

Fractional cointegration rank estimation

Katarzyna Lasak and Carlos Velasco

CREATES Research Paper 2013-08

Fractional cointegration rank estimation

Katarzyna Lasak*

Department of Econometrics, FEWEB,
VU University Amsterdam
& CREATES, Aarhus University
E-mail: k.a.lasak@vu.nl

Carlos Velasco

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

March 15, 2013

Abstract

We consider cointegration rank estimation for a p -dimensional Fractional Vector Error Correction Model. We propose a new two-step procedure which allows testing for further long-run equilibrium relations with possibly different persistence levels. The first step consists in estimating the parameters of the model under the null hypothesis of the cointegration rank $r = 1, 2, \dots, p - 1$. This step provides consistent estimates of the cointegration degree, the cointegration vectors, the speed of adjustment to the equilibrium parameters and the common trends. In the second step we carry out a sup-likelihood ratio test of no-cointegration on the estimated $p - r$ common trends that are not cointegrated under the null. The cointegration degree is re-estimated in the second step to allow for new cointegration relationships with different memory. We augment the error correction model in the second step to control for stochastic trend estimation effects from the first step. The critical values of the tests proposed depend only on the number of common trends under the null, $p - r$, and on the interval of the cointegration degrees b allowed, but not on the true cointegration degree b_0 . Hence, no additional simulations are required to approximate the critical values and this procedure can be convenient for practical purposes. In a Monte Carlo study we analyze the finite sample properties of different specifications of the correction terms and compare our procedure with alternative methods.

Keywords: Error correction model, Gaussian VAR model, Likelihood ratio tests, Maximum likelihood estimation. **JEL:** C12, C15, C32.

*corresponding author's address: VU Amsterdam – FEWEB, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.

1 Introduction

Fractional cointegration is a generalization of standard models with $I(1)$ integrated time series and $I(0)$ cointegration relationships. In general, observed time series can display different degrees of integration, while equilibrium relationships can be just characterized by a lower persistence or integration degree than the levels, perhaps allowing different values if there is more than one equilibrium. Much focus of the literature has been placed on parameter estimation, using both semiparametric (e.g. Marinucci and Robinson (2001)) or parametric methods, which specify also short run dynamics (e.g. Robinson and Hualde (2003); Johansen and Nielsen (2012)). However, the estimation of the parameters of the cointegrated model assumes the knowledge of a positive number of cointegration relationships (and regression based methods also take the dependent variables as given), so the related testing problems on the existence of cointegration and the cointegration rank have been also investigated in the literature.

Fractional cointegration testing has been analyzed from different perspectives. One approach focuses on the estimation of the memory parameters, see e.g. Marinucci and Robinson (2001), Nielsen (2004), Gil-Alaña (2003), Robinson (2008). Marmol and Velasco (2004) and Hualde and Velasco (2008) compare OLS and different GLS-type estimates of the cointegrating vector to construct a test statistic. Lasak (2010) directly exploits a Fractional Vector Error Correction Model (FVECM) to propose Likelihood Ratio (LR) tests.

Recent work has proposed fractional cointegration tests inspired by multivariate methods. Breitung and Hassler (2002) solve a generalized eigenvalue problem of the type considered in the Johansen's procedure for developing multivariate score tests of fractional integration, see Johansen (1988, 1991, 1995) and Nielsen (2005). Avarucci and Velasco (2009) propose to exploit a parametric FVECM for the development of Wald tests of the cointegration rank. There have been several semi-parametric proposals that focus on spectral matrix estimates, see Robinson and Yajima (2002), Chen and Hurvich (2003, 2006) and Nielsen and Shimotsu (2007).

We estimate the cointegration rank from a parametric perspective and base it on the specification of a FVECM. We rely on pseudo-LR tests based on restricted maximum likelihood (ML) estimates of the system. It is in contrast to Avarucci and Velasco (2009), who investigate the rank of unrestricted OLS estimates. We propose to perform a sequence of hypothesis tests based on a new two-stage procedure. It extends the results of testing the hypothesis of no-cointegration in Lasak (2010) and of estimating the fractionally cointegration systems in Lasak (2008) and Johansen and Nielsen (2012). The first step of the proposed procedure consists in estimating the parameters of the FVECM under the null hypothesis of the cointegration rank $r = 1, 2, \dots, p - 1$. Under the null of the cointegration rank r , this

estimation step provides consistent estimates of the cointegration degree, the cointegration vectors and the speed of the adjustment to the equilibrium parameters, together with an approximation to the common trends. In the second step, we implement the no-cointegration sup LR tests considered in Lasak (2010) to the estimated common trends. We neglect the order of cointegration estimated in the first step, to allow for different persistences in the cointegration relationships and to permit the use of known asymptotic distribution. The critical values depend on the number of common trends under the null hypothesis of rank r , and the interval of possible cointegration degrees, but not on the true cointegration degree.

However, to obtain the same asymptotic distribution as for the no-cointegration tests in Lasak (2010), we need to augment the error correction model in the second step to account for terms spanned by the cointegrating residuals from the first step. For this purpose, we include additional regressors and show that, when they are included, parameter estimation from the first step is asymptotically negligible. We consider different feasible specifications of the correction terms depending on the first step estimation and analyze the performance of the proposed new procedures in finite samples. We compare our approach with the LR rank test of Johansen and Nielsen (2012). The asymptotic distribution of their LR test depends on the true degree of cointegration and it is imposed that all cointegration relationships share the same memory. We also compare our tests with the benchmark LR tests based on the standard VECM that assumes that the degree of cointegration is known and equal to one, see Johansen (1988, 1991, 1995).

The remainder of the paper is organized as follows. Section 2 describes ML estimation of the basic FVECM and testing the null of no cointegration. Section 3 presents our new two-step procedure for testing the higher ranks with possibly different memory. In Section 4 we present the rank testing methods for the model with short run dynamics. Section 5 presents results of the Monte Carlo analysis. Section 6 contains the empirical analysis of the term structure of the interest rates. Section 7 concludes. Appendix contains the proofs of Theorems 1, 3 and Corollary 2. Separate Technical Appendix available at <http://creates.au.dk/research/research-papers/supplementary-downloads/rp-2013-08/> contains Tables 7-14 with Monte Carlo results discussed in Section 5.

2 Analysis of fractional systems

We first discuss the inference of the basic Fractional Vector Error Correction Model (FVECM). It includes the fractional representations proposed in Granger (1986), Johansen (2008, 2009) and Avarucci and Velasco (2009). For a $p \times 1$ vector time series X_t , we consider the following representation

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

where $\Delta = 1 - L$, L being the lag operator, d and b are respectively orders of integration and cointegration satisfying $0 < b \leq d$, $L_b = 1 - \Delta^b$, α and β are $p \times r$ full rank matrices, $0 \leq r \leq p$, ε_t is a $p \times 1$ vector of independent and identically distributed errors with zero mean and positive definite variance-covariance matrix Ω . Note that L_b filters the series X_t in such a way, that the filtered series depends on all lagged values of X_t and does not depend on the current value in period t . We assume that all initial values are set to zero, $X_t = \varepsilon_t = 0$, $t \leq 0$, so Δ^d could be replaced by Δ_+^d , denoting the truncated version of the fractional filter to positive values, $\Delta_+^d X_t = \Delta^d X_t 1\{t > 0\}$, which is necessary when $d \geq 0.5$ to accommodate non square summable filters.

Equation (1) corresponds to a (zero mean) fractional vector autoregressive $\text{VAR}_{d,b}(0)$ model in Johansen and Nielsen (2012) and implies under some further conditions that there exists r , $0 < r < p$, different linear combinations β of time series X_t , that are integrated of order $d - b$, $I(d - b)$, while X_t is integrated of order d , $I(d)$. In Johansen and Nielsen (2012) the time series X_t is said to be a cofractional process of order $d - b$ with $r > 0$ being the cofractional or cointegration rank. The matrix α contains the speed of adjustment to the equilibrium coefficients and β the cointegrating relationships. If $r = 0$ this implies that $\Pi = \alpha\beta' = 0$ or $b = 0$, so that X_t is integrated of order d and no trivial linear combination of X_t has smaller order of integration. In the special case $r = p$, the matrix $\Pi = \alpha\beta'$ is unrestricted.

We denote the true values of the parameters with a 0 subscript and consider for presentation the case $d_0 = 1$, which is the most relevant in applications. It does not affect the generality, since all arguments would apply directly to $\Delta^{\hat{d}}X_t$ instead of ΔX_t . However the parameter d could be estimated jointly with the other model parameters if $d_0 \neq 1$ and it is unknown. We restrict the true value of b to $b_0 \in (0.5, 1]$ when $r > 0$, as in Łasak (2010), so all potential cointegrating relationships are (asymptotically) stationary. Note that when $r > 1$ all cointegrating relationships implied by the $\text{VAR}_{1,b}(0)$ model of Johansen and Nielsen (2012) have the same order of integration $1 - b$. We do not maintain this restriction in the second step of our testing procedure.

