Fractional cointegration rank estimation

Katarzyna Łasak
CREATES
School of Economics and Management
Aarhus University
Building 1322, Bartholins Allé 10
DK-8000 Aarhus C, Denmark
E-mail: klasak@creates.au.dk

Carlos Velasco
Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid 126
28903 Getafe, Spain
E-mail: carlos.velasco@uc3m.es

November 3, 2010

Abstract

We consider the problem of cointegration rank estimation in the framework of fractional Vector Error Correction Mechanism (FVECM). We describe and compare different methods available up to date, namely four LR tests based on different assumptions on the model and a new two-step procedure. In the new two-step procedure, the first step consists in estimation of the FVECM under the null hypothesis of cointegration rank $r = r_0$. This provides consistent estimates of the cointegration degree d, cointegration vectors β and speed of the adjustment to the equilibrium parameters α and also (super) consistent estimates of $\beta_\perp$, orthogonal to β, such that $\beta_\perp' X_t$ is not cointegrated in any direction. In the second step, taking $\hat{\beta}_\perp$ as given, we propose to implement the sup tests considered in Łasak (2010), that are based on the $p - r_0$ vector series $\hat{\beta}_\perp' X_t$, in this case reestimating d again. We analyse the performance of the proposed new procedures in finite samples and compare them with all the LR tests we discuss. These include cases when the cointegration degree is unknown and estimated under the null or under the alternative.

Further we propose a procedure to detect extra cointegrating relations with different memory, which can be seen as generalization of Johansen and Nielsen's (2010b) test. This procedure also uses the sup norm in the spirit of Łasak's (2010) cointegration test.

Keywords: Error correction model, Gaussian VAR model, Maximum likelihood estimation, Likelihood ratio tests, Fractional cointegration rank

JEL: C12, C15, C32.

1 Introduction

Cointegration is commonly thought of as a stationary relation between nonstationary variables. It has become a standard tool in econometrics since the seminal paper of Granger (1981). Following the initial suggestion of Engle and Granger (1987), when the series of interest are $I(1)$, testing for cointegration in a single-equation framework can be conducted by means of residual based tests (cf. Phillips and Ouliaris (1990)). Residual-based tests rely on initial regressions among the levels of the relevant time series. They are designed to test the null of no cointegration by testing whether there is a unit root in the residuals against the alternative that the regression errors are $I(0)$.

Alternatively fully parametric inference on $I(1)/I(0)$ cointegrated systems in the framework of Error Correction Mechanism (ECM) representation has been developed by Johansen (1988, 1991, 1995). He suggests a maximum likelihood procedure based on reduced rank regressions. His methodology consists in identifying the number of cointegration vectors within the VAR by means of performing a sequence of likelihood ratio tests. If the variables are cointegrated, after selecting the rank, the cointegration vectors, the speed of adjustment to the equilibrium coefficients and short-run dynamics are estimated. So-called Johansen's procedure can be preferred to the residual-based approach because it provides a simple way of testing for the cointegration rank and making inference on the parameters of complex cointegrated systems.

However the assumption that deviations from equilibrium are integrated of order zero is far too restrictive. In a general set up, where errors with fractional degree of integration are allowed, it is possible to permit the cointegration residuals to be integrated of order greater than zero. The case of fractionally cointegrated processes has the same economic implications, i.e. exist long-run equilibrium among variables, as when the processes are integer-valued cointegrated, except for the slower rate of convergence to the equilibrium in the former situation. Since a standard setup of $I(1)/I(0)$ cointegrated systems ignores the fractional cointegration parameter, a fractionally integrated equilibrium error will result in a misspecified likelihood function.

There has been relatively few other work dedicated to inference on cointegration rank in fractional systems. Breitung and Hassler (2002) suggest a new variant of efficient score tests against fractional alternatives for univariate time series that generalizes to multivariate cointegration tests. It allows to determine the cointegration rank of fractionally integrated time series by solving a generalized eigenvalue problem of the type proposed by Johansen (1988). Robinson and Yajima (2002) develop methods of investigating the existence and extent of cointegration in fractionally integrated systems with stationary series, in semiparametric setting, with some discussion of extension to nonstationarity.

Nielsen and Shimotsu (2007) propose to extend the cointegration rank determination procedure of Robinson and Yajima (2002) to accommodate both (asymptotically) stationary and nonstationary fractionally integrated processes as the common stochastic trends and cointegrating errors by applying the exact local Whittle analysis of Shimotsu and Phillips (2005). The proposed method estimates the cointegrating rank by examining the rank of the spectral density matrix of the d-th differenced process around the origin, where the fractional integration order d is estimated by the exact local Whittle estimator. Their method does not require estimation of the cointegrating vector(s) and is easier to implement than regression-based approaches, but it only provides a consistent estimate of the cointegration rank, and formal tests of the cointegration rank or levels of confidence are not available except for the special case of no cointegration.

Chen and Hurvich (2003b) consider the semiparametric estimation of fractional cointegration in a multivariate process of cointegrating rank $r > 0$. They estimate the cointegrating relationships by the eigenvectors corresponding to the r smallest eigenvalues of an averaged periodogram matrix of tapered, differenced observations. They determine the rate of convergence of the r smallest eigenvalues of the periodogram matrix and present a criterion that allows for consistent estimation of r. Chen and Hurvich (2006) consider a common-components model for multivariate fractional cointegration, in which different memory of components is allowed and the cointegrating rank may exceed one. They decompose the true cointegrating vectors into orthogonal fractional cointegrating subspaces and estimate each cointegrating subspace separately, using appropriate sets of eigenvectors of an averaged periodogram matrix of tapered, differenced observations, based on the first m Fourier frequencies, with m fixed. They obtain a consistent and asymptotically normal estimate of the memory parameter for the given cointegrating subspace and then they use these estimates to test for fractional cointegration and to consistently identify the cointegrating subspaces.

Our approach in this paper is fully parametric instead, based on the specification of a fractional VECM. We analyse different procedures to estimate the cointegration rank of fractionally cointegrated system. Note that in a fractional framework LR tests for cointegration rank loose their straightforward asymptotic properties since we may obtain different estimates for cointegration degree under the null and under the alternative. We propose to perform a sequence of tests based on a new two stage procedure whose motivation comes from the results on cointegration testing in Łasak (2010) and estimation of fractionally cointegration systems in Łasak (2008). We also consider applying sup tests proposed in Łasak (2010) in a naive way although the same asymptotic inference may not be valid, since in case of testing the cointegration rank the cointegration degree parameter is identified both under the null and under the alternative. This method implies estimation of cointegration degree under the alternative. Further we consider estimation of cointegration degree under the null, which lead us to the construction of a LR test similar to Lyhagen's (1998) trace test in some sense. Further we consider LR tests based on the standard VECM that assumes that the degree of cointegration is known and equal to one, like in Johansen (1988, 1991, 1995), as a benchmark case for comparison. Finally we propose a procedure to detect extra cointegrating relations with different memory, which can be seen as generalization of Johansen and Nielsen's (2010b) test. This procedure also uses the sup norm in the spirit of Łasak's (2010) cointegration test.

The rest of the paper is organized as follows. Section 2 describes ML analysis of fractional system. Section 3 discusses the problem of testing for cointegration. Section 4 presents the problem of rank testing in the fractional framework. Section 5 describes the new two-step procedure. In section 6 we present rank testing in the model with short run dynamics. Section 7 discusses testing procedure in case we have cointegrating relations with different memory. Section 8 presents results of Monte Carlo analysis. Section 9 concludes. Appendix A contains the proof of the fact that replacing $\beta_\perp$ by $\hat{\beta}_\perp$ makes no difference asymptotically in LR test statistics, which we use in our analysis. Appendix B provides theoretical justification of the procedure described in Section 7.

2 Analysis of the fractional system

As a first natural research step we consider the simplest version of the fractional VECM, model without lagged differences and deterministic terms, that is a special case of fractional representations proposed in Granger (1986), Johansen (2008, 2009) and Avarucci (2007),

$$\Delta^{\delta} X_t = \alpha\beta'(\Delta^{-d} - 1)\Delta^{\delta} X_t + \varepsilon_t, \tag{1}$$

where X_t is a $p \times 1$ vector of variables fractionally integrated of order δ and
$\Delta = 1 - L$, L being the lag operator. We assume there exist r linear combinations β of original variables X_t that are of order $\delta - d$, where r is a cointegration rank, however r is unknown. α is a $p \times r$ matrix of the speed of the adjustment to the equilibrium coefficients, ε_t is a $p \times 1$ vector of Gaussian errors with variance-covariance matrix Ω. Note that we assume the Gaussianity of the errors only to define the likelihood function.

We consider the case $\delta_0 = 1$ to easy the notation and we set the true value of d to $d_0 \in (0.5, 1]$ when $r > 0$. Note that when $r > 1$ then all cointegrating relationships have the same order of integration, $1 - d_0$. δ could be estimated consistently if unknown, either from univariate ML from components of X_t or by ML estimation under the null hypothesis of cointegration with rank $r_0 > 0$.

