Econometrics*, 105(1), 225–247.
- MARMOL, F., AND C. VELASCO (2004): “Consistent testing of cointegrating relationships,” *Econometrica*, 72(6), 1809–1844.
- (2005): “Consistent testing of cointegration of unknown rank,” Universidad Carlos III de Madrid. Preprint.
- NIELSEN, M. Ø. (2004a): “Optimal residual-based test for fractional cointegration and exchange rate dynamics,” *Journal of Business Economic Statistics*, 22, 331–345ne.
- (2004b): “Spectral analysis of fractionally cointegrated systems,” *Economics Letters*, 83, 225–231.
- (2005): “Multivariate Lagrange multiplier test for fractional integration,” *Journal of Financial Econometrics*, 3(3), 372–398.
- NIELSEN, M. Ø., AND K. SHIMOTSU (2007): “Determining the cointegrating rank in nonstationary fractional systems by exact local Whittle approach,” *Journal of Econometrics*, forthcoming.
- ROBINSON, P. M. (1994): “Efficient tests of nonstationary hypotheses,” *Journal of the American Statistical Association*, 89, 1420–1437.
- (1995): “Gaussian semiparametric estimation of long range dependence,” *Annals of Statistic*, 23(5), 1630 – 1661.
- (2007): “Diagnostic testing for cointegration,” *Journal of Econometrics*, forthcoming.
- ROBINSON, P. M., AND J. HUALDE (2003): “Cointegration in fractional systems with unknown integration orders,” *Econometrica*, 71(6), 1727–1766.
- ROBINSON, P. M., AND D. MARINUCCI (2003): “Semiparametric frequency domain analysis of fractional cointegration,” in *Time series with long memory*, ed. by P. M. Robinson, pp. 334–373. Oxford University Press.

- ROBINSON, P. M., AND Y. YAJIMA (2002): “Determination of cointegrating rank in fractional systems,” *Journal of Econometrics*, 106, 217–241.
- SHIMOTSU, K., AND P. C. B. PHILLIPS (2005): “Exact local Whittle estimation of fractional integration,” *The Annals of Statistics*, 33(4), 1890–1933.
- VELASCO, C. (2003): “Gaussian semiparametric estimation of fractional cointegration,” *Journal of Time Series Analysis*, 24, 345–378.
- WOOLDRIGE, J. (1991): “Notes on regression with difference-stationary data,” Preprint, Michigan State University.