Łasak (2010) has solved the problem of testing whether the system (1) is cointegrated. The estimation under the assumption that the cointegration rank r , $r > 0$, is known is considered in Łasak (2008) and Johansen and Nielsen (2012) adapting Johansen's (1995) procedure to the fractional VECM. For that, let's define $Z_{0t} = \Delta X_t$ and $Z_{1t-1}(b) = \left(\Delta_+^{-b} - 1\right) \Delta X_t = \Delta_+^{1-b} L_b X_t$. Note that $Z_{1t-1}(b)$ does not depend on data at time t . The model (1) expressed in these variables becomes

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

Then, the log-likelihood function \mathcal{L}_r for the model (1), under the hypothesis of r cointegrating relationships and the gaussianity of ε_t , is given, apart from a constant, by

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

We define the sample cross moments by

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

where S_{ij} is a function of b except when $i = j = 0$. For given values of $r > 0$, b and β , parameters α and Ω are estimated by regressing Z_{0t} on $\beta'Z_{1t-1}(b)$ and

$$\hat{\alpha}(\beta(b)) = S_{01}(b)\beta(\beta'S_{11}(b)\beta)^{-1}, \quad (2)$$

while

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

Plugging these estimates into \mathcal{L}_r we get

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

where $L_r = \exp(\mathcal{L}_r)$ is the likelihood up to a multiplicative constant. For fixed b the maximum of the likelihood is obtained by solving the eigenvalue problem

$$|\lambda(b)S_{11}(b) - S_{10}(b)S_{00}^{-1}S_{01}(b)| = 0 \quad (3)$$

for eigenvalues $\lambda_i(b)$ (ordered by decreasing magnitude for $i = 1, \dots, p$) and eigenvectors $v_i(b)$, such that

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

and $v_j'(b)S_{11}(b)v_i(b) = 1$ if $i = j$ and 0 otherwise. The eigenvectors diagonalize the matrix $S_{10}(b)S_{00}^{-1}S_{01}(b)$, i.e.,

$$v_j'(b)S_{10}(b)S_{00}^{-1}S_{01}(b)v_i(b) = \lambda_i(b)$$

for $i = j$ and they are equal to 0 otherwise. Thus by simultaneously diagonalizing the matrices $S_{11}(b)$ and $S_{10}(b)S_{00}^{-1}S_{01}(b)$ 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 estimate b by maximizing the likelihood in a compact set $\mathcal{B} \subset (0.5, 1]$, i.e.

$$\hat{b}_r = \arg \max_{b \in \mathcal{B}} L_r(b),$$

where

$$L_r(b) = \left[|S_{00}| \prod_{i=1}^r (1 - \lambda_i(b)) \right]^{-T/2}. \quad (4)$$

Expression (4) can be used to construct the sequence of LR tests for testing the fractional cointegration rank in the model (1). The first step is to test the null of no cointegration,

$$H_0 : \text{rank}(\Pi) = 0.$$

We can test it against two different alternatives, full cointegration rank of the impact matrix $\Pi = \alpha\beta'$,

$$H_p : \text{rank}(\Pi) = p,$$

or one extra cointegrating relationship,

$$H_1 : \text{rank}(\Pi) = 1.$$

Lasak (2010) has described how to test H_0 against H_p and H_1 . The LR statistic for testing H_0 against H_p (sup trace test) is defined by

$$\mathcal{LR}_T^p(0|p) = -2 \log \left[L_0 / L_p(\hat{b}_p) \right] = -T \sum_{i=1}^p \log[1 - \lambda_i(\hat{b}_p)], \quad (5)$$

where

$$\hat{b}_p = \arg \max_{b \in \mathcal{B}} L_p(b),$$

with L_p being defined as the likelihood under the hypothesis of rank p and $L_0 = |S_{00}|^{-T/2}$ is the likelihood when $r = 0$.

Alternatively, the LR statistic for testing H_0 against H_1 (sup maximum eigenvalue test) is defined by

$$\mathcal{LR}_T^p(0|1) = -2 \log \left[L_0 / L_1(\hat{b}_1) \right] = -T \log[1 - \lambda_1(\hat{b}_1)], \quad (6)$$

where

$$\hat{b}_1 = \arg \max_{b \in \mathcal{B}} L_1(b)$$

and L_1 denotes the likelihood under the hypothesis of rank 1. Recall that under the null of no cointegration ($r = 0$) we cannot hope that \hat{b}_1 or \hat{b}_p estimate consistently a nonexisting

true value of b , and because of that the LR tests (5) and (6) could be interpreted as sup LR tests, in the spirit of Davies (1977) and Hansen (1996).

We state Assumption 1 as in Łasak (2010).

Assumption 1 ε_t are independent and identically distributed vectors with mean zero, positive definite covariance matrix Ω , and $E\|\varepsilon_t\|^q < \infty$, $q \geq 4$, $q > 2/(2\underline{b} - 1)$, $\underline{b} = \min \mathcal{B} > 0.5$, where $\mathcal{B} \subset (0.5, 1]$ is a compact set.

Under Assumption 1, Łasak (2010) has proved that tests (5) and (6) have the following asymptotic distributions

$$\mathcal{LR}_T^p(0|p) \xrightarrow{d} \sup_{b \in \mathcal{B}} \text{trace}[\mathcal{L}(b)] \stackrel{def}{=} J_p \quad (7)$$

and

$$\mathcal{LR}_T^p(0|1) \xrightarrow{d} \sup_{b \in \mathcal{B}} \lambda_{\max}[\mathcal{L}(b)] \stackrel{def}{=} E_p, \quad (8)$$

where

$$\mathcal{L}(b) = \int_0^1 (dB) B'_b \left[\int_0^1 B_b B'_b du \right]^{-1} \int_0^1 B_b (dB)', \quad (9)$$

B_b is a p -dimensional standard fractional Brownian motion with parameter $b \in \mathcal{B}$,

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

$B = B_1$ is a standard Brownian motion on the unit interval. Note the difference in notation, B_b corresponds to B_{b-1} in Johansen and Nielsen (2012).

Łasak (2010) has simulated quantiles of the asymptotic distributions in (7) and (8), which we reproduce in Tables 1 and 2 for the convenience of the reader. Note that the asymptotic distributions in (7) and (8) and thus critical values of the tests (5) and (6) depend on the interval $\mathcal{B} = [0.5 + \varepsilon; 1]$, with $\varepsilon > 0$ and small, of possible values of b , on which we maximize the likelihood. The bounds of interval \mathcal{B} are determined in order to allow deviations from equilibrium (cointegrating residuals) to be of all possible orders of integration that would be asymptotically stationary. For practical purposes we can simulate the tables of critical values for each d with $\mathcal{B} = [d - 0.5 + \varepsilon; d]$.

Tables 1-2 have been simulated with 100,000 repetitions and using the approximation of fractional Brownian motion by fractionally integrated series based on i.i.d. Gaussian noise of length 1000. To maximize the likelihood function, the MaxSQPF procedure has been used in Ox, see Doornik and Ooms (2007) and Doornik (2009 a,b), and optimization has been done on the interval $\mathcal{B} = [0.5; 1]$, since the critical values converge to a limit when $\varepsilon \rightarrow 0$ (and also when $d \rightarrow \infty$), see Łasak (2010).

Table 1. Quantiles of J_p for the sup trace test.

p	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
1	0.0006	0.0035	0.012	0.045	0.87	3.71	4.98	6.28	8.07
2	0.42	0.71	1.07	1.65	5.37	10.92	12.84	14.67	16.90
3	2.64	3.50	4.41	5.77	13.44	21.73	24.30	26.67	29.64
4	7.06	8.68	10.25	12.53	25.81	36.72	39.95	42.86	46.52
5	13.92	16.23	18.63	22.09	42.59	55.88	59.76	63.24	67.49
6	23.38	26.62	29.95	35.11	63.67	78.87	83.19	87.14	91.93
7	35.56	40.02	44.60	52.37	88.81	105.87	110.84	115.32	120.89
8	50.74	56.53	63.18	77.58	117.84	136.83	142.47	147.61	153.80
9	69.15	76.69	87.20	124.92	150.68	171.61	177.81	183.49	190.27
10	91.42	102.65	128.49	163.12	187.27	210.32	217.29	223.54	230.94

Table 2. Quantiles of E_p for the sup maximum eigenvalue test.

p	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
1	0.0006	0.0035	0.012	0.045	0.87	3.71	4.98	6.28	8.07
2	0.37	0.62	0.93	1.44	4.73	9.86	11.72	13.45	15.67
3	1.87	2.50	3.16	4.15	9.37	15.85	18.01	19.98	22.53
4	4.16	5.07	6.06	7.47	14.26	21.81	24.27	26.49	29.38
5	6.85	8.08	9.35	11.14	19.36	27.72	30.40	32.83	35.91
6	10.00	11.52	13.00	15.11	24.60	33.47	36.28	38.87	41.90
7	13.16	14.92	16.67	19.17	29.89	39.49	42.36	45.06	48.45
8	16.69	18.66	20.64	23.36	35.29	45.29	48.48	51.20	54.62
9	20.41	22.48	24.59	27.83	40.72	51.21	54.35	57.30	60.78
10	24.08	26.30	28.72	32.31	46.19	57.02	60.31	63.43	67.21

When we reject the null hypothesis H_0 of no cointegration we only get the information that the system (1) is cointegrated, but we do not know how many cointegration relationships has X_t , so we need to proceed further and solve the problem of the cointegration rank estimation. For testing the cointegration rank r against rank p , $r = 1, \dots, p-1$ in model (1) we can use the general LR tests proposed by Johansen and Nielsen (2012) based on the solutions of the

eigenvalue problem (3) under both hypothesis, i.e.:

$$\mathcal{LR}_T^p(r|p) = -2T \log \left[L_r(\hat{b}_r) / L_p(\hat{b}_p) \right] = -T \left\{ \sum_{i=1}^p \log[1 - \lambda_i(\hat{b}_p)] - \sum_{i=1}^r \log[1 - \lambda_i(\hat{b}_r)] \right\}, \quad (10)$$

where estimates of the cointegration degree under the null (\hat{b}_r) and under the alternative (\hat{b}_p) might be different in general. The asymptotic distribution of the test statistic $\mathcal{LR}_T(r|p)$,

$$\mathcal{LR}_T(r|p) \xrightarrow{d} \text{trace} \{ \mathcal{L}(b_0) \},$$

depends on the true cointegration degree b_0 and Johansen and Nielsen (2012) suggest using the computer program by MacKinnon and Nielsen (2013) to obtain critical values for the tests. Lyhagen (1998) has also tabulated the asymptotic distribution of the trace test statistic for testing H_r against H_p under the assumption that the true cointegration degree b_0 is known. In the next section we propose a new two-step procedure that leads to the tests with the same asymptotic distributions as tests (5) and (6) in Lasak (2010) that do not depend on any nuisance parameters other than the number of the common trends $p - r$ and the interval \mathcal{B} which can be fixed in our case to be arbitrarily close to $(0.5, 1]$. Moreover this new procedure does not require the assumption that all cointegrating relationships share the same memory.