In order to estimate parameters of the fractionally cointegrated system given by (1) we can follow the procedure described in Johansen (1995), but adjusted for the case of fractional VECM that has been already presented in Łasak (2010) and 3. Let's define $Z_{0t} = \Delta X_t$, $Z_{1t}(d) = \left(\Delta^{1-d} - \Delta\right) X_t$. The model expressed in these variables becomes

$$Z_{0t} = \alpha\beta' Z_{1t}(d) + \varepsilon_t, \quad t = 1, ..., T.$$

The log-likelihood function apart from a constant for the model (1) is given by

$$L(\alpha, \beta, \Omega, d) = -\frac{1}{2}T \log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}[Z_{0t} - \alpha\beta' Z_{1t}(d)]'\Omega^{-1}[Z_{0t} - \alpha\beta' Z_{1t}(d)].$$

Define as well

$$S_{ij}(d) = T^{-1}\sum_{t=1}^{T} Z_{it}(d) Z_{jt}(d)' \quad i, j = 0, 1,$$

and note that S_{ij} do not depend on d when $i = j = 0$.
For fixed d and β, parameters α and Ω are estimated by regressing Z_{0t} on $\beta' Z_{1t}(d)$ and

$$\hat{\alpha}(\beta(d)) = S_{01}(d)\beta(\beta' S_{11}(d)\beta)^{-1} \tag{2}$$

while

$$\hat{\Omega}(\beta(d)) = S_{00} - S_{01}(d)\beta(\beta' S_{11}(d)\beta)^{-1}\beta' S_{10}(d) = S_{00} - \hat{\alpha}(\beta)(\beta' S_{11}(d)\beta)\hat{\alpha}(\beta)'. \tag{3}$$

Plugging these estimates into the likelihood we get

$$L_{\max}^{-2/T}(\hat{\alpha}(\beta(d)), \beta, \hat{\Omega}(\beta(d)), d) = L_{\max}^{-2/T}(\beta, d) = |S_{00} - S_{01}(d)\beta(\beta' S_{11}(d)\beta)^{-1}\beta' S_{10}(d)|,$$

and finally the maximum of the likelihood is obtained by solving the following eigenvalue problem

$$\left|\lambda(d) S_{11}(d) - S_{10}(d) S_{00}^{-1} S_{01}(d)\right| = 0 \tag{4}$$

for eigenvalues $\lambda_i(d)$ and eigenvectors $\upsilon_i(d)$, for a given d, such that

$$\lambda_i(d)S_{11}(d)\upsilon_i(d) = S_{10}(d)S_{00}^{-1}S_{01}(d)\upsilon_i(d),$$

and $\upsilon_j'(d)S_{11}(d)\upsilon_i(d) = 1$ if $i = j$ and 0 otherwise. Note that the eigenvectors diagonalize the matrix $S_{10}(d)S_{00}^{-1}S_{01}(d)$ since

$$\upsilon_j'(d)S_{10}(d)S_{00}^{-1}S_{01}(d)\upsilon_i(d) = \lambda_i(d)$$

if $i = j$ and 0 otherwise. Thus by simultaneously diagonalizing the matrices $S_{11}(d)$ and $S_{10}(d)S_{00}^{-1}S_{01}(d)$ we can estimate the $r-$dimensional cointegrating space as the space spanned by the eigenvectors corresponding to the r largest eigenvalues.
With this choice of β we can estimate d by maximizing the log-likelihood, i.e.

$$\tilde{d} = \arg\max_{d \in \mathcal{D}} L_{\max}(d), \tag{5}$$

where

$$L_{\max}(d) = \left[|S_{00}| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i(d)\right) \right]^{-\frac{T}{2}} \tag{6}$$

and $\mathcal{D} \subset (0.5, 1]$. We have solved the problem of testing whether the system (1) is cointegrated in Łasak (2010). Further we have analyzed the estimation of the system (1) in Łasak (2008) under the assumption that the cointegration rank has been known already. In this paper we concentrate on the problem how to establish the cointegration rank. We follow the methods developed in Johansen (1988, 1991, 1995) of using the sequence of likelihood ratio tests to test the null hypothesis $H_0 : r = 1$ first, then $H_0 : r = 2$, till we cannot reject the null. We discuss the problem in details in the next two sections.

3 Testing for cointegration

Łasak (2010) have proposed two sup tests to test the null of no cointegration against two different alternatives. Using sup trace we test the null hypothesis of no cointegration,

$$H_0 : rank\ (\Pi) = r_0 = 0,$$

against the alternative of the full rank of the impact matrix Π,

$$H_1 : rank\ (\Pi) = p.$$

Note that in case we reject the null hypothesis we only get the information that system is cointegrated, but we do not know how many cointegration relations has X_t. The LR statistic for testing H_0 against H_1 is defined by

$$\mathcal{LR}_T\,(p) = trace(\hat{d}_p) = -2\ln\left[LR\,(0|p)\right] = -T\sum_{i=1}^{p} \ln[1 - \hat{\lambda}_i(\hat{d}_p)], \tag{7}$$

where

$$\hat{d}_p = \underset{d\in\mathcal{D}}{\arg\max} L_p(d) = \underset{d\in\mathcal{D}}{\arg\max}\ trace(d)$$

and L_p denotes the likelihood under the hypothesis of rank p.

Alternatively we can use sup maximum eigenvalue test and test the null hypothesis of no cointegration

$$H_0 : rank\ (\Pi) = r_0 = 0,$$

against the alternative of the cointegration with rank 1

$$H_1 : rank\ (\Pi) = 1.$$

Note again that in case we reject the null hypothesis we only get the information that the system is cointegrated, however the cointegration rank can be different than 1. The LR statistic for this case is defined by

$$\mathcal{LR}_T\,(1) = \lambda_{\max}(\hat{d}_1) = -2\ln\left[LR\,(0|1)\right] = -T\ln[1-\lambda_1(\hat{d}_1)], \tag{8}$$

where

$$\hat{d}_1 = \underset{d\in\mathcal{D}}{\arg\max} L_1(d) = \underset{d\in\mathcal{D}}{\arg\max} \lambda_{\max}(d)$$

and L_1 denotes the likelihood under the hypothesis of rank 1.

Recall that under the null of no cointegration ($r_0 = 0$) we cannot hope that $\hat{d}_1$ or $\hat{d}_p$ estimate consistently a nonexisting true value of d, and because of that tests (7) and (8) could be interpreted as sup LR tests, in the spirit of Davies (1977) and Hansen (1996).

In Łasak (2010) we demonstrated that

$$\mathcal{LR}_T\,(p) = trace(\hat{d}_p) \xrightarrow{d} \sup_{d\in\mathcal{D}} trace\left[\pounds(d)\right] = J_p \tag{9}$$

and

$$\mathcal{LR}_T\,(1) = \lambda\max(\hat{d}_1) \xrightarrow{d} \sup_{d\in\mathcal{D}} \lambda_1\left[\pounds(d)\right] = E_p \tag{10}$$

where $\mathcal{D} \subset (0.5, 1]$ is a compact set,

$$\pounds(d) = \int_0^1 (dB_d)\, B_d' \left[\int_0^1 B_d B_d' du\right]^{-1} \int_0^1 B_d\,(dB_d)',$$

and B_d is a p-dimensional standard fractional Brownian motion with parameter $d \in (0.5, 1]$,

$$B_d\,(x) = \Gamma^{-1}\,(d) \int_0^x (x-z)^{d-1}\, dB\,(z)\,,$$

B being standard Brownian motion.

We might apply directly sup tests for testing the rank $r > 0$. However in Łasak (2008) it is found that under the null hypothesis H_0 of the positive cointegration rank $r = r_0 > 0$, $\hat{d}_{r_0}$ is root-T consistent and asymptotically normal. By contrast, when the null hypothesis H_0 is true, $\hat{d}_r$ computed under the alternative for some $r > r_0 > 0$, can be expected either to be consistent for d, though with a different asymptotic distribution, or to be random, but with an asymptotic distribution that would depend on the data generating process (as happens when $r_0 = 0$). Our simulations support the first situation, which is consistent with the fact that in this case d is actually identified (by the $r_0 > 0$ cointegration relationships).

The second possibility based on testing the significance of the eigenvalues $\lambda_{r_0+1}\left(\hat{d}_r\right), \ldots, \lambda_p\left(\hat{d}_r\right)$, $r > r_0$, would render non pivotal asymptotic distributions and no feasible critical values tabulation.

4 Testing the cointegration rank

Recall we consider the problem of rank estimation in the fractionally cointegrated model (1), when d_0 is unknown. The main idea is to establish the cointegration rank using a sequence of the likelihood ratio tests, estimating d in every step of the sequence. We can perform the sequence of trace tests[1], where we test the null hypothesis H_0 of the cointegration rank r_0, where $r_0 = 1$, 2, etc.

$$H_0 : rank\ (\Pi) = r_0 > 0,$$

against the alternative hypothesis H_1 of the full rank of the impact matrix Π

$$H_1 : rank\ (\Pi) = p.$$

Note that the hypothesis of the full rank of the impact matrix Π means that the VAR is stationary in levels rather than that we have cointegration in the system. However if we reject the null of a certain number of cointegrating vectors we move to the next step in the sequence of tests rather than accept the information given by the alternative hypothesis.

Another option we consider is to perform a sequence of the maximum eigenvalue tests[2], where we test the null hypothesis H_0 of the cointegration rank r_0, where $r_0 = 1$, 2, etc.

$$H_0 : rank\ (\Pi) = r_0 > 0,$$

against the alternative hypothesis H_1 of the cointegration rank $r_1 = r_0 + 1$,

$$H_1 : rank\ (\Pi) = r_0 + 1.$$

Recall that again we do not get complete information about the rank by a separate usage of one test from the sequence. Only if at a certain stage we cannot reject the null hypothesis of the cointegration rank r_0 we can interpret the result as an information about the rank. If we reject also the null hypothesis of cointegration rank $r_0 = p - 1$ then the system was stationary in levels rather than cointegrated.

Note that in general LR tests for testing the cointegration rank r_0 against r_1 can be derived based on the solutions of the eigenvalue problem (4) as

$$\mathcal{LR}_T\left(r_0|r_1\right) = -2\ln\left[LR\left(r_0|r_1\right)\right] = -T\left\{\begin{array}{c} |S_{00}(\hat{d}_{r_1})| + \sum_{i=1}^{r_1}\ln[1-\hat{\lambda}_i(\hat{d}_{r_1})] \\ -|S_{00}(\hat{d}_{r_0})| - \sum_{i=1}^{r_0}\ln[1-\hat{\lambda}_i(\hat{d}_{r_0})] \end{array}\right\}, \tag{11}$$

where estimates of the cointegration degree under the null $(\hat{d}_{r_0})$ and under the alternative $(\hat{d}_{r_1})$

[1] Note that we denote by trace test every LR test with the alternative hypothesis of the full rank of the matrix Π.

[2] Note that we denote by maximum eigenvalue test every LR test with the cointegration rank under the alternative hypothesis higher by 1 than under the null

will be in general different. However test statistic $\mathcal{LR}_T(r_0|r_1)$ has unknown so far asymptotic distribution which could be difficult to derive and hardly useful in practice.