3 New tests for the cointegration rank

In this section we propose a new two-step procedure to establish the cointegration rank in the FVECM given in (1) when b_0 is unknown and $d_0 = 1$. This procedure has two main features: (i) it leads to the known asymptotic distribution (9); and (ii) it allows the cointegrating relationships under the alternative to have different memory compared to the null ones.

The first step of our test procedure consists in the parameter estimation for model (1) under the null hypothesis H_r of cointegration rank r . This provides consistent estimates of b and of the decomposition $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices, as in Theorem 10 of Johansen and Nielsen (2012). Then we compute (super) consistent estimates of the full rank matrix $p \times (p - r)$ β_\perp satisfying $\beta'_\perp\beta = 0$ and exploit the fact that under the null $\beta'_\perp X_t$ is not cointegrated in any direction. From Granger's Representation Theorem for the cofractional VAR model, see Theorem 2 in Johansen and Nielsen (2012),

$$X_t = C\Delta_+^{-d}\varepsilon_t + \Delta_+^{b-d}Y_t^+,$$

where $C = \beta_\perp (\alpha'_\perp\beta_\perp)^{-1} \alpha'_\perp$ and Y_t^+ is fractional of order zero, with initial conditions set to

zero and $\det(\alpha'_\perp \beta_\perp) \neq 0$. Then

$$\beta'_\perp X_t = \beta'_\perp C \Delta_+^{-d} \varepsilon_t + \Delta_+^{b-d} \beta'_\perp Y_t^+, \quad (11)$$

where $\beta'_\perp C$ is of rank $p - r$ under the null, and $\beta'_\perp X_t$ is just a rotation of the $p - r$ common trends $\alpha'_\perp \Delta_+^{-d} \varepsilon_t$ plus $I(d - b)$ components.

For given $\hat{\beta}_\perp$, we propose to implement sup tests (5) and (6) described in Section 2 based on the $p - r$ series $\hat{\beta}'_\perp X_t$ and with critical values from the E_{p-r} and J_{p-r} distributions (see (7) and (8)) as for no-cointegration testing in Lasak (2010). Given the superconsistency of $\hat{\beta}$ and therefore of $\hat{\beta}_\perp$, the estimation of $\hat{\beta}$ in the first step does not alter the asymptotic distribution of the test statistic in the second step. Under the alternative, $\hat{\beta}'_\perp X_t$ contains at least one further cointegrating relationship since $\hat{\beta}$ only estimates an r dimensional cointegrating space, so the sup tests should be able to detect it consistently, without imposing any a priori cointegration order.

This approach has two particular characteristics. First, when searching for further cointegration relationships among the estimated common trends, it does not use the information provided by the estimate \hat{b}_r of the persistence of the existing cointegrating relationships under the null. Second, the linear combinations $\beta'_\perp X_t$ are not cointegrated $I(1)$ series under the null, but they are also not pure $I(1)$ processes as is implied by (1) for the original series X_t when $\text{rank}(\Pi) = 0$. In fact, they are contaminated by $I(1 - b)$ series, as in (11). Test procedures should take into account this new feature of the projections $\beta'_\perp X_t$ compared to the data generated under (1) if the same asymptotic null distribution is to be preserved. This correction is derived for the case of a triangular systems, which are easier to handle as we will show next.

Consider a triangular representation of fractionally cointegrated $I(1)$ series with rank r ,

$$\begin{aligned} \beta' X_t &= \Delta_+^{b-1} u_{1t}, \\ \gamma' X_t &= \Delta_+^{-1} u_{2t}, \end{aligned} \quad (12)$$

where β and γ are, respectively, $p \times r$ and $p \times (p - r)$ matrices, $u_t = (u'_{1t}, u'_{2t})'$ is *iid* $(0, \Sigma)$, with $\Sigma > 0$, and $\Theta = (\beta \dot{\vdash} \gamma)$ has full rank p . Then we have that

$$\begin{aligned} \beta' \Delta X_t &= \left(\Delta - \Delta_+^{1-b} \right) \beta' X_t + u_{1t}, \\ \gamma' \Delta X_t &= u_{2t}, \end{aligned}$$

so that from the identity $\gamma_{\perp} (\beta' \gamma_{\perp})^{-1} \beta' + \beta_{\perp} (\gamma' \beta_{\perp})^{-1} \gamma' = I_p$, we obtain that

$$\Delta X_t = \alpha \beta' \left(\Delta_+^{-b} - 1 \right) \Delta X_t + \varepsilon_t,$$

where $\alpha = -\gamma_{\perp} (\beta' \gamma_{\perp})^{-1}$ and $\varepsilon_t = K u_t$, $K = \left(\gamma_{\perp} (\beta' \gamma_{\perp})^{-1} : \beta_{\perp} (\gamma' \beta_{\perp})^{-1} \right)$. Therefore

$$X_t = \Theta^{-1'} \begin{pmatrix} \Delta_+^{b-1} u_{1t} \\ \Delta_{-1} u_{2t} \end{pmatrix},$$

and hence that

$$\beta'_{\perp} X_t = M_1 \Delta_+^{b-1} u_{1t} + M_2 \Delta_+^{-1} u_{2t}, \quad (13)$$

where M_2 is a $(p-r) \times (p-r)$ full rank matrix and there is no β_1 such that $\beta'_1 (\beta'_{\perp} X_t)$ is an $I(1-b_1)$ process, any $b_1 > 0$, i.e. a process less integrated than $\beta'_{\perp} X_t$. As far as $M_1 \neq 0$, $\beta'_{\perp} X_t$ contains some $I(1-b)$ terms, by contrast with equation (1) when $r=0$ and $\Pi=0$. The interesting feature of the triangular model is that these $I(1-b)$ terms are spanned by the cointegrating residuals $\beta' X_t = \Delta_+^{b-1} u_{1t}$.

Therefore, from representation (11), the common trends $\beta'_{\perp} X_t$ do not follow exactly a FVECM (1) under H_r , despite non being cointegrated. Then in a reduced rank regression of $\hat{V}_{0t} = \hat{\beta}'_{\perp} \Delta X_t$ on $\hat{V}_{1t-1}(b_1) = (1 - \Delta_+^{-b_1}) \hat{\beta}'_{\perp} \Delta X_t$ with true coefficient $\Pi_1 = 0$, we need to incorporate additional regressors to correct for terms that are at most $I(0)$. For this purpose we consider the following decomposition based on the triangular model in (12),

$$\beta'_{\perp} \Delta X_t = M_1 \left(\Delta_+^b - 1 \right) u_{1t} + M_1 u_{1t} + M_2 u_{2t},$$

which indicates that we have to correct for the predictable part of $\beta'_{\perp} \Delta X_t$, i.e. $(\Delta_+^b - 1) u_{1t}$. This correction can be done using directly proxies for u_{1t} and then transforming them adequately.

A first possibility is to fit the model

$$\hat{V}_{0t} = \Pi_1 \hat{V}_{1t-1}(b_1) + \Phi \left(1 - \Delta_+^{-\hat{b}} \right) \hat{\beta}'_{\perp} \Delta X_t + \text{error}_t \quad (14)$$

by reduced rank regression, exploiting that under H_r and (12), $(1 - \Delta_+^{-b}) \beta' \Delta X_t = (\Delta_+^b - 1) u_{1t}$ incorporates the predictable component of $\beta'_{\perp} \Delta X_t$, see (13). Equation (14) takes into account both the directions in $\hat{\beta}$ and $\hat{\beta}_{\perp}$ of the original series ΔX_t . We could use estimate of $\beta' \Delta X_t = \Delta_+^b u_{1t} = (\Delta_+^b - 1) u_{1t} + u_{1t}$ instead of $(1 - \Delta_+^{-b}) \beta' \Delta X_t$ in (14), as it also incorporates the contemporaneous value of u_{1t} , however it can be correlated with $e_t = u_{2t}$, leading to biased estimates of Π_1 .

A second alternative to augment the FECVM of \hat{V}_{0t} is to use the first step residuals $\varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta})$ under H_r ,

$$\varepsilon_t(b, \alpha, \beta) = \left(I_p - \alpha\beta'(\Delta_+^{-b} - 1) \right) \Delta X_t.$$

Then, for identifying the components u_{1t} out of the rotated errors $\varepsilon_t = Ku_t$, we could just compute the linear projection of $\hat{\beta}'\Delta X_t$ given $\varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta})$,

$$\tilde{u}_{1t} = \sum_{t=1}^T \hat{\beta}'\Delta X_t \varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta})' \left(\sum_{t=1}^T \varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta}) \varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta})' \right)^{-1} \varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta}),$$

which proxies the contemporaneous contribution of u_{1t} in $\beta'\Delta X_t$. Then we use the filtered series $(\Delta_+^{\hat{b}} - 1)\tilde{u}_{1t}$, to estimate the regression

$$\hat{V}_{0t} = \Pi_1 \hat{V}_{1t-1}(b_1) + \Phi \left(\Delta_+^{\hat{b}} - 1 \right) \tilde{u}_{1t} + \text{error}_t, \quad (15)$$

without imposing $b_1 = \hat{b}$.

A third possibility is to use in (15) directly the whole $p \times 1$ vector $\varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta})$ instead of the $p - r$ linear combinations \tilde{u}_{1t} ,

$$\hat{V}_{0t} = \Pi_1 \hat{V}_{1t-1}(b_1) + \Psi \left(\Delta_+^{\hat{b}} - 1 \right) \varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta}) + \text{error}_t. \quad (16)$$

Then the corresponding LR test statistics on the rank of Π_1 from either (14), (15) or (16) are denoted as $\mathcal{LR}_T^{p-r}(0|1)$ and $\mathcal{LR}_T^{p-r}(0|p-r)$, where we replace X_t by the $p - r$ series $\hat{\beta}'_{\perp} X_t$ and we first project the involved series with a regression on proxies of $(\Delta_+^b - 1)u_{1t}$. Theorem 1 shows that the asymptotic distributions of these test statistics are E_{p-r} and J_{p-r} , respectively, since replacing β_{\perp} by $\hat{\beta}'_{\perp}$ and b_0 by \hat{b} has no asymptotic impact on the test statistics under the assumption

$$\hat{\beta}_{\perp} - \beta_{\perp} = O_p(T^{-b_0}), \quad \hat{b} - b_0 = O_p(T^{-1/2}), \quad (17)$$

see the proof in the Appendix. Similarly, the proxies of $(\Delta_+^b - 1)u_{1t}$ based on consistent estimates correct for the additional term in the second step of FVECM.

Theorem 1 *Under Assumption 1, model (12) and (17), the LR tests based on regressions (14), (15) or (16) for testing $\text{rank}(\Pi_1) = 0$, satisfy under H_r ,*

$$\begin{aligned} \mathcal{LR}_T^{p-r}(0|1) &\xrightarrow{d} E_{p-r}, \\ \mathcal{LR}_T^{p-r}(0|p-r) &\xrightarrow{d} J_{p-r}. \end{aligned}$$

The proof of Theorem 1 is given in the Appendix. The asymptotic distribution is the same as the no-cointegration tests in Łasak (2010), only the degrees of freedom need to be adapted for the dimension of $\hat{\beta}'_{\perp} X_t$ under H_r . It does not depend on any further nuisance parameter other than the set \mathcal{B} . The requirement (17) on the rates of convergence is fulfilled for parametric maximum likelihood estimates, but other estimates could also be considered, see the discussion in Johansen and Nielsen (2012).

If the parameter d has to be estimated, then we replace \hat{V}_{0t} and $\hat{V}_{1t-1}(b_1)$ by $\hat{V}_{0t}(\hat{d}) = \hat{\beta}'_{\perp} \Delta^{\hat{d}} X_t$ on $\hat{V}_{1t-1}(b_1, \hat{d}) = (1 - \Delta_+^{-b_1}) \hat{\beta}'_{\perp} \Delta^{\hat{d}} X_t$ in the test statistics and possibly readjust the set \mathcal{B} . Then the following corollary applies.

Corollary 2 *The conclusions of Theorem 1 remain valid if $\Delta^{d_0} X_t$ is replaced by $\Delta^{\hat{d}} X_t$ and $\hat{d} - d_0 = O_p(T^{-1/2})$.*

The proof of Corollary 2 is given in the Appendix. Theorem 1 in Robinson and Hualde (2003) presents a similar result for the case where memory parameters are estimated.

For the analysis of the consistency of our tests we can consider the alternative hypothesis H_{r+r_1} generated by the model

$$\Delta X_t = (\alpha\beta' + \alpha_1\beta_1') \left(\Delta_+^{-b} - 1 \right) \Delta X_t + \varepsilon_t, \quad (18)$$

where the $p \times r$ matrices α and β are of rank r , while the $p \times r_1$ matrices α_1 and β_1 are of rank r_1 , $p - r \geq r_1 > 0$. When parameter estimation for the model (18) is carried out under the null hypothesis of rank r , the vector $\hat{\beta}$ can capture at most r out of the $r + r_1$ existing equilibrium relationships. Hence $\hat{\beta}'_{\perp} X_t$ still contains at least one further cointegration relationship. Then, the consistency of the test would follow from the correlation between $\hat{\beta}'_{\perp} \Delta X_t$ and $\left(\Delta_+^{-b_1} - 1 \right) \hat{\beta}'_{\perp} \Delta X_t$ under H_{r+r_1} for a range of values of b_1 and any full rank $p \times (p - r)$ matrix $\hat{\beta}_{\perp}$.

Interestingly, the analysis can be extended by considering the model

$$\Delta X_t = \alpha\beta' \left(\Delta_+^{-b_0} - 1 \right) \Delta X_t + \alpha_1\beta_1' \left(\Delta_+^{-b_1} - 1 \right) \Delta X_t + \varepsilon_t,$$

where $b_1 \neq b_0$, so the new cointegrating relationships have different persistence compared to the first r ones. This model has been first proposed in Lyhagen (1998), however the properties of this model have yet to be established.

4 Rank testing in ECM with short run noise

Model (1) can be enlarged to incorporate short run dynamics by introducing fractional lags of $\Delta_+^d X_t$ to produce a $\text{VAR}_{d,b}(k)$ model,

$$\Delta_+^d X_t = \Delta_+^{d-b} L_b \alpha \beta' X_t + \sum_{i=1}^k \Gamma_i L_b^i \Delta_+^d X_t + \varepsilon_t, \quad (19)$$

as in Johansen (2008). In this case $\Delta_+^d X_t$ follows a VAR model in the lag operator $L_b = (1 - \Delta_+^b)$ rather than in the usual lag operator $L = L_1$. Johansen and Nielsen (2012) show that the existence of a Granger representation for X_t depends on $\det(\alpha'_\perp \Gamma \beta_\perp) \neq 0$ with $\Gamma = I_p + \sum_{i=1}^k \Gamma_i$, $\Gamma_k \neq 0$, and on the properties of the matrix polynomial $\Psi(y) = (1 - y) I_p - \alpha \beta' y - \sum_{i=1}^k \Gamma_i (1 - y) y^i$. For this result it is assumed that $\det[\Psi(y)] = 0$ implies either $y = 1$ or $y \in \mathbb{C}_b$, where \mathbb{C}_b is the image of the mapping $y = 1 - (1 - z)^b$ for z in the unit disk. Then estimation of (19) requires initial regression of $\Delta_+^d X_t$ and $\Delta_+^{d-b} L_b X_t$ on the lags $L_b^i \Delta_+^d X_t$, $i = 1, \dots, k$ and usual reduced rank regression optimizing in b (and d).

Following Avarucci and Velasco (2009), we can also allow for short run correlation in the levels of X_t using ordinary lags, by assuming that the prewhitened series $X_t^\dagger = A(L) X_t$ satisfy the model (1), but we actually observe X_t , i.e.

$$\Delta_+^d X_t = \Delta_+^{d-b} L_b \alpha \beta' A(L) X_t + (I - A(L)) \Delta_+^d X_t + \varepsilon_t, \quad (20)$$

where $A(L) = I - A_1 L - \dots - A_k L^k$. This model can be shown to encompass triangular models used in the literature (cf. Robinson and Hualde (2003)) and has also nice representations if the roots of the equation $\det[A(z)] = 0$ are out of the unit circle, $d > b$. In fact, if X_t^\dagger is cointegrated with cointegrating vector β , X_t is also cointegrated with cointegrating vector in the same space spanned by β given that $A(1)$ is full rank.

Even under the assumption of known d , model (20) is nonlinear in $\Pi = \alpha \beta'$ and A_1, \dots, A_k , so ML estimation can not be performed using the usual two step procedure of Johansen to prewhiten first the differenced levels $Z_{0t} = \Delta X_t$ and the fractional regressor $Z_{1t-1}(b) = \Delta_+^{1-b} L_b X_t$ given a particular value of b . However, it is easier to estimate the unrestricted linear model (in A_j and A_j^*)

$$Z_{0t} = \alpha \beta^{*'} Z_{1t-1}(b) + \sum_{j=1}^k A_j^* \Delta Z_{1t-j}(b) + \sum_{j=1}^k A_j Z_{0t-j} + \varepsilon_t, \quad (21)$$

under the assumption of α and β^* being $p \times r$, without imposing $A_j^* = -\Pi \tilde{A}_j$. We have used the decomposition $A(L) = A(1) - \Delta \tilde{A}(L)$ so that the coefficients of $\tilde{A}(L) = \sum_{j=0}^{p-1} \tilde{A}_j L^j$

satisfy $\tilde{A}_j = \sum_{i=1+j}^p A_i$, $j = 0, \dots, p-1$. Here $\beta^* = A(1)' \beta$ spans the same cointegration space as β . The estimation procedure follows as in usual reduced rank regression but with an initial step to prewhiten the series Z_{0t} and $Z_{1t-1}(b)$ on lags of Z_{0t} and $\Delta Z_{1t-1}(b)$, which are at most $I(0)$. This estimate could be inefficient compared with the actual ML estimate, but is much simpler to compute and analyze.

Once we have estimated the model under the null H_r , $r > 0$, we need to build up a valid second step regression on $\beta'_\perp X_t$ to identify further cointegrating relationships. In a triangular model set up with the VAR modelization $A(L)X_t = X_t^\dagger$ in levels and X_t^\dagger generated by (12) we have that

$$X_t = (I - A(L))X_t + \Theta^{-1} \begin{pmatrix} \Delta_+^{b-1} u_{1t} \\ \Delta_+^{-1} u_{2t} \end{pmatrix}, \quad (22)$$

and therefore

$$\beta'_\perp X_t = \sum_{j=1}^k \beta'_\perp A_j X_{t-j} + M_1 \Delta_+^{b-1} u_{1t} + M_2 \Delta_+^{-1} u_{2t}, \quad (23)$$

where M_2 is full rank under H_r , so that $\beta'_\perp X_t$ contains some $I(1-b)$ terms if $M_1 \neq 0$.