The inference would simplify if we decide to use a common estimate $\hat{d} = \hat{d}_{r_1} = \hat{d}_{r_0}$ for both of them. Such an assumption seem not to influence the generality of our results in any sense, since the model (1) we consider assumes that all the cointegrating relations share the same memory. However we could choose whether to estimate the cointegration degree under the null or under the alternative hypothesis.

We consider both estimating d under the null and under the alternative. Note that Lyhagen (1998) has tabulated the asymptotic distribution of the trace test statistic in a fractional framework under the assumption that we know the true cointegration degree d_0[3]. Our first proposal is to apply his results for the case when we do not know d and to pre-estimate d under the null hypothesis and use this estimate ($\hat{d}_{r_0}$) as the true value of the cointegration degree d_0. Then for this kind of test we can use the critical values tabulated in Lyhagen (1998), since $\hat{d}_{r_0} \to_p d_0$ under the null H_0. To easy the notation we will call this test as Lyhagen's trace test.

Our second proposal is to estimate d under the alternative hypothesis so that we could extrapolate naively the asymptotic distributions of the sup tests derived in Łasak (2010) to test also for higher ranks. We call these tests naive sup trace and naive sup maximum eigenvalue test further in the paper.

In Section 4.7 we check and compare by Monte Carlo simulation the finite samples performance of Lyhagen's trace, naive sup tests, LR tests based on the standard VECM (called as Johansen's tests to easy the notation) and a new two-step procedure that we propose and describe in the next section.

5 Two-step procedure to establish the rank

In this section we propose a new two-step procedure to establish the cointegration rank. The first step consists in the estimation of the model (1) under the null hypothesis H_0 of cointegration rank $r = r_0$. This provides consistent estimates of d, Π and of the decomposition $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices. Then we can compute (super) consistent estimates of $\beta_\perp$, orthogonal to β, so that $\beta'_\perp X_t$ is not cointegrated (in any direction). Further, taking $\hat{\beta}_\perp$ as given, we propose to implement sup tests, described in Section 4.3, based on the $p - r_0$ vector series $\hat{\beta}'_\perp X_t$. In this case d is reestimated again, by contrast with the alternative procedures that would fix $d = \hat{d}$ from a first step. Then the sup tests statistics would be compared to critical values from the E_{p-r_0} and J_{p-r_0} distributions (see (9) and (10)) and given the superconsistency of $\hat{\beta}$ and therefore of $\hat{\beta}_\perp$ there should be no estimation effects for the first stage. Under the alternative $\hat{\beta}'_\perp X_t$ contains at least one further cointegrating relationship and the sup tests should be able to detect it consistently.

This procedure has two potential drawbacks. First, it ignores the information on the true d provided by $\hat{d}_{r_0}$. Second, under the null $\beta'_\perp X_t$ is a $(p - r_0) \times 1$ vector of not cointegrated $I(1)$ series, but they are not pure $I(1)$ processes as (1) would indicate for X_t when $rank(\Pi) = 0$, but are contaminated by the r_0 -rank cointegrating residuals, which are $I(1-d)$ series. Therefore, test procedures should take into account this new feature of the data under (1). We discuss this issue

[3] And the same did Johansen and Nielsen in JN (2008), although they obtained slightly different critical values. So the behaviour of the so-called by us Lyhagen's test will us the flavour of the behaviour of JN(2008) LR test with an extra step, where the cointegration degree is estimated under the null.

with two examples.

Example 1 ***Triangular model.***
Consider a triangular representation of cointegrated $I(1)$ series with rank r, with conformable partition of matrices with $X_t' = (y_t', x_t')$ and $\beta' = \begin{pmatrix} I_r & -\gamma \end{pmatrix}$, all matrices with full rank,

$$\begin{pmatrix} I_r & -\gamma \\ 0 & I_{p-r} \end{pmatrix} \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \Delta^{d-1} & 0 \\ 0 & \Delta^{-1} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Then we have that

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} I_r & \gamma \\ 0 & I_{p-r} \end{pmatrix} \begin{pmatrix} \Delta^{d-1}\varepsilon_{1t} \\ \Delta^{-1}\varepsilon_{2t} \end{pmatrix}$$

and therefore

$$\begin{aligned} \beta_\perp' X_t &= \beta_\perp' \begin{pmatrix} I_r & \gamma \\ 0 & I_{p-r} \end{pmatrix} \begin{pmatrix} \Delta^{d-1}\varepsilon_{1t} \\ \Delta^{-1}\varepsilon_{2t} \end{pmatrix} \\ &= \begin{pmatrix} M_1 & M_2 \end{pmatrix} \begin{pmatrix} \Delta^{d-1}\varepsilon_{1t} \\ \Delta^{-1}\varepsilon_{2t} \end{pmatrix}, \end{aligned}$$

where M_2 is a matrix with full rank and therefore there is no b such that $b'\left(\beta_\perp' X_t\right)$ is an $I(1-d)$ process, a process less integrated than $\beta_\perp' X_t$. As far as $M_1 \neq 0$ we can see that $\beta_\perp' X_t$ contains some $I(1-d)$ terms, by contrast with (1) when $r = 0$ and $\Pi = 0$. The interesting feature is that these $I(1-d)$ terms are spanned by $\Delta^{d-1}\varepsilon_{1t}$, which are the cointegrating residuals of $\beta' X_t$.

Looking at the representation of $\beta_\perp' \Delta X_t$, the substitution of the past values of $M_1 \Delta^{d_0} \varepsilon_{1t}$, i.e. $M_1 \left(\Delta^{d_0} - 1\right) \varepsilon_{1t}$ by $\beta' \Delta X_{t-1} = \Delta^{d_0} \varepsilon_{1t-1}$ amounts to comparing the sequences $\sum_{j=1}^{t} \pi_j(d_0)\, \varepsilon_{1t-j}$ and $\sum_{j=1}^{t} \pi_{j-1}(d_0)\, \varepsilon_{1t-j}$, which are also present in the discussion of Lobato and Velasco (2007). The main difference is that $\pi_j(d_0) < 0$ for $j > 0$, whereas $\pi_0(d_0) = 1$, so some innovations have reversed sign in each series, despite the weights being asymptotically similar for large j.

Example 2 ***General model.***
From (1) we can write the following representation using Theorem 8 of Johansen (2008)

$$X_t = C\Delta^{-\delta}\varepsilon_t + \Delta^{d-\delta} Y_t^+$$

where $C = \beta_\perp \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp'$ and $Y_t^+ \sim I(0)$. Then

$$\beta_\perp' X_t = C^* \Delta^{-\delta}\varepsilon_t + \Delta^{d-\delta} \beta_\perp' Y_t^+$$

where $C^* = \beta_\perp' C$ is full rank under the null hypothesis H_0, so that $\beta_\perp' X_t$ is $I(\delta)$ and not cointegrated. However the term $\Delta^{d-\delta}\beta_\perp' Y_t^+$ is $I(\delta - d)$ and since $\beta' X_t$ is also $I(\delta - d)$ we could use $\hat{\beta}' X_t$ as a proxy for this term.

To capture this effect at $I(1-d)$, $\delta = 1$ level when testing for $r = r_0 > 1$ we propose to estimate either the following fractional error correction model

$$\Delta \hat{\beta}_\perp' X_t = ab'(1 - \Delta^{-d})\Delta \hat{\beta}_\perp' X_t + \gamma \left(\hat{\beta}' \Delta X_{t-1}\right) + e_t,$$

or alternatively

$$\Delta\hat{\beta}_{\perp}^{\prime}X_t = ab^{\prime}(1-\Delta^{-d})\Delta\hat{\beta}_{\perp}^{\prime}X_t + \gamma\Delta X_{t-1} + e_t,$$

where $\hat{\beta}$ is the ML estimate of β under the null hypothesis of the cointegration rank $r = r_0$, and $\hat{\beta}_{\perp}$ is in its null space. The second equation, considering ΔX_{t-1} as an additional regressor instead of the increments of the cointegration residuals, $\hat{\beta}^{\prime}\Delta X_{t-1}$, takes into account both the directions in $\hat{\beta}^{\prime}$ and in $\hat{\beta}_{\perp}^{\prime}$. This may improve size, but of course can have effects on power, because of a possible correlation of ΔX_{t-1} with the regressor $(1-\Delta^{-d})\Delta\hat{\beta}_{\perp}^{\prime}X_t$.

Then the test statistics are $\mathcal{LR}_T(1)$ and $\mathcal{LR}_T(p-r_0)$, where we replace X_t by $\hat{\beta}_{\perp}^{\prime}X_t$ and we prefilter the involved series with a regression on $\hat{\beta}^{\prime}\Delta X_t$ or ΔX_t. Our proposal is to approximate the asymptotic distribution of these test statistics by E_{p-r_0} and J_{p-r_0} respectively since we check that replacing $\beta_{\perp}$ by $\hat{\beta}_{\perp}^{\prime}$ has no asymptotic impact on the test statistics under the assumption that $\hat{\beta}_{\perp} - \beta_{\perp} = O_p\left(T^{-d_0}\right)$, see Appendix A.