Given the pseudo ML estimate of β , we can construct the linear combinations $\hat{V}_{0t} = \hat{\beta}'_\perp \Delta X_t$ and $\hat{V}_{1t-1}(b_1) = (1 - \Delta_+^{-b_1}) \hat{\beta}'_\perp \Delta X_t$ and propose a similar regression equation as for $k=0$, but in this case with the FVECM has to be enlarged by lags of ΔX_t and proxies of the $I(1-b)$ terms from (23),

$$\hat{V}_{0t} = \Pi_1 \hat{V}_{1t-1}(b_1) + \sum_{j=1}^k C_j \Delta X_{t-j} + \Phi \left(\Delta_+^{\hat{b}} - 1 \right) \tilde{u}_{1t} + e_t, \quad (24)$$

where \tilde{u}_{1t} is obtained from a projection of $\hat{\beta}'_\perp \Delta X_t$ on FVECM residuals $\varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta}, \hat{A}^*, \hat{A})$ from (21) to isolate the u_{1t} contribution in $\beta'_\perp \Delta X_t$, which might contain other predictable contributions at time t due to the autoregressive structure. As when $k=0$, we could replace \tilde{u}_{1t} by $\varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta}, \hat{A}^*, \hat{A})$, but, for instance, using $(1 - \Delta_+^{-\hat{b}}) \hat{\beta}'_\perp \Delta X_t$ might not lead to a correctly specified model due to the presence of lags of $\Delta_+^{-b} \Delta X_t$ in $\beta'_\perp \Delta X_t$.

Then, the asymptotic distributions of the maximum eigenvalue and trace test statistics, $\mathcal{LR}_T^{p-r}(0|1)$ and $\mathcal{LR}_T^{p-r}(0|p-r)$, remain E_{p-r} and J_{p-r} , respectively, if the first step estimates are consistent, as stated in Theorem 3.

Theorem 3 *Under Assumptions 1, (22), (17) and*

$$\hat{A}_i^* - A_i^* = O_p\left(T^{-1/2}\right), \quad \hat{A}_i - A_i = O_p\left(T^{-1/2}\right), \quad i = 1, \dots, p,$$

the LR tests for testing $\Pi_1 = 0$ based on regression (24) with \tilde{u}_{1t} possibly replaced by

$\varepsilon_t(\hat{b}, \hat{\alpha}, \hat{\beta}, \hat{A}^*, \hat{A})$, have the same asymptotic distribution as in Theorem 1 under H_r .

The proof of Theorem 3 is given in the Appendix.

5 Finite sample properties of cointegration rank tests

In this section we analyze the performance of the proposed new procedure in finite samples. We simulate a cointegrated trivariate system ($p = 3$), with $d = 1$, using the following triangular representation

$$X_t = \begin{pmatrix} I_r & \delta \\ 0 & I_{p-r} \end{pmatrix} \begin{pmatrix} \Delta_+^{b_0-1} u_{0t} \\ \Delta_+^{b_1-1} u_{1t} \\ \Delta_+^{-1} u_{2t} \end{pmatrix}, \quad t = 1, \dots, T, \quad (25)$$

for model (1). The innovations $u_t = (u'_{0t}, u'_{1t}, u'_{2t})'$ are standard Gaussian *iid*. Note that the triangular representation (25) implies FVECM (1) with

$$\alpha = \begin{pmatrix} -I_r \\ 0 \end{pmatrix} \text{ and } \beta' = (I_r \quad -\delta).$$

To investigate the empirical size of the tests we simulate (25) with cointegration rank $r = 1$ and cointegrating vector $\beta = [1 \ 0 \ -1]'$ and for the power study we add an extra cointegrating relationship $\beta_1 = [0 \ 1 \ -0.5]'$. Further we also consider the model with short run dynamics (20) and with $k = 1$. For this model we add to (25) the autoregression

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

with $Z_0 = 0$ and $A_1 = a I_p$, where $a = 0.5$ or $a = 0.8$.

We simulate the systems with the memory of the first cointegrating relationship determined by $b_0 = 0.51, 0.6, 0.7, 0.8, 0.9, 0.99$ and for power analysis we set the memory of the second cointegrating relationship with $b_1 = 0.51, b_1 = 0.9$ or $b_1 = b_0$. This way we can illustrate the power of the procedure when the memory of the second cointegrating relationship is relatively small, big and when it is the same as the memory of the first cointegrating relationship. We consider the cointegrated systems with the sample sizes of $T = 100, 200, 400$ and when needed 1000 observations. For all simulations we use Ox Metrics 6.3, see Doornik and Ooms (2007) and Doornik (2009 a,b) and we make 10,000 repetitions of each experiment.

We compare the performance of the tests discussed in this paper, i.e.:

1. New two step procedures that use either trace test $\mathcal{LR}_T^{p-r}(0|p-r)$ or maximum eigenvalue test $\mathcal{LR}_T^{p-r}(0|1)$ based on the FVECM for $\hat{\beta}'_{\perp} X_t$, with the following corrections:
 - (a) $\Delta \hat{\beta}' X_t$ as in (14).
 - (b) $(\Delta^{\hat{b}_0} - 1) \tilde{u}_{1t}$ as in (15).
 - (c) $(\Delta^{\hat{b}_0} - 1) \varepsilon_t (\hat{b}_0, \hat{\alpha}_0, \hat{\beta}_0)$ as in (16).
2. Trace and maximum eigenvalue LR tests based on the standard VECM with $d = b = 1$ like in Johansen (1988, 1991, 1995), called Johansen's trace and Johansen's maximum eigenvalue tests.
3. Trace LR test $\mathcal{LR}_T^p(r|p)$ in (10) proposed by Johansen and Nielsen (2012), where estimation is restricted to $d = 1$ and critical values are obtained from the computer program of MacKinnon and Nielsen (2013) with ML estimate of b_0 rounded to a decimal point.

The tests in 1.(a) are not justified when $k = 1$, since $\Delta \hat{\beta}' X_t$ is affected by the autoregressive structure. Johansen's tests in 2. are not justified for the data generating process (25), as they are based on the misspecified model. However we check their performance, since they are included in most of econometric packages and they are routinely used by practitioners. Johansen and Nielsen (2012) test in 3. is correctly specified when $k = 0$, and it is a natural benchmark for comparison.

The results of our Monte Carlo simulations when $k = 0$ are presented in Tables 3-4 in this section and when $k = 1$ in Tables 7-14 in the Technical Appendix. Tables 3 and Tables 7-8 provide 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 Table 4 and Tables 9-14.

When $k = 0$ the new two step procedures are undersized (empirical size around 3%) for all sample sizes considered. Trace LR test by Johansen and Nielsen (2012) is usually oversized, but size distortions are decreasing with sample size T and true value b_0 . Johansen's LR tests have size close to the nominal 5% in all considered cases, see Table 3.

When $k = 0$ all the procedures except for 1.(c) have very good power for all sample sizes T and all true values of b_0, b_1 . The power of 1.(c), like the power of all other procedures, is increasing with sample size T , true value b_0 and true value b_1 , see Table 4.

When $k = 1$ Johansen's tests are undersized for small values of b_0 in smaller samples and size distortions in these cases increase with correlation a . The LR test of Johansen and Nielsen (2012) heavily overrejects in most cases considered and size distortions increase with sample size T and correlation a , but decrease with true value b_0 . When $k = 1$ the two step

procedures have higher empirical size than when $k = 0$, they are even slightly oversized in smaller samples, but their size tends to decrease with T , see Tables 7-8 in the Technical Appendix.

When $k = 1$ all procedures are less powerful than for $k = 0$, especially when the correlation a increases, see Tables 9-14 in the Technical Appendix. However the power increases with sample size T and true value of b_0 and b_1 in all cases. The LR test of Johansen and Nielsen has largest power among all in many cases, but it is not relevant as this test does not keep the size in this experiment. We find that the procedure 1.(c) has poorest power among all two-step tests, possibly because it uses a too large correction term that is highly correlated with the fractional correction term. Procedures 1.(a) and 1.(b) have a similar behavior to the one-step LR test with $d = b = 1$, however they seem to be more powerful when b_0 or b_1 is small. Therefore, our tests 1.(a)-1.(b) seem to behave very well in relative terms being able to exploit the difference between b_0 and b_1 .

Table 3. Percentage of rejections by trace test $\mathcal{LR}_T^{p-r}(r|p-r)$ and maximum eigenvalue test $\mathcal{LR}_T^{p-r}(r|r+1)$ in the two step procedure with corrections (a), (b) and (c), Johansen's trace and Johansen's maximum eigenvalue tests with $d = b = 1$ and trace test $\mathcal{LR}_T^p(r|p)$ of Johansen and Nielsen (2012) under the null hypothesis of cointegration rank $r = 1$ in $p = 3$ dimensional system with true $d_0 = 1$ and no lagged differences, i.e. $k = 0$. Nominal size 5%.

T	Test	b_0					
		0.51	0.60	0.70	0.80	0.90	0.99
100	$\mathcal{LR}_T^{p-r}(r r+1), d=1$ (a)	2.9	3.2	3.2	3.0	3.0	3.2
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$ (a)	2.7	2.9	3.2	2.8	3.0	3.1
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$ (b)	3.1	3.4	3.3	3.0	3.1	3.2
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$ (b)	2.8	3.0	3.2	3.0	3.1	3.2
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$ (c)	3.9	4.5	4.2	3.7	3.7	3.7
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$ (c)	3.6	4.3	4.1	3.6	3.6	3.6
	$\mathcal{LR}_T^p(r r+1), d=b=1$	4.6	4.9	5.1	4.9	4.9	5.4
	$\mathcal{LR}_T^p(r p), d=b=1$	4.5	4.8	5.1	5.1	5.0	5.4
	$\mathcal{LR}_T^p(r p)$	10.6	11.3	9.5	7.7	6.2	5.4

Table 3 continued.