6 Rank testing in ECM with short run noise

Following Avarucci (2007), we allow now for short run correlation in the fractional cointegration relationship and in the levels and use the model

$$\Delta^{\delta}X_t = \Pi(\Delta^{-d}-1)A(L)\Delta^{\delta}X_t + (I - A(L))\Delta^{\delta}X_t + \varepsilon_t, \tag{12}$$

where $A(L) = I - A_1L - \cdots - A_kL^k$. This model can be showed to encompass triangular models used in the literature (cf. Robinson and Hualde (2003)) and has nice representations if the roots of the equation $|A(z)| = 0$ are out of the unit circle, $\delta > d$. Basically this model implies that there is fractional cointegration among the prewhitened series $X_t^{\dagger} = A(L)X_t$, for which (1) holds. It can be seen as a multivariate extension of Hualde and Robinson's (2007) bivariate cointegrated model. In fact, if $X_t^{\dagger}$ is cointegrated with cointegrating vector β, X_t is also cointegrated, with cointegrating vector $\beta^* = A(1)^{\prime}\beta$, using the representation $A(L)^{-1} = A(1)^{-1} + \Delta\tilde{A}(L)$, with $\sum_{j=1}^{\infty}||\tilde{A}_j|| < \infty$. That is,

$$\begin{aligned} X_t &= A(L)^{-1}C\Delta^{-\delta}\varepsilon_t + \Delta^{d-\delta}A(L)^{-1}Y_t^{+} \\ &= A(1)^{-1}C\Delta^{-\delta}\varepsilon_t + \tilde{A}(L)C\Delta^{1-\delta}\varepsilon_t + \Delta^{d-\delta}A(L)^{-1}Y_t^{+} \end{aligned}$$

and therefore

$$\beta^{*\prime}X_t = \beta^{\prime}A(1)\tilde{A}(L)C\Delta^{1-\delta}\varepsilon_t + \Delta^{d-\delta}\beta^{\prime}A(1)A(L)^{-1}Y_t^{+}$$

is $I(\min\{\delta-1,\delta-d\})$. If $\delta = 1$ the cointegrating residuals are then $I(1-d)$ as before.

This way of allowing for lags in the model imposes a VAR(k) structure on the $I(1)$ variables. Alternative specifications use VAR models in the fractional lag operator $L_d = 1-\Delta^d$, so that $L_1 = L$, see Johansen (2008, 2009). Both approaches may have advantages and disadvantages, but inference for the model proposed in Johansen (2008, 2009) complicates because both short and long run parameters depend on the parameter d, which is always identified, even in absence of cointegration.

Model (12) is nonlinear in Π and $A_1,\ldots,A_k$, so ML estimation can not be performed using the usual two step procedure of Johansen to prewhiten first the differenced levels $\Delta^{\delta}X_t$ and the fractional regressor $(1-\Delta^{-d})\Delta^{\delta}X_t$ given a particular value of d. Instead we could estimate the unrestricted linear model

$$\Delta^{\delta}X_t = \alpha\beta'(\Delta^{-d}-1)\Delta^{\delta}X_t + \sum_{j=1}^{k} L^j B_j \left\{(\Delta^{-d}-1)\Delta^{\delta}X_t\right\} + \sum_{j=1}^{k} L^j A_j \Delta^{\delta}X_t + \varepsilon_t,$$

under the assumption of Π being of rank r, but we do not impose $B_j = \Pi A_j$. The estimation procedure then follows as in Johansen's method but with an initial step to prewhiten the main series $\left\{(1-\Delta^{-d})\Delta^{\delta}X_t\right\}$ and $\Delta^{\delta}X_t$ on k lags of each. This estimate could be inefficient compared with the ML estimate, but much simpler to compute and analyze.

Given the pseudo ML estimate of β, we can construct the projection $\hat{\beta}_{\perp}'X_t$ and propose a similar second step, but in this case with the ECM enlarged by lags of ΔX_t,

$$\Delta\hat{\beta}_{\perp}'X_t = ab(1-\Delta^{-d})\Delta\hat{\beta}_{\perp}'X_t + \sum_{j=1}^{k} C_j \Delta X_{t-j} + e_t. \quad (13)$$

To justify such model, we note that in a triangular model set up, including a VAR modelization $A(L)X_t = X_t^{\dagger}$,

$$X_t = (I - A(L))X_t + \begin{pmatrix} I_r & \gamma \\ 0 & I_{k-r} \end{pmatrix} \begin{pmatrix} \Delta^{d-1}\varepsilon_{1t} \\ \Delta^{-1}\varepsilon_{2t} \end{pmatrix}$$

and therefore

$$\beta_{\perp}'X_t = \sum_{j=1}^{k} \beta_{\perp}'A_j X_{t-j} + \begin{pmatrix} M_1 & M_2 \end{pmatrix} \begin{pmatrix} \Delta^{d-1}\varepsilon_{1t} \\ \Delta^{-1}\varepsilon_{2t} \end{pmatrix}$$

where M_2 is full rank, with $\beta_{\perp}'X_t$ containing some $I(1-d)$ terms.

Example 3 ***General model.*** *From the representation in Theorem 8 of Johansen (2008),*

$$\beta_{\perp}'X_t = C^*\Delta^{-\delta}\varepsilon_t + \sum_{j=1}^{k} \beta_{\perp}'A_j X_{t-j} + \Delta^{d-\delta}\beta_{\perp}'Y_t^{+}$$

where $C^* = \beta_{\perp}'C$ *is full rank, so that* $\beta_{\perp}'X_t$ *is* $I(\delta)$ *and not cointegrated. However the term* $\Delta^{d-\delta}\beta_{\perp}'Y_t^{+}$ *is* $I(\delta - d)$.

In the augmented regression (13) we could impose the structure $\hat{\beta}_{\perp}'A_j$ in the coefficients of ΔX_{t-j}, but it can be preferable to let the coefficients unrestricted using the whole vector ΔX_{t-j}. In this way we take into account simultaneously the cointegrating directions, $\hat{\beta}'\Delta X_{t-j}$, that will serve to take into account the contribution of $\Delta^{d}\beta_{\perp}'Y_t^{+}$, and the orthogonal directions to these ones, $\hat{\beta}_{\perp}'\Delta X_{t-j}$.

Then, the asymptotic distribution of the maximum eigenvalue and test test statistics, $\mathcal{LR}_T(1)$ and $\mathcal{LR}_T(p-r_0)$, is approximated by E_{p-r_0} and J_{p-r_0} respectively, since the proof that the randomness of $\hat{\beta}_{\perp}$ does not affect test statistics in Appendix A can easily be extended to the augmented set up.

7 Rank Testing with different memory

In this section we propose a testing procedure to establish the cointegrating rank in case we have cointegrating relations with different memory. This procedure can be seen as a generalisation of the LR test for cointegration rank derived by Johansen and Nielsen, see Johansen and Nielsen (2010b), in the sense that our procedure is also LR test. However the asymptotic distribution of our test appears to be different and does not depend on the true cointegration degree, which is the case in Johansen and Nielsen (2010b). The asymptotic distribution of our procedure, that we derive in the following section, proves to have the same form as the asymptotic distribution of sup tests in Łasak (2010).

Tests we propose are LR tests based on the FVECM model that allows different memory of different cointegrating relations. Particularly we are interested in testing whether exists any extra cointegrating relation with a memory that is smaller than memory of the cointegrating relations under the null. Note that in practice we would rather detect stronger memory first. An extra cointegrating relation with the same memory could be possibly found by Johansen-Nielsen procedure. Recall Johansen and Nielsen (2010a) consider the following model

$$\Delta^{\delta} X_t = \Pi \Delta^{\delta - d} L_d X_t + \sum_{i=0}^{k} \Gamma_i \Delta^{\delta} L_d^i X_t + \varepsilon_t \tag{14}$$

and tests that $\Pi = \alpha\beta'$. They demonstrate that likelihood ratio test that $\Pi = \alpha\beta'$ has rank r has asymptotic distribution

$$-2 \log LR(\Pi = \alpha\beta') \xrightarrow{d} trace \left[\int_0^1 (dB)\, B_{d_0}{}' \left[\int_0^1 B_{d_0} B_{d_0}' \right]^{-1} \int_0^1 B_{d_0}\, (dB)' \right],$$

where B and B_{d_0} are both $p - r$ dimensional. (Note the difference in notation, B_{d_0} is denoted as B_{d^0-1} by Johansen and Nielsen (2010b)).

The model we consider is the following

$$\Delta X_t = \alpha_0 \beta_0' \left(\Delta^{-d^0} - 1 \right) \Delta X_t + \alpha_1 \beta_1' \left(\Delta^{-d^1} - 1 \right) \Delta X_t + \varepsilon_t, \tag{15}$$

where X_t is a vector of $I(1)$ series of order $p \times 1$, ε_t is a $p \times 1$ vector of Gaussian error with variance-covariance matrix Ω. The matrices α_0 and β_0 are $p \times r_0$ of rank r_0 and represent the error correction and cointegrating coefficients matrix, respectively, that share the same memory d^0. They could be estimated under the rank r_0 established by Johansen-Nielsen's procedure if we knew d^0. Matrices α_1 and β_1 are $p \times r_1$ of rank r_1 and represent the error correction and cointegrating coefficients matrix, respectively, that correspond to the relations with the memory d^1, that can be possibly smaller than d^0.

Based on the representation (15) we would like to test the null hypothesis of r_0 cointegrating relations β_0 with memory d^0 against the alternative hypothesis that there are extra r_1 cointegrating relations β_1 with a different memory d^1.
The procedure of testing and estimation is the following.

1. Estimate the model

$$\Delta X_t = \alpha_0 \beta_0' \left(\Delta^{-d^0} - 1 \right) \Delta X_t + \varepsilon_t, \tag{16}$$

under the null of r_0 cointegrating relations sharing the same memory d^0. We know from Lasak (2008) that this will give us consistent estimates of $\hat{\alpha}_0$, $\hat{\beta}_0$ and $\hat{d}^0$ if the rank r_0 is correctly specified. (According to Stakenas (2008) they are consistent even if the rank is not correctly specified)

2. Consider the model

$$\Delta X_t = \alpha_0 \beta_0' \left(\Delta^{-d^0} - 1\right) \Delta X_t + \alpha_1 \beta_1' \left(\Delta^{-d^1} - 1\right) \Delta X_t + \varepsilon_t, \tag{17}$$

and plug in estimated $\hat{\alpha}_0$, $\hat{\beta}_0$ and $\hat{d}^0$. (Or use the true ones α_0, β_0 and d^0 in case they are known). Define

$$Z_{0t}\left(\hat{d}^0\right) = \Delta X_t - \hat{\alpha}_0 \hat{\beta}_0' \left(\Delta^{-\hat{d}^0} - 1\right) \Delta X_t, \tag{18}$$

$$Z_{1t}(d^1) = \left(\Delta^{-d^1} - 1\right) \Delta X_t \tag{19}$$

and

$$S_{ij}(d^i, d^j) = T^{-1} \sum_{t=1}^{T} Z_{it}(d^i) Z_{jt}(d^j)' \quad i, j = 0, 1$$

3. Construct and solve the following eigenvalue problem

$$\left| \lambda(\hat{d}^0, d^1) S_{11}(d^1) - S_{10}(\hat{d}^0, d^1) S_{00}^{-1}(\hat{d}^0) S_{01}(\hat{d}^0, d^1) \right| = 0. \tag{20}$$

Note that $S_{11}(d^1)$ depends only on d^1, the memory of the extra cointegrating relations under the alternative, $S_{00}(\hat{d}^0)$ depends only on $\hat{d}^0$, the estimated memory of cointegrating relations under the null, while $S_{10}(\hat{d}^0, d^1)$, $S_{01}(\hat{d}^0, d^1)$ depends on both $\hat{d}^0$ and d^1.

4. Next solve the eigenvalue problem (20) and choose the solution that corresponds to the d^1 for which the likelihood function is maximized. Note that value of the likelihood function depends on the rank imposed under the alternative, so in general we have different solutions for two tests described below.

We consider two types LR tests, trace test and maximum eigenvalue test, that will be called sup-tests because of their asymptotic distribution derived in the next section.
Using trace test we test the null hypothesis

$$H_0 : rank\ (\Pi) = r_0$$

against the alternative hypothesis

$$H_1 : rank\ (\Pi) = p$$

using the test statistic defined by

$$\sup \text{trace} = trace(\hat{d}^0, d^1) = -2 \ln \left[LR\left(r_0 | p\right)\right] = -T \sum_{i=1}^{p} \ln[1 - \hat{\lambda}_i(\hat{d}^0, d^1)], \tag{21}$$

where

$$\begin{aligned}\hat{d}^0 &= \arg\max_{\hat{d}^0 \in \mathcal{D}} L_{r_0}(d^0) \\ \hat{d}^1_p &= \arg\max_{d^1 \in \mathcal{D}} L_p(\hat{d}^0, d^1) = \arg\max_{d^1 \in \mathcal{D}}\ trace(\hat{d}^0, d^1)\end{aligned}$$

and L_r denotes the likelihood under the hypothesis of rank r.[Discussion on the set $\mathcal{D}$, especially why it is the same for both $\hat{d}^0$ and d^1 to be included]

By maximum eigenvalue statistic we test cointegrating rank r_0 against rank $r_0 + 1$, i.e. we test the null hypothesis

$$H_0 : rank\ (\Pi) = r_0$$

against

$$H_1 : rank\ (\Pi) = r_0 + 1$$

and the test statistic is defined by

$$\sup \text{lambda} \max = \lambda_{\max}(\hat{d}^0, d^1) = -2 \ln \left[LR\left(r_0 | r_0 + 1\right)\right] = -T \ln[1 - \hat{\lambda}_1((\hat{d}^0, d^1)] \tag{22}$$

with

$$\begin{aligned}\hat{d}^0 &= \arg\max_{\hat{d}^0 \in \mathcal{D}} L_{r_0}(d^0) \\ \hat{d}^1_{r_0+1} &= \arg\max_{d^1 \in \mathcal{D}} L_{r_0+1}(\hat{d}^0, d^1) = \arg\max_{d^1 \in \mathcal{D}}\ \lambda_{\max}(\hat{d}^0, d^1)\end{aligned}$$

and L_r denotes the likelihood under the hypothesis of rank r.

Recall that while $\hat{d}^0$ estimates consistently the memory of the first r_0 cointegrating relations, we cannot hope that $\hat{d}^1_p$ or $\hat{d}^1_{r_0+1}$ estimate consistently a nonexisting, under the null, true value of d^1. Because of that our tests can be interpreted as sup LR tests, in the spirit of Davies (1977) and Hansen (1996), similarly to tests proposed in Lasak (2010) to test the null of no cointegration.

7.1 Asymptotic distribution

In this section we derive the asymptotic distribution of the likelihood ratio tests that we have proposed in (21) and (22). First let's state assumptions about the innovations, necessary to derive the asymptotic distributions of our likelihood ratio tests.

Assumption 1 *ε_t are independent and identically distributed vectors with mean zero, positive definite covariance matrix Ω, and $E||\varepsilon_t||^q < \infty,\ q \geq 4,\ q > 2/\left(2d - 1\right),\ d \in \mathcal{D}$*

Note that under H_0 we have:

$$\Delta X_t - \hat{\alpha}_0 \hat{\beta}_0' \left(\Delta^{-\hat{d}^0} - 1\right) \Delta X_t = \varepsilon_t, \tag{23}$$

or

$$Z_{0t}\left(\hat{d}^0\right) = \varepsilon_t, \ \ t = 1, ..., T$$

so it can be easily seen that

$$S_{00}\left(\hat{d^0}\right) \xrightarrow{P} \Omega.$$

Further using the methods of Marinucci and Robinson (2000) we obtain that under Assumption 1

$$T^{0.5-d^1} Z_{1[T\tau]} \xrightarrow{\omega} W_{d^1}(\tau), \qquad \text{for } d^1 > 0.5,$$

where $\xrightarrow{\omega}$ means convergence in the Skorohod J_1 topology of $\mathcal{D}[0,1]$, W_{d^1} is a fractional Brownian motion called by Marinucci and Robinson (1999) "Type II" fractional Brownian motion and defined as

$$W_d(\tau) = \int_0^\tau \frac{(\tau - s)^{d-1}}{\Gamma(d)} dW(s),$$

and $W(s)$ is vector Brownian motion with covariance matrix Ω.
Then by the Continuous Mapping Theorem we have the following convergence for each $d^1 > 0.5$

$$T^{1-2d^1} S_{11}(d^1) \xrightarrow{d} \int_0^1 W_{d^1}(\tau) W_{d^1}(\tau)' d\tau \tag{24}$$

and, similarly to Robinson and Hualde (2003), Proposition 3,

$$T^{1-d^1} S_{10}(\hat{d^0}, d^1) \xrightarrow{d} \int_0^1 W_{d^1}(\tau) dW',$$

where $\xrightarrow{d}$ denotes convergence in distribution.

The product moments $T^{1-2d^1} S_{11}(d^1)$, $T^{1-d^1} S_{10}(\hat{d^0}, d^1)$ are $O_p(1)$ uniformly in d^1 since we can show weak convergence for $d^1 \in \mathcal{D}$ in the space $C(\mathcal{D})$ of continuous functions in $\mathcal{D}$ (see Proof of Theorem 1 in the Appendix A in Łasak (2010)), S_{00} is also $O_p(1)$, so the roots $\hat{\lambda}_i(\hat{d^0}, d^1)$ of equation (20) converge to zero like T^{-1} under the null of no cointegration. This implies that

$$-T \sum_{i=1}^{p} \ln[1 - \hat{\lambda}_i(\hat{d^0}, d^1)] = T \sum_{i=1}^{p} \hat{\lambda}_i(\hat{d^0}, d^1) + o_p(1).$$

The sum of the eigenvalues can be found as follows

$$\left|\lambda(\hat{d^0}, d^1) S_{11}(d^1) - S_{10}(\hat{d^0}, d^1) S_{00}^{-1}(\hat{d^0}) S_{01}(\hat{d^0}, d^1)\right| = 0. \tag{25}$$

that is equivalent to solve the equation

$$\left|\lambda(\hat{d^0}, d^1) I - S_{11}^{-1}(d^1) S_{10}(\hat{d^0}, d^1) S_{00}^{-1}(\hat{d^0}) S_{01}(\hat{d^0}, d^1)\right| = 0. \tag{26}$$

which shows that

$$T \sum_{i=1}^{p} \hat{\lambda}_i(\hat{d^0}, d^1) = T\, tr\{S_{11}^{-1}(d^1) S_{10}(\hat{d^0}, d^1) S_{00}^{-1}(\hat{d^0}) S_{01}(\hat{d^0}, d^1)\}.$$

From the above reasoning we find that for each d^1 the product

$$S_{11}^{-1}(d^1) S_{10}(\hat{d^0}, d^1) S_{00}^{-1}(\hat{d^0}) S_{01}(\hat{d^0}, d^1)$$

converges in distribution towards

$$\left(\int_0^1 W_{d^1}(\tau) W_{d^1}(\tau)' d\tau\right)^{-1} \int_0^1 W_{d^1}(\tau) dW'\Omega^{-1} \int_0^1 (dW) W_{d^1}(\tau)',$$

which we can write as

$$\Omega^{-1/2} \left[\int_0^1 B_{d^1}(\tau) B_{d^1}(\tau)' d\tau\right]^{-1} \int_0^1 B_{d^1}(\tau) dB' \int_0^1 (dB) B_{d^1}(\tau)' \left(\Omega^{1/2}\right)', \tag{27}$$

where $B_{d^1}(\tau) = \Omega^{-1/2} W_{d^1}(\tau)$ is the standard fractional Brownian motion. Then we can see that asymptotic distribution of trace and maximum eigenvalue for a fixed d^1 are respectively the trace and the greatest eigenvalue of (27), i.e.

$$trace(\hat{d}^0, d^1) \xrightarrow{d} trace \left[\int_0^1 (dB) B_{d^1}(\tau)' \left[\int_0^1 B_{d^1}(\tau) B_{d^1}(\tau)' d\tau\right]^{-1} \int_0^1 B_{d^1}(\tau) (dB)'\right]$$

$$\lambda \max(\hat{d}^0, d^1) \xrightarrow{d} \lambda_1 \left[\int_0^1 (dB) B_{d^1}(\tau)' \left[\int_0^1 B_{d^1}(\tau) B_{d^1}(\tau)' d\tau\right]^{-1} \int_0^1 B_{d^1}(\tau) (dB)'\right].$$

In the case when d^1 is estimated the following theorem applies.