T	Test		b_0					
			0.51	0.60	0.70	0.80	0.90	0.99
200	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(a)	3.0	3.1	2.9	3.0	3.0	2.8
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(a)	2.9	3.0	2.7	3.0	3.1	2.7
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(b)	3.1	3.2	2.9	3.0	3.1	2.8
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(b)	3.0	3.2	2.8	3.0	3.1	2.7
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(c)	3.6	3.9	3.6	3.3	3.4	3.1
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(c)	3.6	3.8	3.6	3.2	3.2	3.0
	$\mathcal{LR}_T^p(r r+1), d=b=1$		4.5	4.8	4.8	5.2	5.2	4.6
	$\mathcal{LR}_T^p(r p), d=b=1$		4.5	4.9	4.8	5.2	5.3	4.8
	$\mathcal{LR}_T^p(r p)$		8.9	8.3	8.2	7.1	6.3	4.8
400	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(a)	3.1	3.2	3.5	3.4	3.3	3.1
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(a)	3.2	3.3	3.3	3.1	3.3	2.8
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(b)	3.1	3.3	3.5	3.4	3.2	3.2
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(b)	3.2	3.2	3.3	3.1	3.2	2.9
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(c)	3.5	3.8	3.8	3.5	3.1	3.3
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(c)	3.5	3.9	3.6	3.2	2.9	3.2
	$\mathcal{LR}_T^p(r r+1), d=b=1$		4.5	4.8	5.2	5.2	5.0	5.0
	$\mathcal{LR}_T^p(r p), d=b=1$		4.6	4.9	5.1	5.0	5.1	5.1
	$\mathcal{LR}_T^p(r p)$		7.6	7.6	7.0	6.3	5.8	5.1
1000	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(a)	3.4	3.5	3.6	3.4	3.3	3.5
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(a)	3.4	3.3	3.4	3.4	3.1	3.2
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(b)	3.4	3.5	3.6	3.4	3.3	3.5
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(b)	3.4	3.3	3.5	3.3	3.2	3.3
	$\mathcal{LR}_T^{p-r}(r r+1), d=1$	(c)	3.7	3.4	3.5	3.5	3.1	3.4
	$\mathcal{LR}_T^{p-r}(r p-r), d=1$	(c)	3.7	3.4	3.3	3.4	3.1	3.3
	$\mathcal{LR}_T^p(r r+1), d=b=1$		4.8	4.9	5.1	5.1	5.1	4.9
	$\mathcal{LR}_T^p(r p), d=b=1$		4.9	5.0	5.0	5.3	4.9	5.0
	$\mathcal{LR}_T^p(r p)$		6.1	6.5	6.0	6.0	5.4	5.1

Table 4. Percentage of rejections by trace test $\mathcal{LR}_T^{p-r}(r|p-r)$ and maximum eigenvalue test $\mathcal{LR}_T^{p-r}(r|r+1)$ in the two step procedure with corrections (a), (b) and (c), Johansen's trace and Johansen's maximum eigenvalue tests with $d = b = 1$ and trace test $\mathcal{LR}_T^p(r|p)$ of Johansen and Nielsen (2012) under the hypothesis of cointegration rank $r = 2$ in $p = 3$ dimensional system with $T = 100$ observations, true $d_0 = 1$, no lagged differences, i.e. $k = 0$ and 2nd cointegrating relationship with the memory $b_1 = b_0$, 0.51 or 0.9. Nominal size 5%.

$T = 100$		Test	b_0					
			0.51	0.60	0.70	0.80	0.90	0.99
$b_1 = b_0$	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(a)	97.6	99.8	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(a)	96.8	99.8	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(b)	97.7	99.9	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(b)	97.0	99.8	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(c)	51.2	77.6	94.8	99.5	100	100
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(c)	49.5	76.4	94.0	99.3	100	100
	$\mathcal{LR}_T^p(r r+1), d = b = 1$		88.1	98.8	100	100	100	100
	$\mathcal{LR}_T^p(r p), d = b = 1$		88.3	98.9	100	100	100	100
	$\mathcal{LR}_T^p(r p)$		99.7	100	100	100	100	100
$b_1 = 0.51$	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(a)	97.6	99.1	99.5	99.6	99.5	99.5
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(a)	96.8	98.7	99.2	99.4	99.3	99.2
	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(b)	97.7	99.0	99.5	99.6	99.5	99.5
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(b)	97.0	98.8	99.2	99.4	99.4	99.3
	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(c)	51.2	61.8	67.3	71.9	75.1	76.2
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(c)	49.5	59.9	65.1	69.4	72.7	73.8
	$\mathcal{LR}_T^p(r r+1), d = b = 1$		88.1	94.1	95.4	95.6	95.7	95.8
	$\mathcal{LR}_T^p(r p), d = b = 1$		88.3	94.3	95.4	95.9	95.8	96.0
	$\mathcal{LR}_T^p(r p)$		99.7	99.9	99.8	99.5	99.0	98.1
$b_1 = 0.9$	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(a)	99.0	100	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(a)	98.7	100	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(b)	99.0	100	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(b)	98.7	100	100	100	100	100
	$\mathcal{LR}_T^{p-r}(r r+1), d = 1$	(c)	70.5	89.8	98.1	99.8	100	100
	$\mathcal{LR}_T^{p-r}(r p-r), d = 1$	(c)	68.8	88.3	97.7	99.7	100	100
	$\mathcal{LR}_T^p(r r+1), d = b = 1$		95.8	99.6	100	100	100	100
	$\mathcal{LR}_T^p(r p), d = b = 1$		95.9	99.8	100	100	100	100
	$\mathcal{LR}_T^p(r p)$		99.7	100	100	100	100	100

6 Analysis of the term structure of the interest rates

To illustrate the empirical relevance of the described methodology we reconsider the analysis of the term structure of the interest rates by Iacone (2009). There has been a lot of interest in this issue in the current literature, see for example Chen and Hurvich (2003) and Nielsen (2010).

A good model of the term structure of the interest rates is needed to measure the effects of the monetary policy and to price financial assets. It is an important tool for policy evaluation since the Federal Reserve operates in just one market, the one with contracts with very short maturity. Therefore, it is necessary to model the conduction of the monetary policy impulses to the rates of contracts with longer maturities. Modelling the interactions across rates is also important for the economic agents to forecast the effects of future monetary policy decisions on the price of financial assets. Soderlind and Svensson (1997) have discussed a practical example of how to extract the market's expectations on future policy rates from a given term structure, and how to use them to price financial instruments.

Cointegration has the appealing feature in the analysis of the term structure, because it makes possible to distinguish the high persistence of shocks to interest rates from the much lower persistence of shocks to the spreads. Standard cointegration in the context of modelling a vector of US dollar interest rates has been considered by Hall, Anderson and Granger (1992), Engsted and Tangaard (1994), Dominguez and Novales (2000).

However it has been argued that unit root model for the interest rates is often incompatible with monetary and finance theories, because it may imply a unit root model for the expected inflation rate as well. This is the case, for example, if the real interest rate is constant in the long run, or if the central bank sets the interest rate using a linear reaction function like the ones described by Taylor (1993) or by Svensson (1997). Such a strong persistence is hardly acceptable, because it implies that the central bank does not stabilize inflation.

We can allow for fractional integration instead. It permits to combine high persistence with mean reversion in the long run, and it maintains the possibility of the presence of a common stochastic term in multivariate processes. Fractional integration may also be motivated as the result of occasional breaks in an otherwise weakly autocorrelated process. This interpretation seems particularly appealing when modelling the interest rates because changes to the discount rate are infrequent. Granger and Hyung (2004) have shown that fractional integration and occasional breaks may in practice be indistinguishable and, following also a comment by Diebold and Inoue (2001), adopting fractional integration in a model may result in good forecasts.

We analyze the behavior of the US dollar interest rates with maturities of 1, 3 and

6 months (the London InterBank Offered Rate LIBOR) over the period 01/1963-04/2006. The data come from DataStream with identification codes being respectively USI60LDC, USI60LDD, USI60LDE. LIBOR is not affected by any regulation imposed by the central bank, and thus it is a typical measure of the cost of funds in US dollars. For this data set Iacone (2009) has found evidence that the three considered series share the same order of integration with estimated $\hat{d} = 0.88$. The test of Robinson and Yajima (2002) and local Whittle procedure of Robinson (1995) have been used to obtain this result. Iacone (2009) has also concluded the fractional cointegration with rank $r = 2$ in this system using procedures in Phillips and Ouliaris (1988) and Robinson and Yajima (2002).

However the integration order of the cointegrating residuals of two relations found by Iacone (2009) differ significantly, and the transmission of impulses is slower the longer distance (in maturity) from the market where the Federal Reserve is directly present, so a model that allows different b 's would be appropriate for this example. Lasak (2008) has analyzed three bivariate systems and has not imposed the assumption that both cointegration relationships share the same memory. The methodology developed in this paper enables us to test the rank directly in the 3-variate system (1) without imposing such assumption, as we now pursue.

We consider the basic version of the model presented in Section 3, as it seems to be a right choice looking at PACF of the processes. We have tested the existence of the breaks in levels of considered series using the test of Sibbertsen and Kruse (2009) and it has indicated no breaks in the series. All the tests considered in Section 5 have been computed and all confirm that this system is cointegrated with rank 2. The values of the test statistics when testing rank $r = 1$ are presented in Table 5.

Table 5. Values of the tests under the null of cointegration rank $r = 1$.

LR test	(a)	(b)	(c)	Johansen's	J-N
lambda	18.64	19.04	26.79	53.8	-
trace	18.64	19.04	27.28	54.5	30.7

We do not have the critical values for tests (a)-(c) when $\hat{d} = 0.88$, but instead we make use of Tables 1 and 2 that have been simulated for $d = 1$. Thus, the 5% critical values, that we use for the tests, are given in Table 6. Note the difference in the number of degrees of freedom, the two step procedures have just one degree of freedom and the other two tests have two degrees of freedom, in this particular example.

Table 6. Critical values of the tests under the null of cointegration rank $r = 1$.

	(a), (b), (c)	Johansen's	J-N
lambda	11.72	11.23	-
trace	12.84	12.32	10.95

The critical values for tests (a)-(c) when $\hat{d} = 0.88$ would be smaller than the ones we use, so we would reject the null hypothesis anyway. To obtain approximate critical values for tests (a)-(c), the asymptotic distributions (7) and (8) should be simulated with $\mathcal{B} = [\hat{d} - 0.5 + \varepsilon; \hat{d}]$, with $\varepsilon > 0$ and small.