Theorem 1 *When d^1, $d^1 \in \mathcal{D}$, is estimated by the maximum likelihood principle under the model (15) using procedure proposed in the previous section the asymptotic distributions of trace and maximum eigenvalue statistics are given respectively by*

$$\sup trace = trace(\hat{d}^0, \hat{d}^1_p) \xrightarrow{d} \sup_{\hat{d}^1_p \in \mathcal{D}} trace \left[\pounds(d^1)\right],$$

and

$$\sup lambda \max = \lambda \max(\hat{d}^0, \hat{d}^1_{r_0+1}) \xrightarrow{d} \sup_{\hat{d}^1_{r_0+1} \in \mathcal{D}} \lambda_1 \left[\pounds(d^1)\right],$$

where $\mathcal{D} = [0.5 + \varepsilon, 1]$ is a compact set, and

$$\pounds(d^1) = \int_0^1 (dB) B_{d^1}(\tau)' \left[\int_0^1 B_{d^1}(\tau) B_{d^1}(\tau)' d\tau\right]^{-1} \int_0^1 B_{d^1}(\tau) (dB)',$$

where B is a $p - r_0$-dimensional Brownian motion on the unit interval, $B_{d^1}(\tau)$ is the standard fractional Brownian motion.

The proof follows as in the Appendix A of Łasak (2010). And the same comments concerning set $\mathcal{D}$ apply.

Finally let us consider the behavior of our tests under the alternative. Note that if the null hypothesis is not true and we have extra cointegrating relation with memory $d^1 \leq d^0$, then one of the eigenvalues in (20) will be positive in the limit. Then

$$-2 \ln [LR(r_0|p)] \geqslant -T \ \ln\left(1 - \hat{\lambda}_1(\hat{d}^0, \hat{d}^1_p)\right) \xrightarrow{p} \infty$$

and

$$-2\ln\left[LR\left(r_0|r_0+1\right)\right] = -T\ln\left(1-\hat{\lambda}_1(\hat{d}^0,\hat{d}^1_{r_0+1})\right) \xrightarrow{p} \infty.$$

So the asymptotic power of both tests is 1.

8 Finite sample properties

(incomplete yet, procedure to test the rank with different memory to be included)

In this section we compare by Monte Carlo simulation the performance in finite samples of all the tests discussed in this paper,

- new procedures based on projections on $\hat{\beta}'_{\perp}X_t$: $\mathcal{LR}_T\left(p-r_0\right)$ for trace and $\mathcal{LR}_T\left(1\right)$ for maximum eigenvalue test;
- trace and maximum eigenvalue tests based on the estimation of the standard VECM like in Johansen (1988, 1991, 1995), called Johansen's trace and Johansen's maximum eigenvalue tests to easy the notation;
- trace test proposed by Lyhagen (1998), to which we add pre-estimation of d under the null, called to simplify as Lyhagen's trace (it can be viewed as a version of Johansen and Nielsen's (2010b) LR test as well);
- naive sup tests, where we estimate d under the alternative and we use sup tables with $p-r_0$ degrees of freedom in line of the standard "Johansen's procedure".

Let us describe the data generating process. We simulate a trivariate system ($p=3$) using the following triangular representation

$$X_t = \begin{pmatrix} I_r & \gamma \\ 0 & I_{p-r} \end{pmatrix} \begin{pmatrix} \Delta^{d_0-1}\varepsilon_{1t} \\ \Delta^{-1}\varepsilon_{2t} \end{pmatrix},\ t=1,...,T \tag{28}$$

for the basic model (1). Note that the triangular representation (28) implies ECM (1) with

$$\alpha = \begin{pmatrix} -I_r \\ 0 \end{pmatrix} \text{ and } \beta' = \left(I_r \ \ -\gamma\right).$$

We consider the model (12) with $k=1$. For this model we add to (28) the autoregression

$$Z_t = A_1 Z_{t-1} + X_t,$$

with $Z_0 = 0$ and $A_1 = a\ I_p$, where 3 different values for a are considered $a=-0.6$, $a=0$ or $a=0.6$. We simulate systems (28) cointegrated of order $d_0 = 0.55$, 0.75 and 0.95.

Note that the case $A_1 = 0$ is useful to check on one hand the effect of overspecification of k in terms of size and power, and also can be used to check whether it is better to incorporate the whole vector ΔX_t or cointegrating residuals $\Delta\hat{\beta}'X_t$ in the regression to control the $I\left(-d_0\right)$ terms.

To check the size of the considered tests we simulate the system (28) with 1 cointegrating relation $\beta = [1\ 1\ 1]'$, whereas to check the power we simulate the same system (28) with 2 cointegrating relations

$$\beta = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}'.$$

The innovations $\varepsilon_t = (\varepsilon'_{1t}, \varepsilon'_{2t})'$ are Gaussian IID $(0, \Omega)$ where

$$\Omega = \begin{pmatrix} \omega^2 & \omega\rho & \omega\rho \\ \omega\rho & 1 & 0 \\ \omega\rho & 0 & 1 \end{pmatrix}$$

with $\omega^2 = 0.5$ and $\rho = 0$ or $\rho = 0.4$.

We make all the simulations in Ox 3.40 or Ox 4.04 (see Doornik and Ooms (2001) and Doornik (2002)) with 10,000 replications. We consider the sample sizes of $T = 100$, 200, 300 observations.

The results of Monte Carlos simulation are presented below. Tables 4.1-4.6 demonstrate the percentage of rejections under the null hypothesis of cointegration rank $r = 1$. The percentage of rejections under the alternative hypothesis of cointegration rank $r = 2$ is presented in Tables 4.7-4.12.

Table 4.1. Size of tests: $d_0 = 0.55$, $r_0 = 1$, $r = 1$, $\rho = 0$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T(p - r_0)$	5.4	5.8	5.7	5.1	5.5	5.3	5.3	5.2	4.9
$LR_T(1)$	5.2	5.7	5.4	4.9	5.2	5.1	5.1	5.0	4.7
Johansen's trace	4.6	5.1	4.9	4.0	5.0	4.9	2.2	3.3	4.1
Johansen's lambdamax	4.5	5.0	4.9	3.9	5.0	4.8	2.0	3.3	4.0
Naive sup trace	3.8	3.3	2.7	3.4	3.7	3.2	1.8	2.9	3.5
Naive sup lambdamax	3.6	3.2	2.6	3.3	3.7	3.1	1.6	2.7	3.2
Lyhagen's & JN's trace	9.5	5.9	6.8	4.2	6.7	6.9	2.3	3.6	5.4

Table 4.2. Size of tests: $d_0 = 0.55$, $r_0 = 1$, $r = 1$, $\rho = 0.4$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T(p - r_0)$	6.8	8.1	9.0	6.1	6.7	7.1	6.2	6.0	6.0
$LR_T(1)$	6.8	8.1	9.0	6.0	6.7	7.0	6.0	5.9	5.9
Johansen's trace	5.1	5.2	5.0	4.7	5.2	5.0	2.7	4.2	4.7
Johansen's lambdamax	5.0	5.2	5.0	4.7	5.2	5.0	2.5	4.1	4.6
Naive sup trace	3.8	3.1	2.5	3.9	3.7	3.1	2.2	3.5	3.9
Naive sup lambdamax	3.8	3.1	2.5	3.9	3.6	3.0	2.0	3.4	3.8
Lyhagen's & JN's trace	9.6	6.6	6.5	4.7	6.4	6.7	2.8	4.3	4.6

Table 4.3. Size of tests: $d_0 = 0.75$, $r_0 = 1$, $r = 1$, $\rho = 0$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	6.9	6.2	5.8	6.3	5.1	4.8	5.9	5.3	4.4
$LR_T\,(1)$	7.1	5.8	5.7	6.0	4.2	5.0	5.1	5.3	4.0
Johansen's trace	6.0	5.9	5.6	5.6	5.1	5.2	4.4	5.6	5.3
Johansen's lambdamax	6.0	5.3	5.3	5.3	4.3	4.8	3.5	5.3	5.3
Naive sup trace	5.0	4.5	4.0	5.2	4.1	3.3	3.6	4.6	4.3
Naive sup lambdamax	5.3	3.7	3.4	4.8	3.4	3.8	2.9	4.2	4.2
Lyhagen's & JN's trace	11.3	5.2	4.6	5.7	4.6	4.0	4.5	5.6	5.2

Table 4.4. Size of tests: $d_0 = 0.75$, $r_0 = 1$, $r = 1$, $\rho = 0.4$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	8.2	9.0	10.8	6.3	8.0	8.2	6.0	6.8	5.2
$LR_T\,(1)$	8.2	8.2	10.9	5.9	7.2	7.9	5.1	6.7	4.7
Johansen's trace	4.7	6.2	5.4	4.2	6.5	5.7	3.6	6.0	4.5
Johansen's lambdamax	4.4	5.1	5.2	4.3	5.7	5.6	3.1	5.9	5.1
Naive sup trace	3.3	4.7	3.8	3.3	5.4	3.9	3.5	5.9	4.2
Naive sup lambdamax	3.7	4.1	3.4	3.2	4.9	3.8	1.3	3.7	3.0
Lyhagen's & JN's trace	6.5	5.5	4.2	4.0	6.5	4.8	3.5	5.9	4.2

Table 4.5. Size of tests: $d_0 = 0.95$, $r_0 = 1$, $r = 1$, $\rho = 0$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	7.0	6.2	6.5	6.4	5.0	4.9	5.9	5.0	4.0
$LR_T\,(1)$	7.1	6.0	5.7	6.2	4.0	5.1	5.2	5.0	3.8
Johansen's trace	6.0	6.1	5.6	6.1	5.4	5.2	6.0	6.1	5.5
Johansen's lambdamax	6.2	5.3	5.1	5.9	4.7	4.8	5.7	5.9	5.2
Naive sup trace	5.1	5.0	4.6	5.5	4.8	3.9	5.5	5.2	4.6
Naive sup lambdamax	5.0	4.3	4.3	5.2	3.9	4.3	5.0	4.7	4.4
Lyhagen's & JN's trace	8.7	6.1	5.5	7.4	5.4	5.1	6.4	5.9	5.5

Table 4.6. Size of tests: $d_0 = 0.95$, $r_0 = 1$, $r = 1$, $\rho = 0.4$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	9.1	9.8	12.0	6.2	7.9	7.9	5.5	6.7	5.2
$LR_T\,(1)$	9.0	9.1	11.7	5.9	6.9	7.3	5.5	6.2	4.6
Johansen's trace	4.6	6.0	5.4	4.7	6.4	5.7	4.8	6.5	5.1
Johansen's lambdamax	4.4	5.1	5.4	4.1	5.4	5.6	4.4	6.2	5.1
Naive sup trace	3.8	4.9	4.8	3.8	6.0	4.5	3.7	5.8	4.2
Naive sup lambdamax	3.8	4.0	4.0	3.4	4.8	4.5	3.0	4.7	4.0
Lyhagen's & JN's trace	5.7	5.8	5.3	4.8	6.6	5.6	4.4	6.3	4.9

Two step tests have reasonable size for simpler model ($\rho = 0$), but when $\rho \neq 0$ it is oversized, more as d_0 increases, and for some designs it is not improving as T grows. Johansen's tests and naive sup tests are usually undersized, while Lyhagen's test can be seriously oversized when $a \leq 0$. Overall, naive sup tests seem to be the best ones in the terms of size, being the most conservative ones. The case with $a > 0$ appears to be more difficult.