We also estimate the cointegration vectors on the basis of all considered models, including the VECM with $d = b = 1$ and the FVECM $\Delta^d X_t = \alpha \beta' (\Delta^{-b} - 1) \Delta^d X_t + \varepsilon_t$, with $d = \hat{d} = 0.88$ imposed, which is justified by Corollary 2. The first cointegration relationship is common to all procedures, but the second one can be different. When we focus on the two-step procedure proposed in Section 3, the estimate of the second cointegrating relationship β_1 is found according to the formula $\hat{\beta}_1 = \beta'_\perp \beta_1^*$, where β_1^* comes directly from solving the eigenvalue problem (3) constructed on the basis of the transformed models (14), (15) and (16), denoted respectively as (a), (b) and (c). It turns out that all outcomes of the procedures imply the same cointegrating space spanned by

$$\hat{\beta}^{norm} = \begin{bmatrix} 1 & 1 \\ -0.98 & 0 \\ 0 & -0.96 \end{bmatrix}.$$

We can see that cointegrating parameters are very close to -1 , so the spreads can be computed as $s_t^{(j)} = i_t^{(j)} - i_t^{(1)}$, $j = 3, 6$. Iacone (2009) has estimated the orders of integration of these spreads using Local Whittle estimator of Robinson (1995) to be $s_t^{(3)} = 0.34$ and $s_t^{(6)} = 0.47$ and rejected the hypothesis that these orders are the same. Therefore the rank estimation methodology developed in this paper is suitable for this example, as it takes into account the possibility that the memories of the cointegration relations differ. The evidence of cointegration found is an important result because it means that the transmission of impulses along the term structure is fast enough to let the central bank to conduct an active monetary policy.

7 Conclusions

In this paper we have proposed a new procedure, based on sequential two-step LR tests, to establish the cointegration rank in a fractional system. The main novelty is that it allows

the cointegrating relationships under the alternative to have different memory compared to the null ones. It only needs a small modification of the cointegration testing routines after ML fitting of the model under the null of a certain cointegration rank. The asymptotic distributions of the test statistics in use are the same as for the no-cointegration testing, so there is no need for additional simulations to obtain the critical values, which can be seen as an advantage for empirical work.

We have investigated the performance of our procedure in finite samples for a simple fractionally cointegrated model and have compared it with the LR trace test of Johansen and Nielsen (2012) and with Johansen's LR trace and maximum eigenvalue tests. We have found that our tests control size and have an advantage in terms of power to detect extra cointegrating relationships in situations when the memories of the cointegration relations differ. Our new methodology can be adapted and further developed to include deterministic terms and to allow for the joint estimation of the unknown memory of the original series.

8 Acknowledgements

We are very grateful to Niels Haldrup, Lennart Hoogerheide, Søren Johansen, Siem Jan Koopman, Morten Ørregaard Nielsen and participants at various seminars and conferences for helpful comments and suggestions. The first author acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The second author acknowledges support from the Spanish Plan National de I+D+I (SEJ2007-62908 and ECO2012-31748). A preliminary version of this paper has been included in the Ph.D. Thesis of the first author defended at the Department of Economics and Economic History, UAB, Spain.

Appendix

Proof of Theorem 1. We demonstrate first that replacing β_{\perp} by $\hat{\beta}_{\perp}$ makes no difference asymptotically in the two step LR test statistics. LR tests statistics depend on properly normalized sample moments of dependent and independent variables in the regression models (14), (15) and (16), cf. (3). The result follows from Theorem 1 in Łasak (2010), after controlling for the projection on $(\Delta_{+}^b - 1)u_{1t}$ as in Lemma 10.3 in Johansen (1995), using the representation (13) for the dependent variable $V_{0t} = \beta'_{\perp} \Delta X_t$.

Set $V_{1t-1}(b_1) = (1 - \Delta_{+}^{-b_1})V_{0t}$, recalling the definition of $\hat{V}_{1t-1}(b_1)$ and using the true

β_{\perp} . First, we want to show that

$$T^{-b_1} \sum_{t=1}^T \hat{V}_{1t-1}(b_1) \hat{V}'_{0t} - T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1) V'_{0t} \rightarrow_p 0$$

uniformly for $b_1 \in \mathcal{B}$ if $\hat{\beta}_{\perp} - \beta_{\perp} = O_p(T^{-b_0})$. The difference on the left hand side is

$$T^{-b_1} \sum_{t=1}^T \left\{ \hat{V}_{1t-1}(b_1) - V_{1t-1}(b_1) \right\} V'_{0t} + T^{-b_1} \sum_{t=1}^T \hat{V}_{1t-1}(b_1) (\hat{V}'_{0t} - V'_{0t}). \quad (26)$$

The first term in (26) is equal to

$$\left(\hat{\beta}'_{\perp} - \beta'_{\perp} \right) T^{-b_1} \sum_{t=1}^T \left(1 - \Delta_+^{-b_1} \right) \Delta X_t V'_{0t} = o_p(1),$$

uniformly in $b_1 \in \mathcal{B}$ because $\hat{\beta}_{\perp} - \beta_{\perp} = O_p(T^{-b})$, $b > 0.5$, and $T^{-b_1} \sum_{t=1}^T \left(1 - \Delta_+^{-b_1} \right) \Delta X_t V'_{0t} = O_p(T^{1/2-\epsilon})$ uniformly in b_1 , $b_1 > 0.5$, for some $\epsilon > 0$ from (104) in Lemma A.9 in Johansen and Nielsen (2012).

The second term on the right hand side of (26) is

$$T^{-b_1} \sum_{t=1}^T \hat{V}_{1t-1}(b_1) \Delta X'_t \left(\hat{\beta}_{\perp} - \beta_{\perp} \right) = O_p(T^{-b}) T^{-b_1} \sum_{t=1}^T \hat{V}_{1t-1}(b_1) \Delta X_t,$$

and this is $O_p(T^{-b}) O_p(T^{1/2-\epsilon}) = o_p(1)$, uniformly in b_1 with $b > 0.5$, $\epsilon > 0$, using again Lemma A.9 in Johansen and Nielsen (2012).

Using the same ideas it can be shown that

$$T^{-2b_1} \sum_{t=1}^T \hat{V}_{1t-1}(b_1) \hat{V}'_{1t-1}(b_1) - T^{-2b_1} \sum_{t=1}^T V_{1t-1}(b_1) V'_{1t-1}(b_1) \rightarrow_p 0$$

uniformly for $b_1 \in \mathcal{B}$ and

$$T^{-1} \sum_{t=1}^T \hat{V}_{0t} \hat{V}'_{0t} - T^{-1} \sum_{t=1}^T V_{0t} V'_{0t} \rightarrow_p 0,$$

exploiting (103) and (102), respectively, in Lemma A.9 in Johansen and Nielsen (2012), so that the estimation of β_{\perp} in the first step has no impact on the asymptotic distribution of the test statistics.

We next show that replacing $(\Delta_+^b - 1) u_{1t}$ by $(\Delta_+^{\hat{b}} - 1) \hat{u}_{1t} = (1 - \Delta_+^{-\hat{b}}) \hat{\beta}' \Delta X_t$ in (14)

is also negligible asymptotically under (12). For that, it is enough to consider the differences

$$T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1) \left(\Delta_+^b - 1 \right) u'_{1t} - T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1) \left(\Delta_+^{\hat{b}} - 1 \right) \hat{u}'_{1t} \quad (27)$$

$$T^{-b_1} \sum_{t=1}^T u_t \left(\Delta_+^b - 1 \right) u'_{1t} - T^{-b_1} \sum_{t=1}^T u_t \left(\Delta_+^{\hat{b}} - 1 \right) \hat{u}'_{1t}, \quad (28)$$

since other terms appearing in the projections of $V_{1t-1}(b_1)$ and V_{0t} on $\left(\Delta_+^{\hat{b}} - 1 \right) \hat{u}'_{1t}$ could be dealt with in the same way. Now we can decompose (27) in

$$T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1) \left(\Delta_+^b - \Delta_+^{\hat{b}} \right) \hat{u}'_{1t} + T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1) \left(\Delta_+^{\hat{b}} - 1 \right) \Delta X'_t \left\{ \beta - \hat{\beta} \right\}. \quad (29)$$

The first term in (29) can be shown to be $o_p(1)$ uniformly in b_1 as in Robinson and Hualde (2003, Proposition 9), expanding $\left(\Delta_+^b - \Delta_+^{\hat{b}} \right) u_{1t} = \left(1 - \Delta_+^{\hat{b}-b} \right) \Delta_+^b u_{1t}$ around $b - \hat{b} = 0$, with $b - \hat{b} = O_p(T^{-1/2})$ and noting that the terms in the expansion behave as the derivatives of $\Delta_+^b u_{1t}$ with respect to b , cf. (104) in Lemma A.9 in Johansen and Nielsen (2012), whose sample moments are $O_p(T^{1/2-\epsilon})$ uniformly in b_1 , $\epsilon > 0$. The second term in (29) is $o_p(1)$ using a similar argument for $\left(\Delta_+^{\hat{b}} - 1 \right) \Delta X_t$, being approximately an $I(-b)$ asymptotically stationary process, and the superconsistency of $\hat{\beta}$. Finally, the analysis of (28) being $o_p(1)$ is simpler because it does not depend on b_1 and u_t is i.i.d.