Table 4.7. Power of tests: $d_0 = 0.55$, $r_0 = 1$, $r = 2$, $\rho = 0$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	99.1	100	100	80.2	99.9	100	10.0	45.0	84.1
$LR_T\,(1)$	99.4	100	100	81.0	99.9	100	7.5	44.6	85.5
Johansen's trace	88.8	99.3	100	59.2	95.1	99.2	6.1	34.2	69.3
Johansen's lambdamax	88.6	99.1	100	60.3	95.2	99.4	5.4	34.6	68.8
Naive sup trace	97.0	100	100	67.3	99.8	100	5.7	34.7	74.2
Naive sup lambdamax	97.8	100	100	68.1	99.8	100	4.4	35.4	73.7
Lyhagen's & JN's trace	99.4	100	100	82.3	99.9	100	6.4	45.0	76.9

Table 4.8. Power of tests: $d_0 = 0.55$, $r_0 = 1$, $r = 2$, $\rho = 0.4$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	99.7	100	100	87.2	100	100	18.2	59.2	91.0
$LR_T\,(1)$	99.7	100	100	87.4	100	100	15.5	59.9	91.4
Johansen's trace	92.5	99.6	100	66.0	97.4	99.6	8.6	38.8	72.1
Johansen's lambdamax	91.6	99.5	100	66.2	97.5	99.6	6.1	39.7	72.2
Naive sup trace	99.8	100	100	73.6	99.9	100	8.7	40.1	76.2
Naive sup lambdamax	98.8	100	100	73.3	100	100	5.8	39.8	78.9
Lyhagen's & JN's trace	99.8	100	100	86.7	100	100	9.6	39.6	73.5

Table 4.9. Power of tests: $d_0 = 0.75$, $r_0 = 1$, $r = 2$, $\rho = 0$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	100	100	100	99.9	100	100	42.1	98.6	100
$LR_T\,(1)$	100	100	100	100	100	100	41.5	98.9	100
Johansen's trace	100	100	100	98.9	100	100	35.6	95.8	100
Johansen's lambdamax	100	100	100	99.1	100	100	36.1	95.3	100
Naive sup trace	100	100	100	99.2	100	100	32.6	96.2	100
Naive sup lambdamax	100	100	100	99.8	100	100	32.3	96.2	100
Lyhagen's & JN's trace	100	100	100	99.9	100	100	37.0	96.8	100

Table 4.10. Power of tests: $d_0 = 0.75$, $r_0 = 1$, $r = 2$, $\rho = 0.4$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	100	100	100	100	100	100	56.3	98.8	100
$LR_T\,(1)$	100	100	100	100	100	100	55.3	99.0	100
Johansen's trace	100	100	100	99.5	100	100	40.1	96.6	100
Johansen's lambdamax	100	100	100	99.8	100	100	40.9	97.0	100
Naive sup trace	100	100	100	99.8	100	100	37.2	97.0	100
Naive sup lambdamax	100	100	100	99.8	100	100	35.7	97.9	100
Lyhagen's & JN's trace	100	100	100	100	100	100	41.7	97.9	100

Table 4.11. Power of tests: $d_0 = 0.95$, $r_0 = 1$, $r = 2$, $\rho = 0$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T\,(p - r_0)$	100	100	100	100	100	100	90.8	100	100
$LR_T\,(1)$	100	100	100	100	100	100	91.9	100	100
Johansen's trace	100	100	100	100	100	100	87.6	100	100
Johansen's lambdamax	100	100	100	100	100	100	89.6	100	100
Naive sup trace	100	100	100	100	100	100	85.1	100	100
Naive sup lambdamax	100	100	100	100	100	100	86.2	100	100
Lyhagen's & JN's trace	100	100	100	100	100	100	88.2	100	100

Table 4.12. Power of tests: $d_0 = 0.95$, $r_0 = 1$, $r = 2$, $\rho = 0.4$

T	100	200	300	100	200	300	100	200	300
a		-0.6			0.			0.6	
$LR_T(p - r_0)$	100	100	100	100	100	100	93.8	100	100
$LR_T(1)$	100	100	100	100	100	100	94.7	100	100
Johansen's trace	100	100	100	100	100	100	89.5	100	100
Johansen's lambdamax	100	100	100	100	100	100	92.5	100	100
Naive sup trace	100	100	100	100	100	100	86.3	100	100
Naive sup lambdamax	100	100	100	100	100	100	89.3	100	100
Lyhagen's & JN's trace	100	100	100	100	100	100	89.4	100	100

Power in general increases with d_0 as expected, as well as in ρ and decreases with a. The two step procedure appears as the most powerful in terms of raw power, but it is not clear it would be the best in terms of size-adjusted power. Among other procedures, naive sup tests perform better for small d_0 ($d_0 = 0.55$), but Johansen's tests are better when d_0 gets closer to $d_0 = 1$ as is then the appropriate LR test. Lyhagen's version of the LR shows a behaviour close to the best one of these single step tests.

9 Conclusions

We propose a new two-step procedure to establish cointegration rank in a fractional system. We investigate the performance of the proposed procedure in finite samples for a simple fractionally cointegrated model and compare it with appropriate versions of sup LR tests, Lyhagen's tests and Johansen's tests. All of them present important size problems to different extent, but naive sup LR tests seem to be the most reliable and fares well in comparison with the most powerful alternatives.

We also proposed a new testing procedure that allows to find extra cointegrating relations with different memory. This procedure can be seen as an extension of both Johansen-Nielsen's (2010b) procedure and Łasak's (2010) test for no fractional cointegration.

Methodology developed in this paper allows to complete the basic likelihood analysis of fractionally cointegrated systems with higher rank and can be adapted and further developed to include deterministic terms and to allow for unknown memory of the original series, among other extensions.

10 Appendix A

Proof. We demonstrate here that replacing $\beta_\perp$ by $\hat{\beta}_\perp$ makes no difference asymptotically in two step LR test statistics.

Setting $\hat{V}_t = \hat{\beta}_\perp' X_t$ and $\hat{V}_t(d) = \left(1 - \Delta^{-d}\right)\Delta \hat{V}_t$ and defining V_t and $V_t(d)$ with the true $\beta_\perp$, we want to show that

$$T^{-d}\sum_{t=1}^{T} \hat{V}_t(d)\,\Delta\hat{V}_t' - T^{-d}\sum_{t=1}^{T} V_t(d)\,\Delta V_t' \to_p 0$$

uniformly for $d \in D$ if $\hat{\beta}_\perp - \beta_\perp = O_p\left(T^{-d_0}\right)$. We first have that

$$T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta\hat{V}_t' = T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta V_t' + T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\left(\Delta\hat{V}_t - \Delta V_t\right)'.$$

The first term is

$$T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta V_t' = T^{-d}\sum_{t=1}^{T}V_t(d)\,\Delta V_t' + T^{-d}\sum_{t=1}^{T}\left\{\hat{V}_t(d) - V_t(d)\right\}\Delta V_t'$$

where the first term on the right hand side is $O_p(1)$ uniformly in d and

$$T^{-d}\sum_{t=1}^{T}\left\{\hat{V}_t(d) - V_t(d)\right\}\Delta V_t' = \left(\hat{\beta}_\perp' - \beta_\perp'\right)T^{-d}\sum_{t=1}^{T}X_t(d)\,\Delta V_t',$$

which is $o_p(1)$ uniformly in d because $T^{-d}\sum_{t=1}^{T}X_t(d)\,\Delta V_t'$ is $O_p(1)$ uniformly in d.

The second term on the right hand side of (13) is

$$T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta X_t'\left(\hat{\beta}_\perp - \beta_\perp\right) = O_p\left(T^{-d_0}\right)T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta X_t,$$

and this can be seen easily to be $O_p\left(T^{d_0-d}\right) = o_p(1)$, uniformly in d, because

$$T^{-1}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta X_t' \to_p \lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^{T}V_t(d)\,\Delta X_t'\right] = \sum_{j=0}^{\infty}\pi_j(1-d)\,\Omega\psi_j' < \infty$$

where ψ_j are the Wold decomposition weights of ΔX_t, which is $I(0)$. Then

$$T^{-d}\sum_{t=1}^{T}\hat{V}_t(d)\,\Delta X_t'\left(\hat{\beta}_\perp - \beta_\perp\right)' = O_p\left(T^{1-d_0-d}\right) = o_p(1),$$

uniformly in d, because $d_0, d > 0.5$, and the estimation effect of $\beta_\perp$ is negligible.

11 Appendix B

In this appendix we justify the two step procedure described in Section 7.

1. ■ **Size.** Proof under the null with general FVECM (no necessary Triangular m. here):

$$\Delta X_t = \alpha_0\beta_0'\left(\Delta^{-b_0} - 1\right)\Delta X_t + \varepsilon_t$$

Proofs. We need to show that the estimation effect is negligible, i.e.

$$\begin{aligned}\sup_{b_1\in\mathcal{B}} T^{1-2b_1}\left\{S_{11}(b_1) - \hat{S}_{11}(b_1)\right\} &= o_p(1)\\ \sup_{b_1\in\mathcal{B}} T^{1-b_1}\left\{S_{10}(b_1) - \hat{S}_{10}(b_1)\right\} &= o_p(1)\\ S_{00}(b_1) - \hat{S}_{00}(b_1) &= o_p(1)\end{aligned}$$

where, e.g.

$$\hat{S}_{11}(b_1) = \frac{1}{T}\sum_{t=1}^{T}\hat{Z}_{1t}(b_1)\hat{Z}_{1t}(b_1)'$$
$$\hat{S}_{10}(b_1) = \frac{1}{T}\sum_{t=1}^{T}\hat{Z}_{1t}(b_1)\hat{R}_t'$$

with

$$\begin{aligned}\hat{Z}_{1t}(b_1) &= \left(\Delta^{-b_1}-1\right)\hat{R}_t \\ \hat{R}_t &= \Delta X_t - \hat{\alpha}_0\hat{\beta}_0'\left(\Delta^{-\hat{b}_0}-1\right)\Delta X_t \\ &= \varepsilon_t + \left\{\alpha_0\beta_0'\left(\Delta^{-b_0}-1\right) - \hat{\alpha}_0\hat{\beta}_0'\left(\Delta^{-\hat{b}_0}-1\right)\right\}\Delta X_t.\end{aligned}$$

Here the critical point is the proof of the difference $\left\{S_{10}(b_1) - \hat{S}_{10}(b_1)\right\}$ being negligible, in particular the last term in this expansion,

$$\begin{aligned}\sup_{b_1\in\mathcal{B}} T^{1-b_1}\left\{S_{10}(b_1)-\hat{S}_{10}(b_1)\right\} &= \sup_{b_1\in\mathcal{B}} T^{-b_1}\sum_{t=1}^{T}\left\{Z_{1t}(b_1)\varepsilon_t' - \hat{Z}_{1t}(b_1)\hat{R}_t'\right\} \\ &= \sup_{b_1\in\mathcal{B}} T^{-b_1}\sum_{t=1}^{T}\left\{Z_{1t}(b_1)-\hat{Z}_{1t}(b_1)\right\}\varepsilon_t \\ &\quad + \sup_{b_1\in\mathcal{B}} T^{-b_1}\sum_{t=1}^{T}\hat{Z}_{1t}(b_1)\left\{\varepsilon_t'-\hat{R}_t'\right\},\end{aligned}$$

where $Z_{1t}(b_1) = \left(\Delta^{-b_1}-1\right)\varepsilon_t$ and $R_t = \varepsilon_t$. Now

$$\begin{aligned}\sum_{t=1}^{T}\hat{Z}_{1t}(b_1)\left\{\varepsilon_t'-\hat{R}_t'\right\} &\sim \sum_{t=1}^{T}Z_{1t}(b_1)\left\{\varepsilon_t'-\hat{R}_t'\right\} \\ &= \sum_{t=1}^{T}Z_{1t}(b_1)\Delta X_t'\left\{\alpha_0\beta_0'\left(\Delta^{-b_0}-1\right) - \hat{\alpha}_0\hat{\beta}_0'\left(\Delta^{-\hat{b}_0}-1\right)\right\}' \\ &\sim \sum_{t=1}^{T}Z_{1t}(b_1)\left\{\left(\Delta^{-b_0}-1\right)\Delta X_t'\beta_0\right\}(\alpha_0-\hat{\alpha}_0)' + (\ldots),\end{aligned}$$

where $Z_{1t}(b_1) \sim I(b_1)$ and $\left\{\left(\Delta^{-b_0}-1\right)\Delta X_t'\beta_0\right\} \sim I(0)$. Then for

$$W_T(b_1) = \sum_{t=1}^{T}Z_{1t}(b_1)\left\{\left(\Delta^{-b_0}-1\right)\Delta X_t'\beta_0\right\}$$

we have that $E[W_T(b_1)] = T$, while $Var[W_T(b_1)] = T^2$ [tightness comes from Lasak (2010) or Johansen-Nielsen] so $W_T = O_p(1)$ for every b_1, and the contribution of this term is $O_p\left(T^{-b_1}T^{-1/2}T\right) = O_p\left(T^{1/2-b_1}\right) = o_p(1)$ given that $(\alpha_0-\hat{\alpha}_0) = T^{-1/2}$ and $b_1 > \frac{1}{2}$.

■ **Analysis of power properties** under the alternative: perhaps only do a preliminary analysis justifying the alternative with a triangular model,

$$\Delta X_t = \alpha_0 \beta_0' \left(\Delta^{-b_0} - 1\right) \Delta X_t + \alpha_1 \beta_1' \left(\Delta^{-b_1} - 1\right) \Delta X_t + \varepsilon_t.$$

Here the key point is to know the behavior of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{b}_0$. It is more obvious that (irrespective of $b_0 =\neq b_1$) $\hat{\beta}_0 \in sp(\beta)$, $\beta = (\beta_0, \beta_1)$, and $\hat{\alpha}$ will converge to the appropiate scaling given the limit of the (normalized) $\hat{\beta}_0$. The limit value of $\hat{b}_0$ seems to be in the range $[b_{\min}, b^{\max}]$ depending on the model parameters, but there is a tendency to estimate $b_{\max}$ which is the value which would provide a better fit, everything else the same (e.g. fixed components of $\Sigma_{\beta\beta}$). Then for some $\beta_0^{*\prime} = \delta_0 \beta_0 + \delta_1 \beta_1$

$$\begin{aligned} \hat{R}_t &= \Delta X_t - \hat{\alpha}_0 \hat{\beta}_0' \left(\Delta^{-\hat{b}_0} - 1\right) \Delta X_t \\ &= \alpha_0 \beta_0' \left(\Delta^{-b_0} - 1\right) \Delta X_t + \alpha_1 \beta_1' \left(\Delta^{-b_1} - 1\right) \Delta X_t \\ &\quad - \alpha_0^* \beta_0^{*\prime} \left(\Delta^{-b_0^*} - 1\right) \Delta X_t + \varepsilon_t + o_p(1) \end{aligned}$$

so it is pretty clear that $\hat{R}_t$ is going to be cointegrated, possible still in both directions β_0 and β_1, and any consistent cointegration test will detect that.

References

[1] Avarucci, M. (2007), Three Essays on Fractional Cointegration, Ph.D. Thesis, University of Rome "Tor Vergata".

[2] Avarucci, M., Velasco C. (2009), A Wald Test for the Cointegration Rank in Nonstationary Fractional Systems, Journal of Econometrics, 151 (2), 178-189.

[3] Breitung, J., Hassler U. (2002), Inference on the Cointegration Rank in Fractionally Integrated Processes, Journal of Econometrics, Elsevier, 110 (2), 167-185.

[4] Chen, W., Hurvich, C. (2003a), Estimating fractional cointegration in the presence of polynomial trends, Journal of Econometrics, Elsevier, 117 (1), 95-121.

[5] Chen, W., Hurvich, C. (2003b), Semiparametric Estimation of Multivariate Fractional Cointegration, Journal of American Statistical Association, 98, 629-642.

[6] Chen, W., Hurvich, C. (2006), Semiparametric Estimation of Fractional Cointegrating Subspaces, Annals of Statistics, 34, 2939-2979.

[7] Davies, R.B. (1977), Hypothesis Testing when a Nuisance is Present Only under the Alternative, Biometrika, 64, 247-254.

[8] Doornik, J. A. (2002), Object-Oriented Matrix Programming Using Ox, 3rd ed. London, Timberlake Consultants Press and Oxford, www.doornik.com.

[9] Doornik, J. A., Ooms, M. (2001), Introduction to Ox, www.nuff.ox.ac.uk/Users/Doornik.

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

[11] Hansen, B.E. (1996), Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, Econometrica, 64, 413-430.

[12] Hualde, J. (2009), Consistent estimation of cointegrating subspaces, work in progress.

[13] Hualde, J., Robinson, P.M. (2007), Root-N-Consistent Estimation Of Weak Fractional Cointegration, Journal of Econometrics, Elsevier, 140, 450-484.

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

[15] Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica, 59 (6), 1551-1580.

[16] Johansen, S. (1995), Likelihood-based inference in cointegrated Vector Auto-Regressive Models. Oxford University Press, Oxford.

[17] Johansen, S. (2008), A representation theory for a class of vector autoregressive models for fractional processes, Econometric Theory, 24, 651-676.

[18] Johansen, S. (2009), Representation of cointegrated autoregressive processes with application to fractional cointegration, Econometric Reviews, 28, 121-145.

[19] Johansen, S., Nielsen, M. Ø. (2010a), Likelihood Inference for a Nonstationary Fractional Autoregressive Model, Journal of Econometrics, in press.

[20] Johansen, S., Nielsen, M. Ø. (2010b), Likelihood inference for a vector autoregressive model which allows for fractional and cofractional processes. Work in progress.

[21] Łasak, K. (2008), Maximum likelihood estimation of fractionally cointegrated systems, CREATES Research Paper 2008-53.

[22] Łasak, K. (2010), Likelihood based testing for no fractional cointegration, Journal of Econometrics, in press.

[23] Lobato, I., Velasco, C. (2007), Efficient Wald Tests for Fractional Unit Roots, Econometrica, 75, 575-589.

[24] Lyhagen J. (1998), Maximum likelihood estimation of the multivariate fractional cointegrating model, Working Paper Series in Economics and Finance, 233, Stockholm School of Economics.

[25] Nielsen, M. Ø., Shimotsu, K. (2007), Determining the cointegrating rank in nonstationary fractional systems by exact local Whittle approach, Journal of Econometrics 141, 574–596.

[26] Robinson, P.M., Hualde, J. (2003), Cointegration in Fractional Systems with Unknown Integration Orders, Econometrica, Econometric Society, 71 (6), 1727-1766.

[27] Robinson, P.M., Yajima, Y. (2002), Determination of cointegrating rank in fractional systems, Journal of Econometrics, 106, 217-241.

[28] Stakėnas, P. (2008), Likelihood-based estimation of cointegration rank in fractional systems, University of Amsterdam, work in progress.