To show the validity of the other correction alternatives in (15) and (16), it is only necessary to observe that the vector u_t is just a rotation of the vector ε_t , so all previous approximations and bounds can be used similarly. ■

Proof of Corollary 2. We have to additionally show that terms like

$$T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1, \hat{d}) V'_{0t}(\hat{d}) - T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1, d_0) V'_{0t}(d_0)$$

are $o_p(1)$ uniformly for $b_1 \in \mathcal{B}$ if $\hat{d} - d_0 = O_p(T^{-1/2})$. This follows now from a similar analysis as that of the first term in (29), writing this difference as

$$T^{-b_1} \sum_{t=1}^T V_{1t-1}(b_1, \hat{d}) \left\{ V'_{0t}(\hat{d}) - V'_{0t}(d_0) \right\} + T^{-b_1} \sum_{t=1}^T \left\{ V_{1t-1}(b_1, \hat{d}) - V_{1t-1}(b_1, d_0) \right\} V'_{0t}(d_0)$$

and using a Taylor expansion of $1 - \Delta_+^{\hat{d}-d_0}$ around $\hat{d} - d_0 = 0$ in $V'_{0t}(\hat{d}) - V'_{0t}(d_0) = \left(1 - \Delta_+^{\hat{d}-d_0} \right) \Delta^{d_0} \beta'_\perp X_t$ and $V_{1t-1}(b_1, \hat{d}) - V_{1t-1}(b_1, d_0) = \left(1 - \Delta_+^{-b_1} \right) \left(1 - \Delta_+^{\hat{d}-d_0} \right) \Delta^{d_0} \beta'_\perp X_t$,

and then using uniform bounds for the corresponding sample moments on (derivatives of) fractionally integrated processes. ■

Proof of Theorem 3. The proof follows the lines of the proof of Theorem 1, since the additional lags ΔX_{t-j} , $j = 1, \dots, k$ in regression (24) pose no additional problem compared to the projection of \hat{V}_{0t} and $\hat{V}_{1t-1}(b_1)$ on $(\Delta_+^{\hat{b}} - 1)\hat{u}_{1t}$, because the former are observed and $I(0)$. ■

References

- [1] Avarucci, M., Velasco C. (2009), A Wald Test for the Cointegration Rank in Nonstationary Fractional Systems, *Journal of Econometrics*, 151, 178-189.
- [2] Breitung, J., Hassler U. (2002), Inference on the Cointegration Rank in Fractionally Integrated Processes, *Journal of Econometrics*, 110, 167-185.
- [3] Chen, W., Hurvich, C. (2003), Semiparametric Estimation of Multivariate Fractional Cointegration, *Journal of American Statistical Association*, 98, 629-642.
- [4] Chen, W., Hurvich, C. (2006), Semiparametric Estimation of Fractional Cointegrating Subspaces, *Annals of Statistics*, 34, 2939-2979.
- [5] Davies, R. B. (1977), Hypothesis Testing when a Nuisance is Present Only under the Alternative, *Biometrika*, 64, 247-254.
- [6] Diebold, F. X., Inoue, A. (2001), Long memory and regime switching, *Journal of Econometrics*, 105, 131-159.
- [7] Dominguez, E., Novales, A. (2000), Testing the expectations hypothesis in Eurodeposits, *Journal of International Money and Finance*, 19, 713-736.
- [8] Doornik, J. A. (2009a), *An Introduction to OxMetrics 6*, Timberlake Consultants Press, London.
- [9] Doornik, J. A. (2009b), *An Object-Oriented Matrix Language Ox 6*, Timberlake Consultants Press, London.
- [10] Doornik, J. A., Ooms, M. (2007), *Introduction to Ox: An Object-Oriented Matrix Language*, Timberlake Consultants Press, London.
- [11] Engsted, T., Tangaard, C. (1994), Cointegration and the U.S. term structure, *Journal of Banking and Finance*, 18, 167-181.

- [12] Gil-Alaña, L. A. (2003), Testing of fractional cointegration in macroeconomic time series, *Oxford Bulletin of Economics and Statistics*, 65, 517-529.
- [13] Granger, C. W. J. (1986), Developments in the Study of Cointegrated Economic Variables, *Oxford Bulletin of Economics and Statistics*, 48, 213-28.
- [14] Granger, C. W. J., Hyung, N. (2004), Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance*, 11, 399-421.
- [15] Hall, A. D., Anderson, H. M., Granger, C. W. J. (1992), *The Review of Economics and Statistics*, 74, 116-126.
- [16] Hansen, B. E. (1996), Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, *Econometrica*, 64, 413-430.
- [17] Hualde, J., Velasco, C. (2008), Distribution-Free Tests Of Fractional Cointegration, *Econometric Theory*, 24 , 216-255.
- [18] Iacone, F. (2009), A Semiparametric Analysis of the Term Structure of the US Interest Rates, *Oxford Bulletin of Economics and statistics*, 71, 475-490.
- [19] Johansen, S. (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.
- [20] Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59, 1551-1580.
- [21] Johansen, S. (1995), *Likelihood-based inference in cointegrated Vector Auto-Regressive Models*. Oxford University Press.
- [22] Johansen, S. (2008), A representation theory for a class of vector autoregressive models for fractional processes, *Econometric Theory*, 24, 651-676.
- [23] Johansen, S. (2009), Representation of cointegrated autoregressive processes with application to fractional cointegration, *Econometric Reviews*, 28, 121-145.
- [24] Johansen, S., Nielsen, M. Ø. (2012), Likelihood inference for a fractionally cointegrated vector autoregressive model, *Econometrica*, 80, 2667-2732.
- [25] Łasak, K. (2008), Maximum likelihood estimation of fractionally cointegrated systems, *CREATES Research Paper 2008-53*.

- [26] Lasak, K. (2010), Likelihood based testing for no fractional cointegration, *Journal of Econometrics*, 158, 67-77.
- [27] Lyhagen, J. (1998), Maximum likelihood estimation of the multivariate fractional cointegrating model, *Working Paper Series in Economics and Finance*, 233, Stockholm School of Economics.
- [28] MacKinnon, J.G., Nielsen, M. Ø. (2013), Numerical distribution functions of fractional unit root and cointegration tests, *Journal of Applied Econometrics*, forthcoming.
- [29] Marinucci, D., Robinson, P. M. (2001), Semiparametric fractional cointegration analysis, *Journal of Econometrics*, 105, 225–247.
- [30] Marmol, F., Velasco, C. (2004), Consistent Testing of Cointegrating Relationships, *Econometrica*, 72, 1809-1844.
- [31] Nielsen, M. Ø. (2004), Optimal Residual-Based Tests for Fractional Cointegration and Exchange Rate Dynamics, *Journal of Business & Economic Statistics*, 22, 331-345.
- [32] Nielsen, M. Ø. (2005), Multivariate Lagrange Multiplier Tests for Fractional Integration, *Journal of Financial Econometrics*, 3, 372-398.
- [33] Nielsen, M. Ø. (2010), Nonparametric cointegration analysis of fractional systems with unknown integration orders, *Journal of Econometrics*, 155, 170-187.
- [34] 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.
- [35] Phillips P. C. B., Ouliaris, S. (1988), Testing for Cointegration Using Principal Component Measures, *Journal of Economic Dynamics and Control*, 12, 205-230.
- [36] Robinson, P. M (1995), Gaussian Semiparametric Estimation of Long Range Dependence, *Annals of Statistics*, 23, 1630-1661.
- [37] Robinson, P. M (2008), Diagnostic testing for cointegration. *Journal of econometrics*, 143, 206-225.
- [38] Robinson, P. M., Hualde, J. (2003), Cointegration in Fractional Systems with Unknown Integration Orders, *Econometrica*, 71, 1727-1766.
- [39] Robinson, P. M., Yajima, Y. (2002), Determination of cointegrating rank in fractional systems, *Journal of Econometrics*, 106, 217-241.

- [40] Sibbertsen, P., Kruse, R. (2009), Testing for a break in persistence under long-range dependencies, *Journal of Time Series Analysis*, 30, 263-285.
- [41] Soderlind, P., Svensson, L. (1997), New techniques to extract market expectations from financial instruments, *Journal of Monetary Economics*, 40, 383-429.
- [42] Svensson, L. E. O. (1997), Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets, *European Economic Review*, 41, 1111-1146.
- [43] Taylor, J. B. (1993), Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

Research Papers 2013



- 2012-49: Peter Christoffersen, Christian Dorion , Kris Jacobs and Lotfi Karoui: Nonlinear Kalman Filtering in Affine Term Structure Models
- 2012-50: Peter Christoffersen, Kris Jacobs and Chayawat Ornthanalai: GARCH Option Valuation: Theory and Evidence
- 2012-51: Tim Bollerslev, Lai Xu and Hao Zhou: Stock Return and Cash Flow Predictability: The Role of Volatility Risk
- 2012-52: José Manuel Corcuera, Emil Hedeveg, Mikko S. Pakkanen and Mark Podolskij: Asymptotic theory for Brownian semi-stationary processes with application to turbulence
- 2012-53: Rasmus Søndergaard Pedersen and Anders Rahbek: Multivariate Variance Targeting in the BEKK-GARCH Model
- 2012-54: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperature Trends and Co-Shifting: A Shifting Mean Vector Autoregressive Analysis
- 2012-55: Daniel J. Nordman, Helle Bunzel and Soumendra N. Lahiri: A Non-standard Empirical Likelihood for Time Series
- 2012-56: Robert F. Engle, Martin Klint Hansen and Asger Lunde: And Now, The Rest of the News: Volatility and Firm Specific News Arrival
- 2012-57: Jean Jacod and Mark Podolskij: A test for the rank of the volatility process: the random perturbation approach
- 2012-58: Tom Engsted and Thomas Q. Pedersen: Predicting returns and rent growth in the housing market using the rent-to-price ratio: Evidence from the OECD countries
- 2013-01: Mikko S. Pakkanen: Limit theorems for power variations of ambit fields driven by white noise
- 2013-02: Almut E. D. Veraart and Luitgard A. M. Veraart: Risk premia in energy markets
- 2013-03: Stefano Grassi and Paolo Santucci de Magistris: It's all about volatility (of volatility): evidence from a two-factor stochastic volatility model
- 2013-04: Tom Engsted and Thomas Q. Pedersen: Housing market volatility in the OECD area: Evidence from VAR based return decompositions
- 2013-05: Søren Johansen and Bent Nielsen: Asymptotic analysis of the Forward Search
- 2013-06: Debopam Bhattacharya, Pascaline Dupasand Shin Kanaya: Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets
- 2013-07: Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns
- 2013-08: Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation