

Semiparametric inference in multivariate fractionally cointegrated systems[☆]J. Hualde^{a,*}, P.M. Robinson^b^a Universidad Pública de Navarra, Spain^b London School of Economics, United Kingdom

ARTICLE INFO

Article history:

Received 26 October 2006

Received in revised form

21 January 2009

Accepted 2 April 2010

Available online 13 May 2010

JEL classification:

C32

Keywords:

Fractional cointegration

Semiparametric model

Unknown integration orders

Standard inference

ABSTRACT

A semiparametric multivariate fractionally cointegrated system is considered, integration orders possibly being unknown and $I(0)$ unobservable inputs having nonparametric spectral density. Two estimates of the vector of cointegrating parameters ν are considered. One involves inverse spectral weighting and the other is unweighted but uses a spectral estimate at frequency zero. Both corresponding Wald statistics for testing linear restrictions on ν are shown to have a standard null χ^2 limit distribution under quite general conditions. Notably, this outcome is irrespective of whether cointegrating relations are “strong” (when the difference between integration orders of observables and cointegrating errors exceeds $1/2$), or “weak” (when that difference is less than $1/2$), or when both cases are involved. Finite-sample properties are examined in a Monte Carlo study and an empirical example is presented.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Semiparametric modelling has become popular in cointegration analysis of $I(1)$ time series with $I(0)$ cointegrating errors. In the simplest parametric setting, observables follow a random walk and cointegrating errors are serially uncorrelated. Vector autoregressive (VAR) extensions have been developed (e.g. Johansen, 1991), but optimal inference on the unknown cointegrating relations loses validity if the VAR order is under-specified, or if the process lies outside the VAR class. Phillips and Hansen (1990) and Phillips (1991a) and others showed that one can do as well allowing the $I(0)$ inputs to have nonparametric autocorrelation, under suitable conditions on the bandwidth employed in the smoothed nonparametric spectrum estimate.

Another source of possible misspecification is the basic $I(1)/I(0)$ framework itself. Recently, optimal inference has developed in a fractional setting (see e.g. Jeganathan, 1999; Robinson and Hualde, 2003). Here, integration orders were allowed to be unknown, to non-trivially generalize the $I(1)/I(0)$ assumption, but theory was developed only in a fully parametric setting, incurring the familiar concern about misspecification, and just for a bivariate situation, hence avoiding the complexity in simultaneously

dealing with multiple cointegrating relations where integration orders of errors may differ. Dolado and Marmol (1996) and Kim and Phillips (2000) allowed nonparametric autocorrelation in $I(0)$ inputs but, respectively assuming knowledge of integration orders, and proposing sub-optimal procedures.

The present paper develops inference on cointegrating relations in a semiparametric fractional setting, with unknown integration orders. To describe our model, we introduce the following definitions corresponding to ones in Robinson and Hualde (2003) (hereafter RH). For any scalar or vector sequence v_t , $t = 0, \pm 1, \dots$, we denote

$$v_t^\# = v_t 1(t > 0),$$

where $1(\cdot)$ is the indicator function. Defining the difference operator $\Delta = 1 - L$, where L is the lag operator, the fractional difference operator is given formally, for any real α , $\alpha \neq -1, -2, \dots$, by

$$\Delta^{-\alpha} = \sum_{j=0}^{\infty} a_j(\alpha) L^j, \quad a_j(\alpha) = \frac{\Gamma(j + \alpha)}{\Gamma(\alpha)\Gamma(j + 1)},$$

with Γ denoting the gamma function. Denoting by a_{it} the i th component of an arbitrary vector process a_t , we say that a scalar process ζ_t is integrated of order d , $\zeta_t \sim I(d)$, if for any $l \times 1$ ($l < \infty$) covariance stationary process ξ_t whose spectral density matrix is continuous and nonsingular at all frequencies,

$$\zeta_t = \sum_{k=1}^l \Delta^{-d_k} \xi_{kt}^\# \quad (1)$$

[☆] This paper is an extensive development of Hualde and Robinson (2006). We are grateful for the comments of two referees, which have led to considerable improvements.

* Corresponding author. Tel.: +34 948169674; fax: +34 948169721.
E-mail address: javier.hualde@unavarra.es (J. Hualde).

for $d = \max_{1 \leq k \leq l} d_k$. Robinson and Gerolimetto (2006) refer to each summand of (1) as “basic fractional”, so our $I(d)$ process, like theirs, is a linear combination of “basic fractional” processes, with maximal order d . We say that a vector process is $I(d)$ if at least one of its components is $I(d)$, the rest having integration orders no greater than d . This definition is identical to that of Hualde (2008) and resembles that of Johansen (2008), which allows $I(d)$ vectors to have individual components of smaller order than d .

For any $r_0 \times 1$ vector $d = (d_1, \dots, d_{r_0})'$, the prime denoting transposition, denote $\Delta(d) = \text{diag} \{ \Delta^{d_1}, \dots, \Delta^{d_{r_0}} \}$. Let $u_t, t = 0, \pm 1, \dots$, be an $r_0 \times 1$ covariance stationary unobservable process with zero mean and nonparametric spectral density matrix $f(\lambda)$, given by

$$E(u_0 u_t') = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda,$$

that is at least continuous and nonsingular at all frequencies. For a $r_0 \times r_0$ nonsingular matrix Υ and a $r_0 \times 1$ possibly unknown vector $\delta = (\delta_1, \dots, \delta_{r_0})'$, we define the $r_0 \times 1$ vector observable process $z_t, t = 0, \pm 1$, by

$$\Upsilon z_t = \Delta^{-1}(\delta) u_t^\# \tag{2}$$

where, without loss of generality we set

$$0 \leq \delta_1 \leq \delta_2 \leq \dots \leq \delta_{r_0}, \tag{3}$$

which implies $z_t \sim I(\delta_{r_0})$. We exclude antipersistent processes (having negative integration order), which seem to have limited economic relevance, except in an over-differencing context.

The system (2) is the general framework in which we will discuss the cointegrating properties of z_t . We say that a vector $\xi_t \sim I(d)$ is cointegrated if there exists a vector $\alpha \neq 0$ such that $\alpha' \xi_t \sim I(c)$ with $c < d$. This definition covers the standard notion of cointegration where the observables share the same integration order (as in, e.g., Engle and Granger, 1987), but also many others, where observables with different integration orders might combine in the cointegrating relation. The definition is identical to that in Hualde (2008), which is similar to that of Johansen (1996), and more general than those in Flores and Szafarz (1996) and Robinson and Yajima (2002) or Robinson and Marinucci (2003).

The system (2) is very general, but unless δ and Υ are further restricted, it does not necessarily represent a cointegrated model and moreover, is nonidentifiable. As will be seen in Section 2, additional restrictions on δ and Υ are motivated by the particular cointegrating structure which characterizes our observables. Apart from other possible constraints, we will assume in all cases that Υ is upper-triangular with 1's in the main diagonal. We justify below that there is no loss of generality in this restriction, especially in connection to (3). As a very relevant improvement over existing literature we do not assume that all observables share the same integration order (unlike, e.g., Kim and Phillips, 2000; Chen and Hurvich, 2003, 2006; and effectively also Robinson and Yajima, 2002), although this possibility is considered and materializes if the last column of Υ^{-1} contains no zeroes. In addition, we allow the orders of the cointegrating errors to possibly vary, unlike in the standard cointegration literature where all these orders are assumed zero. We will deal with the case of an arbitrary cointegrating rank (that is the number of linearly independent cointegrating vectors) $r_1 \in \{1, \dots, r_0 - 1\}$, but our results are new even if $r_1 = 1$, for any $r_0 \geq 2$. Given there are no prior restrictions on f , those which will be imposed on Υ and δ ensure identification, and imply that (given consistent estimates of Υ and δ) consistent estimation of f is possible, which is fundamental to our approach.

The truncation in (2) is motivated by systems where at least δ_{r_0} falls in the nonstationary region, $\delta_{r_0} > 1/2$. This version of fractional integration (“Type II” process) and cointegration accords

with that in RH. An alternative one (“Type I” process), for which the procedures developed below nevertheless apply, was used by Dolado and Marmol (1996), Jeganathan (1999) and Kim and Phillips (2000). None of these references covers δ_{r_0} within the stationary region, $\delta_{r_0} \in (0, 1/2)$, which will be permitted by our setting; in this case we say that our relations display “stationary cointegration”. This arose in Robinson (1994a), and it has been stressed in a finance context by Bandi and Perron (2004) and Christensen and Nielsen (2006). A larger class (where $\delta_{r_0} > 1/2$ is possible) consists of cases where the cointegrating gap (the difference between the integration order of the observables and cointegrating error) falls in the $(0, 1/2)$ region, which we denote “weak cointegration”. Empirical evidence of this, with nonstationary observables, was found by Robinson and Marinucci (2003), and it has been further discussed by Hualde and Robinson (2007). The case where the gap is greater than $1/2$, which includes the usual $I(1)/I(0)$ situation, is called “strong cointegration”.

It is desired to conduct inference on the unknown elements of Υ , in the presence of unknown δ . The present paper does not merely extend nontrivially the bivariate model in RH to a richer multivariate framework, and allow also for nonparametric f , but simultaneously covers relations of weak and strong cointegration, which, as we understand, has not been attempted before. While asymptotics for point estimates of unknown parameters in Υ differ significantly across these cases, the same rules of inference prevail throughout, with the same Wald test statistic (for a linear hypothesis on these parameters) having a null limit χ^2 distribution. The borderline situation between strong and weak cointegration, with a cointegrating gap of $1/2$, will be excluded largely because it seems too special to warrant the space necessary to present the somewhat separate technical treatment that it would require. However, while the convergence rate of our estimates differs from those under both strong and weak cointegration, it seems that the same limit distribution for the Wald statistic will still hold, so that slight limitation of our analysis can be dispensed with.

We find it convenient to treat the nonparametric autocorrelation in the frequency domain. This prompts consideration of two alternative methods of estimating and testing hypotheses on ν . One involves a ratio of weighted periodogram averages either across all frequencies in the Nyquist band, or only over those within a shrinking neighbourhood of zero frequency. The weighting is inverse with respect to smoothed estimates of f . Because of the concentration of spectral mass around zero frequency, where f changes little, computationally simpler statistics, with the same asymptotic properties, replace the weights by multiplicative factors based on an estimate of $f(0)$.

The plan of the paper is as follows. In Section 2 additional restrictions on δ and Υ and estimates of the cointegrating matrix and test statistics will be introduced. Regularity conditions and asymptotic properties are presented in Section 3. Section 4 contains a Monte Carlo study of finite-sample behaviour, and Section 5, the analysis of an empirical example. Some concluding remarks are made in Section 6. Proofs are relegated to an Appendix.

2. Estimation of cointegrating parameters and test statistics

As previously mentioned, we need to introduce additional restrictions on Υ, δ , which ensure cointegration and identification, and then propose estimates of the unrestricted parameters in Υ . Our basic assumption (which materializes in Assumption 1) is that the cointegrating properties of z_t are characterized by the following structure. First, $S_{r_1}^{(1)} \subset \mathbb{R}^{r_0}$ represents the cointegrating space of dimension $r_1 < r_0$. This implies the existence of a full rank $r_0 \times r_1$ matrix $\beta(1)$ (whose columns are cointegrating vectors) such that $\beta(1)' z_t \sim I(\delta_{r_1})$, with $\delta_{r_1} < \delta_{r_0}$. Hualde (2008) shows that δ_{r_1} is uniquely identified, in the sense that for any other arbitrary $r_0 \times$

r_1 matrix whose columns are linearly independent cointegrating vectors, say $\bar{\beta}(1)$, $\bar{\beta}(1)'z_t \sim I(\delta_{r_1})$.

In addition, we also assume that $S_{r_1}^{(1)}$ has a subspace $S_{r_1}^{(2)} \subset S_{r_1}^{(1)}$ of dimension $r_2 < r_1$, such that there exists a $r_0 \times r_2$ matrix $\beta(2)$ (whose columns are linear combinations of those of $\beta(1)$) so $S_{r_2}^{(2)} = sp(\beta(2))$ (where $sp(\cdot)$ denotes column space of a matrix) and $\beta(2)'z_t \sim I(\delta_{r_2})$, with $\delta_{r_2} < \delta_{r_1}$ (again Hualde, 2008, ensures the uniqueness of δ_{r_2} in the previous sense). Thus, the column vectors in $\beta(2)$ represent particular directions in the cointegrating space which achieve a greater reduction in the integration orders of the observables (from δ_{r_0} to δ_{r_2}) instead of the “typical” reduction (from δ_{r_0} to δ_{r_1}) achieved by most cointegrating vectors. Proceeding in this fashion, we assume that the whole cointegrating structure of z_t is characterized by subspaces $S_{r_s}^{(s)} \subset S_{r_{s-1}}^{(s-1)} \subset \dots \subset S_{r_1}^{(1)}$, of dimensions r_s, r_{s-1}, \dots, r_1 , respectively, where $1 \leq r_s < r_{s-1} < \dots < r_1 < r_0$, and letting $\beta(j)$ be $r_0 \times r_j$ matrices such that $S_{r_j}^{(j)} = sp(\beta(j))$, $j = 1, \dots, s$, $\beta(j)'z_t \sim I(\delta_{r_j})$, where $\delta_{r_s} < \delta_{r_{s-1}} < \dots < \delta_{r_1} < \delta_{r_0}$.

Two issues appear to be relevant here. First, Hualde (2008) presents a procedure to infer the dimension of all possible cointegrating subspaces, and, furthermore, proposes consistent estimates of these subspaces. Second, one of the implications of Hualde’s (2008) results is that under such a cointegrating structure for z_t , defining for any p -dimensional vector ξ_t , $a < b < p$, $\xi_t^{(a,b)} = (\xi_{at}, \dots, \xi_{bt})'$, $\xi_{-b,t} = (\xi_{b+1,t}, \dots, \xi_{pt})'$, there exist $(r_j - r_{j+1}) \times (r_0 - r_j)$ matrices $D_{(j)}$, $j = 1, \dots, s$, (setting throughout $r_{s+1} = 0$), such that the individual observables in z_t can be ordered in such a way that

$$\begin{aligned} z_t^{(1,r_s)} + D_{(s)}z_{-r_s,t} &= w_t^{(1,r_s)}, \\ z_t^{(r_s+1,r_{s-1})} + D_{(s-1)}z_{-r_{s-1},t} &= w_t^{(r_s+1,r_{s-1})}, \\ &\vdots \\ z_t^{(r_3+1,r_2)} + D_{(2)}z_{-r_2,t} &= w_t^{(r_3+1,r_2)}, \\ z_t^{(r_2+1,r_1)} + D_{(1)}z_{-r_1,t} &= w_t^{(r_2+1,r_1)}, \\ z_{-r_1,t} &= w_t^{(r_1+1,r_0)}, \end{aligned} \tag{4}$$

where the components of $w_t^{(r_{j+1}+1,r_j)}$ are individually $I(\delta_{r_j})$ and not cointegrated. Also, the procedure in Hualde (2008) specifies which variables should appear on the left and right hand sides of the different equations of system (4). The main implication of (4) is that, setting $w_t = \Delta^{-1}(\delta)u_t^\#$, (2) holds with

$$\mathcal{Y} = \begin{pmatrix} I_{r_s} & & & D_{(s)} \\ 0_{r_{s-1}-r_s,r_s} & I_{r_{s-1}-r_s} & & D_{(s-1)} \\ 0_{r_{s-2}-r_{s-1},r_s} & 0_{r_{s-2}-r_{s-1},r_{s-1}-r_s} & \ddots & \vdots \\ \vdots & \vdots & \vdots & I_{r_1-r_2} & D_{(1)} \\ 0_{r_0-r_1,r_s} & 0_{r_0-r_1,r_{s-1}-r_s} & \dots & 0_{r_0-r_1,r_1-r_2} & I_{r_0-r_1} \end{pmatrix}, \tag{5}$$

where I_s is the s -rowed identity matrix, and $0_{i,j}$ the $i \times j$ matrix of zeroes. We formalize the restrictions on \mathcal{Y} implied by (5) and the equalities and strict inequalities among the elements of δ implied by the previous discussion in Assumption 1.

Assumption 1. The relations (2) hold for a possibly unknown vector δ whose components satisfy

$$0 \leq \delta_1 = \dots = \delta_{r_s} < \delta_{r_s+1} = \dots = \delta_{r_{s-1}} < \dots < \delta_{r_2+1} = \dots = \delta_{r_1} < \delta_{r_1+1} = \dots = \delta_{r_0}, \tag{6}$$

$$\delta_{r_i} - \delta_{r_j} \neq 1/2, \quad i < j, j \in \{1, \dots, s\},$$

and for

$$\mathcal{Y} = \begin{pmatrix} \mathcal{Y}_1 & \mathcal{Y}_2 \\ 0_{r_0-r_1,r_1} & I_{r_0-r_1} \end{pmatrix},$$

where \mathcal{Y}_1 is a $r_1 \times r_1$ matrix consisting of $s \times s$ blocks $\mathcal{Y}_1^{(i,j)}$ of dimensions $(r_{s-i+1} - r_{s-i+2}) \times (r_{s-j+1} - r_{s-j+2})$, $i, j = 1, \dots, s$, such that $\mathcal{Y}_1^{(i,i)} = 0$ if $i > j$, $\mathcal{Y}_1^{(i,i)} = I_{r_{s-i+1}-r_{s-i+2}}$, but \mathcal{Y}_1 , like \mathcal{Y}_2 (which is an $r_1 \times (r_0 - r_1)$ matrix), are otherwise unknown and unrestricted.

Note that when

$$\delta_1 = \delta_2 = \dots = \delta_{r_1} = 0, \quad \delta_{r_1+1} = \delta_{r_1+2} = \dots = \delta_{r_0} = 1, \tag{7}$$

(2) is the usual $I(1)/I(0)$ cointegrated system with cointegrating rank r_1 .

Denote by v_k the vector of unrestricted parameters in the k th row of \mathcal{Y} , for $k \in \{1, \dots, r_1\}$, so v_k is a $(r_0 - r_j) \times 1$ vector if $k \in \{r_{j+1} + 1, \dots, r_j\}$, $j = 1, \dots, s$. Hence form the unknown elements of \mathcal{Y} into the $q \times 1$ vector $v = (v'_1, \dots, v'_{r_1})'$, where $q = \sum_{k=1}^s (r_0 - r_k)(r_k - r_{k+1})$. We equivalently write

$$vec \mathcal{Y} = Cv - c, \tag{8}$$

where c is a $r_0^2 \times 1$ vector of zeroes except for -1 in positions $1, r_0 + 2, 2r_0 + 3, 3r_0 + 4, \dots, r_0^2$ and C is the $r_0^2 \times q$ matrix $C = (C_1, \dots, C_{r_1})$, where for $k \in \{r_{j+1} + 1, \dots, r_j\}$, $j = 1, \dots, s$,

$$C_k = (\bar{i}_{j r_0+k}, \bar{i}_{(j+1)r_0+k}, \dots, \bar{i}_{(r_0-1)r_0+k}),$$

with \bar{i}_j a $r_0^2 \times 1$ vector of zeroes except for 1 in the j th component. Define the $r_0^2 \times r_0$ matrix-valued process z_t^d , whose i th $r_0 \times r_0$ submatrix is diagonal with (j, j) th element $\Delta^{d_j} z_{it}$. For generic, possibly matrix-valued, sequences ξ_t, χ_t , define the discrete Fourier transform

$$w_\xi(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_{t=1}^n \xi_t e^{it\lambda},$$

and the cross-periodogram and periodogram

$$I_{\xi\chi}(\lambda) = w_\xi(\lambda)w_\chi(-\lambda)', \quad I_\xi(\lambda) = I_{\xi\xi}(\lambda),$$

for real λ . Denote by $\lambda_j = 2\pi j/n$, $j = 0, \dots, [n/2]$, the Fourier frequencies, where $[\cdot]$ means integer part.

Given observations z_t , $t = 1, \dots, n$, and a nonsingular estimate $\hat{f}(\lambda)$ of $f(\lambda)$, based on these data, define the statistics

$$\begin{aligned} \hat{a}_m(d) &= C' \sum_{j=0}^m s_j \text{Re} \{ w_{z^d}(-\lambda_j) \hat{f}^{-1}(\lambda_j) w'_{z^d}(\lambda_j) \} c, \\ \hat{b}_m(d) &= C' \sum_{j=0}^m s_j \text{Re} \{ w_{z^d}(-\lambda_j) \hat{f}^{-1}(\lambda_j) w'_{z^d}(\lambda_j) \} C, \\ \hat{a}_m^\circ(d) &= C' \sum_{j=0}^m s_j \text{Re} \{ w_{z^d}(-\lambda_j) \hat{f}^{-1}(0) w'_{z^d}(\lambda_j) \} c, \\ \hat{b}_m^\circ(d) &= C' \sum_{j=0}^m s_j \text{Re} \{ w_{z^d}(-\lambda_j) \hat{f}^{-1}(0) w'_{z^d}(\lambda_j) \} C, \end{aligned} \tag{9}$$

for an integer m such that

$$m \rightarrow \infty \quad \text{as } n \rightarrow \infty, \quad 1 \leq m \leq n/2, \tag{10}$$

and for $s_j = 1, j = 0, n/2, s_j = 2$, otherwise.

Defining

$$\hat{v}_m(d) = \hat{b}_m^{-1}(d) \hat{a}_m(d), \quad \hat{v}_m^\circ(d) = \hat{b}_m^\circ(d)^{-1} \hat{a}_m^\circ(d),$$

we consider the two sets of estimates of v ,

$$\begin{aligned} W \text{ (“weighted”) } &: \hat{v}_m(\delta), \hat{v}_m(\hat{\delta}); \\ Z \text{ (“zero-frequency”) } &: \hat{v}_m^\circ(\delta), \hat{v}_m^\circ(\hat{\delta}), \end{aligned} \tag{11}$$

where $\hat{\delta}$ is an estimate of δ . The estimates (11) are of generalized least squares type, and effectively “whiten” the data, as their

desirable asymptotic properties will confirm. The estimates $\widehat{v}_m(\delta)$, $\widehat{v}_m^\circ(\delta)$, treat δ as known (as is usual in the $I(1)/I(0)$ case), but in our context they are generally regarded as infeasible, and included in part for completeness and to demonstrate that estimation of δ makes no asymptotic difference, as well as to imply, with $\widehat{v}_m(\widehat{\delta})$, $\widehat{v}_m^\circ(\widehat{\delta})$, cases where δ is partly known (e.g. $\delta_{r_0} = 1$ is known but $\delta_1, \dots, \delta_{r_1}$ are unknown).

When $m = [n/2]$, (11) are semiparametric multivariate counterparts of the parametric estimates in (2.18) and (3.13) of RH, because the real operators and s_j can be dropped and summations over $[0, [n/2]]$ replaced by ones over $[1, n]$, due to symmetry properties. As noted there, the computational simplicity of the Z estimates over the W estimates is due not only to having to estimate f at only frequency zero, but to

$$\begin{aligned} \widehat{a}_{[n/2]}^\circ(d) &= \frac{1}{2\pi} C' \sum_{t=1}^n z_t^d \widehat{f}^{-1}(0)(z_t^d)' c, \widehat{b}_{[n/2]}^\circ(d) \\ &= \frac{1}{2\pi} C' \sum_{t=1}^n z_t^d \widehat{f}^{-1}(0)(z_t^d)' C. \end{aligned}$$

However, RH found, in their parametric setting with strong cointegration, that “zero-frequency” estimates only do as well as “weighted” ones when the cointegrating gap is larger than 1; when it is 1, a “second-order bias” appears, and when it is less than 1 (but greater than 1/2) convergence is slower due to sub-optimal weighting, and in each case the mixed-normal asymptotics underlying the desirable null χ^2 limit distribution of Wald test statistics is lost. As seen in the following section, limiting the increase of m appropriately repairs this defect (see (26)). In case of weak cointegration, we also limit the increase of m , which, as in the previous case, corrects for simultaneity bias due to correlation between regressors and cointegrating errors. This idea was already present in Robinson (1994a), who found that $m/n \rightarrow 0$ is necessary for consistency of the narrow band least squares estimate in case of stationary cointegration, whereas Robinson and Marinucci (2001, 2003) found that $m/n \rightarrow 0$ reduces the bias of this estimate in nonstationary cointegration situations where the collective memory of regressor and cointegrating error is smaller than 1. Earlier, Phillips (1991a) considered similar estimates to $\widehat{v}_m(\delta)$, $\widehat{v}_m^\circ(\delta)$ in the standard $I(1)/I(0)$ situation (7) with δ assumed known and with additional conditions imposed on Υ , namely $\Upsilon_1 = I_{r_1}$, which are natural restrictions if $S_{r_1}^{(1)}$ describes the whole cointegrating structure, but not in our more complex and general setting.

Alternatively, by elementary row operations on the $r_1 \times r_0$ sub-matrix (Υ_1, Υ_2) , we can transform our Assumption 1 set-up with upper-triangular Υ_1 to that with $\Upsilon_1 = I_{r_1}$, but this would in general reduce the “achievement” of a cointegration analysis, since the cointegrating relations would in general each have integration order $\max_{i \in \{1, \dots, r_1\}} \delta_i$. In particular, this transformation would lead to estimation of the cointegrating space $S_{r_1}^{(1)}$, but all information contained in the different cointegrating subspaces allowed by Assumption 1 would be, in general, lost. Thus, while this type of transformation seems innocuous in the traditional cointegrating framework (where $\max_{i \in \{1, \dots, r_1\}} \delta_i = 0$), in our present fractional setting it implies a very strong simplification.

There have been numerous attempts to validate economic hypotheses on long-run relationships among economic variables by testing the parameters of potentially cointegrated models, in particular, whether these satisfy certain linear restrictions. Without being exhaustive, we briefly describe three examples in distinct economic environments. First, Johansen and Juselius (1992) (see also MacDonald and Marsh, 1997) studied jointly the purchasing power parity (PPP) and uncovering interest rate hypotheses by examining relations among five observables (domestic and foreign

prices, exchange rate, domestic and foreign interest rates), testing for example whether the PPP holds in every cointegrating relation, and also whether the presence of the interest rates in the cointegrating relations is simply constrained to their differential. Second, Johansen and Juselius (1994), in an ISLM multivariate model with five observables (money, income, price, three-month commercial bill and ten-year bond rates) tested restrictions that cointegrating relations are consistent with a money demand equation together with interest rate differential and the deviation of the bond rate from an inflation rate measure. Finally, in a different setting, Jacobson et al. (1998) investigated, in a multivariate cointegrated model with four observables (real output, employment, labour force, real wage), the behaviour of real wages and unemployment, testing (by examining restrictions among cointegrating parameters) the joint hypothesis that unemployment and labour’s share of value added are $I(0)$.

Thus consider the null hypothesis

$$H_0 : Rv = v, \tag{12}$$

where v is a given $J \times 1$ vector and R is a given $J \times q$ matrix. For $j = 1, \dots, s$ and $k \in \{r_{j+1} + 1, r_j\}$, partition $v_k = (v_k^{(j)'}, \dots, v_k^{(1)'})'$, where $v_k^{(l)}$ are the parameters in v_k corresponding to regressors $z_t^{(r_{j+1}, r_{j-1})}$, $l = 1, \dots, j$. Then R must satisfy the following condition:

Assumption 2. R has full row rank and, partitioning it column-wise corresponding to the partition of v into the $v_k^{(l)}$, $k \in \{r_{j+1} + 1, r_j\}$, $l = 1, \dots, j$, $j = 1, \dots, s$, in each row of R the only non-null subvectors correspond to a unique choice of $j = 1, \dots, s$ and $l = 1, \dots, j$.

This assumption avoids commutativity problems due to the different convergence rates of the estimates of particular linear combinations of the v_k . As it stands, it permits linear restrictions among coefficients corresponding to regressors $z_t^{(r_{j+1}, r_{j-1})}$, $j = 1, \dots, s$, within each block of equations in (4), and also covers exclusion restrictions.

The W and Z Wald statistics on which the test of (12) is based are given by

$$W_m(d) = (R\widehat{v}_m(d) - v)' (R\widehat{b}_m^{-1}(d)R')^{-1} (R\widehat{v}_m(d) - v), \tag{13}$$

$$W_m^\circ(d) = (R\widehat{v}_m^\circ(d) - v)' (R\widehat{b}_m^\circ(d)^{-1}R')^{-1} (R\widehat{v}_m^\circ(d) - v), \tag{14}$$

with $d = \delta$ or $\widehat{\delta}$.

3. Main results

We introduce further regularity conditions. For the W estimates we require:

Assumption 3. The process u_t , $t = 0, \pm 1, \dots$, has representation

$$u_t = A(L)\varepsilon_t, \quad A(z) = I_{r_0} + \sum_{j=1}^{\infty} A_j z^j, \tag{15}$$

where the A_j are $r_0 \times r_0$ matrices such that

$$\det\{A(z)\} \neq 0, \quad |z| = 1, \tag{16}$$

$A(e^{i\lambda})$ is differentiable in $\lambda \in [-\pi, \pi]$ with derivative in $\text{Lip}(\eta)$, $\eta > 1/2$, and with $\|\cdot\|$ denoting the Euclidean norm, the $r_0 \times 1$ vector variables ε_t are independent and identically distributed vectors with mean zero, positive definite covariance matrix Ω , $E\|\varepsilon_t\|^\mu < \infty$, $\mu \geq 4$, and if $\delta_{r_i} - \delta_{r_j} > 1/2$ for some $i < j$, $j \in \{1, \dots, s\}$, $\mu > 2/(2 \min_{i,j:i < j, \delta_{r_i} - \delta_{r_j} > 1/2} (\delta_{r_i} - \delta_{r_j}) - 1)$.

This is a multivariate extension of Assumption 1 of RH and is easily satisfied if u_t is a stationary and invertible ARMA, imposing a global smoothness condition on $f(\lambda)$ which implies that

$$\sum_{j=1}^{\infty} j \|A_j\| < \infty, \quad \sum_{j=-\infty}^{\infty} |j| \|\Gamma(j)\| < \infty,$$

where $\Gamma(j) = E(u_t u_{t-j}')$. It is imposed even under $m/n \rightarrow 0$ because it enables the use of the functional limit theorem of Marinucci and Robinson (2000). For the Z estimates we can slightly relax it to:

Assumption 3°. Assumption 3 holds with the condition $\det\{A(1)\} \neq 0$, replacing (16).

Throughout we denote by K a generic, positive, finite constant. Both W and Z estimates use:

Assumption 4.

$$\|\widehat{\delta}\| \leq K, \tag{17}$$

and there exists $\kappa > 0$ such that

$$\widehat{\delta} = \delta + O_p(n^{-\kappa}), \tag{18}$$

and, as $n \rightarrow \infty$,

$$n^{-\kappa} m^{1-\max\{\min(\min_{i,j:i < j}(\delta_{r_i} - \delta_{r_j}), 1), 1/2\}} \log m \rightarrow 0. \tag{19}$$

On \widehat{f} we impose the following two assumptions, for the W and Z estimates respectively.

Assumption 5. Uniformly in $j = 0, 1, \dots, n - 1$, there exist $\kappa > 0, \phi > 0$, such that

$$\widehat{f}(\lambda_j) - f(\lambda_j) = O_p(n^{-\kappa}), \tag{20}$$

$$\widehat{f}(\lambda_{j+1}) - f(\lambda_{j+1}) - (\widehat{f}(\lambda_j) - f(\lambda_j)) = O_p(n^{-\phi}), \tag{21}$$

and, as $n \rightarrow \infty$,

$$n^{-\kappa} m^{1-\max\{\min(\min_{i,j:i < j}(\delta_{r_i} - \delta_{r_j}), 1), 1/2\}} \rightarrow 0, \tag{22}$$

$$n^{-\phi} m^{2-\max\{\min(\min_{i,j:i < j}(\delta_{r_i} - \delta_{r_j}), 1), 1/2\}} \rightarrow 0. \tag{23}$$

Assumption 5°. There exists $\kappa > 0$ such that, as $n \rightarrow \infty$,

$$\widehat{f}(0) - f(0) = O_p(n^{-\kappa}), \tag{24}$$

for which (22) is satisfied.

Finally the following assumptions are additionally imposed on the bandwidth m in the case of the W and Z estimates respectively.

Assumption 6. For all $i < j, j \in \{1, \dots, s\}$ such that $\delta_{r_i} - \delta_{r_j} < 1/2$, for η in Assumption 3,

$$m^{\delta_{r_i} - \delta_{r_j} - 1/2} \log^{1/2} n + m^{3+2\eta} / n^{2+2\eta} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \tag{25}$$

Assumption 6°. Assumption 6 holds, and in addition, for all $i < j, j \in \{1, \dots, s\}$, such that $\delta_{r_i} - \delta_{r_j} > 1/2$,

$$m/n^{\delta_{r_i} - \delta_{r_j}} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \tag{26}$$

Part (17) of Assumption 4 is again taken from RH, where it is discussed. Primitive conditions for (18), (19) and for Assumptions 5 and 5° are presented by Hualde and Robinson (2006), and further

relevant discussions can be found in Chen and Hurvich (2006) and Robinson (2005), but we briefly discuss how they might be satisfied, and how one might proceed in practice.

Regarding the integration orders, in view of (6) we need to estimate $\delta_{r_j}, j = 0, \dots, s$. Denote by $\widehat{\delta}_{r_j}$ the estimate of δ_{r_j} . We first estimate δ_{r_0} , and as this is the common integration order of the individual components of the $(r_0 - r_1) \times 1$ vector $z_{-r_1,t}$, we would likely employ a common estimate. We can obtain such an estimate by univariate semiparametric techniques such as versions of log periodogram regression or local Whittle, applied to only a single element of $z_{-r_1,t}$, or obtain one of these estimates for each element and then estimate δ_{r_0} by a linear combination of these, such as the arithmetic mean. In order to allow for non-stationarity, tapering can be employed, as proposed by Velasco (1999a,b)—he assumed Type I non-stationarity, but according to Robinson (2005) this does not affect convergence rates, at least. An alternative approach is due to Shimotsu and Phillips (2005). All these estimates have nonparametric convergence rates that are consistent with (18). It is possible to impose the equality restriction in a more formal way, to obtain a more efficient estimate of δ_{r_0} than any linear combination of univariate estimates. Robinson (1995a) proposed log-periodogram estimates for a vector process in which integration orders can satisfy linear restrictions, and the same can be done with Lobato's (1999) multivariate local Whittle estimate. They only justified their estimates under stationarity, but non-stationarity can doubtless be covered, as in the scalar case. More important, they assumed their vector observables were not cointegrated, as in our present situation, where the components of $z_{-r_1,t}$ are not cointegrated.

The estimation of $\delta_{r_j}, j = 1, \dots, s$, is necessarily based on residuals. Noting (4), denote by $\overline{D}_{(j)}$ the narrow band least squares (NBLS) estimate (see Robinson, 1994a) of $D_{(j)}, j = 1, \dots, s$, and for two nonnegative real numbers a, b , such that $a > b$, define $g_n(a, b) = n^{a-b} m^{b-a}$, if $a < 1/2$; $= n^{a-b} m^{a+b-1}$, if $a + b < 1$; $= n^{a-b} \log^{-1} m$, if $a + b = 1, b > 0$; $= n^{a-b}$, otherwise. Then, by results in Hualde (2008), $\overline{D}_{(j)} - D_{(j)} = O_p(g_n^{-1}(\delta_{r_{j-1}}, \delta_{r_j}))$, although particular components of $\overline{D}_{(j)}$ could enjoy faster convergence rates. Thus, we can form residuals

$$\widehat{w}_t^{(r_{j+1}+1, r_j)} = z_t^{(r_{j+1}+1, r_j)} + \overline{D}_{(j)} z_{-r_j,t}, \quad j = 1, \dots, s,$$

and apply to them the previously mentioned techniques, noting that, as mentioned before, the individual components of the $(r_j - r_{j+1}) \times 1$ vector $w_t^{(r_{j+1}+1, r_j)}$ are $I(\delta_{r_j})$ and not cointegrated. Under suitable regularity conditions these estimates will have the same first-order asymptotic properties as corresponding estimates based on the unobservable $w_t^{(r_{j+1}+1, r_j)}$ (cf. Robinson, 2005) and will thence have convergence rates consistent with (18).

Next, denoting by \overline{v} the preliminary estimate of v based on $\overline{D}_{(j)}$, we estimate f from the fractionally differenced residual sequence

$$\begin{aligned} \widetilde{u}_t(\widehat{\delta}, \overline{v}) &= (\widehat{w}_t^{(1, r_s)})' (\widehat{\delta}_{r_s}), \dots, \widehat{w}_t^{(r_3+1, r_2)} (\widehat{\delta}_{r_2}), \\ \widehat{w}_t^{(r_2+1, r_1)} (\widehat{\delta}_{r_1}), z'_{-r_1,t} (\widehat{\delta}_{r_0})', \quad t \in \{1, \dots, n\}, \end{aligned}$$

where we introduce the notation

$$v_t(c) = \Delta^c v_t^\#,$$

for a generic sequence v_t . In particular we can form smoothed weighted autocovariance or periodogram estimates $\widehat{f}(\lambda_j)$ from $\widetilde{u}_t(\widehat{\delta}, \overline{v}), t \in \{1, \dots, n\}$. Again, under suitable regularity conditions these will have the same first-order asymptotic properties as corresponding estimates based on the unobservable u_t , whence the literature on smoothed nonparametric spectrum estimation indicates convergence rates that can be consistent with (20), (21) and (24).

However the rates in (18), (20) and (21) have implications for the choice of m , because (19), (22) and (23), (25) and (26) have also to be satisfied. The estimates of integration orders and spectral densities referred to above have convergence rates no better than $n^{2/5}$, so for example (19) and (22) cannot hold when m grows as fast as n , while satisfying (10), and $\min_{i < j} \delta_{r_i} - \delta_{r_j} \leq 3/5$. Since bias tends to increase with m , for a given rate for m satisfying (19), (22), (23), (25) and (26), bias-reduction may be needed in order to satisfy (18), (20) and (21). Various devices are available for achieving this. With respect to \hat{f} , Parzen (1957) employed higher-order kernels in the frequency domain, while with respect to the $\hat{\delta}$'s Andrews and Sun (2004), Hurvich and Brodsky (2001), Moulines and Soulier (1999) and Robinson and Henry (2003) employed various methods (albeit justified only in case of integration orders falling in the stationarity region). All these methods entail an increase in computation, they rely on greater smoothness in f , and in practice they can sometimes exhibit disappointing small-sample properties. Hualde and Robinson (2006) provide practical guidance on appropriate choices of kernels in order to exploit the greater smoothness in f . However, these methods are not needed if we are prepared to contemplate an arbitrarily slow increase of m with n , since then it is possible to satisfy (19), (22), (23) for arbitrarily small κ, \varkappa, ϕ , this is at cost of a slow convergence rate of estimates of cointegration coefficients when $\delta_{r_i} - \delta_{r_j} < 1/2$ (see Proposition 1 in the Appendix) but it does not directly affect our theorems. Assumption 6 applies when for some $i < j$, $\delta_{r_i} - \delta_{r_j} < 1/2$, while Assumption 6° applies when one or more differences $\delta_{r_i} - \delta_{r_j}$ are larger than $1/2$. The first part of (25) holds whenever m increases with n at algebraic rate, while the second standardly reflects the smoothness index, η .

We next show that our ν estimates (11) have mixed normal asymptotics when for all $i, j, i < j, j \in \{1, \dots, s\}$, $\delta_{r_i} - \delta_{r_j} > 1/2$; normal asymptotics when for all $i, j, i < j, j \in \{1, \dots, s\}$, $\delta_{r_i} - \delta_{r_j} < 1/2$; and that when both kinds of relation are present there is a corresponding combination of limiting distributions. Before presenting the general results, we introduce some further notation. Denote by $W(w)$ the $r_0 \times 1$ vector Brownian motion with covariance matrix Ω . Define

$$G(v; \delta) = (G^{(s)}(v; \delta)', G^{(s-1)}(v; \delta)', \dots, G^{(1)}(v; \delta)')',$$

where for $j = 1, \dots, s$, $G^{(j)}(v; \delta) = (0_{(r_0-r_j, r_j)}, H^{(j)}(v; \delta))$, where $H^{(j)}(v; \delta)$ is a $(r_0 - r_j) \times (r_0 - r_j)$ block diagonal matrix with i th block given by

$$v^{\delta_{r_j-i} - \delta_{r_j-1}} \Gamma^{-1} (\delta_{r_j-i} - \delta_{r_j}) I_{r_j-i-r_{j-1}+1}.$$

For $i = 1, \dots, j$, define

$$\tilde{W}(w; \delta) = \int_0^w G(w-v; \delta) A(1) dW(v),$$

$$\bar{W}(w) = P(0)A(1)W(w),$$

where $P(\lambda) = (P^{(s)}(\lambda)', P^{(s-1)}(\lambda)', \dots, P^{(1)}(\lambda)')'$, with

$$P^{(j)}(\lambda) = \begin{pmatrix} 1_{r_0-r_j, 1} p_{r_j+1}(\lambda) \\ \vdots \\ 1_{r_0-r_j, 1} p_{r_j}(\lambda) \end{pmatrix},$$

in which $1_{i,j}$ is the $i \times j$ matrix of ones, and letting i_j be a $r_0 \times 1$ vector of zeroes except for 1 in the j th component,

$$p_a(\lambda) = i_a' f^{-1}(\lambda), \quad a = 1, \dots, r_1.$$

Let T be a $q \times q$ block diagonal matrix with j th block given by $I_{r_s-j+1-r_s-j+2} \otimes \mathcal{Y}_{(r_s-j+1)}, j = 1, \dots, s$, where

$$\mathcal{Y}_{(r_i)} = (0_{r_0-r_i, r_i}, I_{r_0-r_i}) \mathcal{Y}(0_{r_0-r_i, r_i}, I_{r_0-r_i})', \quad \text{for } i = 1, \dots, s.$$

Let $\Lambda_{(n)}^S$ be a $q \times q$ block diagonal matrix with j th block given by $I_{r_s-j+1-r_s-j+2} \otimes \Lambda_{r_s-j+1}^S, j = 1, \dots, s$, where $\Lambda_{r_i}^S, i = 1, \dots, s$, is a block diagonal matrix with k th block given by $n^{\delta_{r_i-k} - \delta_{r_i}} I_{r_i-k-r_{i-k+1}},$ for $k = 1, \dots, i$. Similarly, let $\Lambda_{(n)}^W$ be a $q \times q$ block diagonal matrix with j th block given by $I_{r_s-j+1-r_s-j+2} \otimes \Lambda_{r_s-j+1}^W, j = 1, \dots, s$, where $\Lambda_{r_i}^W, i = 1, \dots, s$, is a block diagonal matrix with k th block given by $m^{1/2} \lambda_m^{\delta_{r_i-k} - \delta_{r_i}} I_{r_i-k-r_{i-k+1}},$ for $k = 1, \dots, i$. Also, let $F(\lambda)$ be a $q \times q$ matrix consisting of $s \times s$ blocks, the (i, j) th one being

$$F_{ij}(\lambda) = \begin{pmatrix} f^{r_s-i+2+1, r_s-j+2+1}(\lambda) & \dots & f^{r_s-i+2+1, r_s-j+1}(\lambda) \\ \vdots & \ddots & \vdots \\ f^{r_s-i+1, r_s-j+2+1}(\lambda) & \dots & f^{r_s-i+1, r_s-j+1}(\lambda) \end{pmatrix} \otimes 1_{r_0-r_s-i+1, r_0-r_s-j+1},$$

$i, j = 1, \dots, s$, where $f^{a,b}$ is the (a, b) th element of f^{-1} . Let D be a $q \times (sr_0 - \sum_{j=1}^s r_j)$ block diagonal matrix with i th block given by $1_{r_s-i+1-r_s-i+2, 1} \otimes I_{r_0-r_s-i+1}, i = 1, \dots, s$, and let A a matrix consisting of $s \times s$ blocks,

$$A = \begin{pmatrix} A_{(s,s)} & \dots & A_{(s,1)} \\ \vdots & \ddots & \vdots \\ A_{(1,s)} & \dots & A_{(1,1)} \end{pmatrix},$$

where for $k, l = 1, \dots, s$,

$$A_{(k,l)} = \begin{pmatrix} A_{(k,l)}^{(r_{k+1}+1, r_{l+1}+1)} & \dots & A_{(k,l)}^{(r_{k+1}+1, r_l)} \\ \vdots & \ddots & \vdots \\ A_{(k,l)}^{(r_k, r_{l+1}+1)} & \dots & A_{(k,l)}^{(r_k, r_l)} \end{pmatrix},$$

in which for $v = r_{k+1} + 1, \dots, r_k, w = r_{l+1} + 1, \dots, r_l, A_{(k,l)}^{(v,w)}$ is a $(r_0 - r_k) \times (r_0 - r_l)$ matrix with (y, z) th element given by

$$\frac{f^{vw}(0) f_{yz}(0)}{2(1 - (\delta_y - \delta_{r_k}) - (\delta_z - \delta_{r_l}))}, \quad y = r_k + 1, \dots, r_0, z = r_l + 1, \dots, r_0,$$

$f_{a,b}$ being the (a, b) th element of f . Finally, let $O = (O^{(s)'}, \dots, O^{(1)'})'$ be the vector which collects the integration orders of the transformed regressors (see Appendix), where

$$O^{(j)} = 1_{r_j-r_{j+1}, 1} \otimes \left(1_{1, r_j-1-r_j} (\delta_{r_j-1} - \delta_{r_j}), 1_{1, r_j-2-r_{j-1}} \times (\delta_{r_j-2} - \delta_{r_j}), \dots, 1_{1, r_0-r_1} (\delta_{r_0} - \delta_{r_j}) \right)'$$

for $j = 1, \dots, s$. By \Rightarrow we mean convergence in the Skorohod J_1 topology of $D[0, 1]$, and by \rightarrow_d convergence in distribution.

Theorem 1. Let Assumptions 1 and 3–6 hold. Then,

(i) if $\delta_{r_i} - \delta_{r_j} > 1/2$ for all $i < j, j \in \{1, \dots, s\}$,

$$\Lambda_{(n)}^S (T')^{-1} (\widehat{\nu}_m(\widehat{\delta}) - \nu) \Rightarrow \left(F(0) \circ D \int_0^1 \tilde{W}(w; \delta) \tilde{W}'(w; \delta) dw D' \right)^{-1} \times \int_0^1 D \tilde{W}(w; \delta) \circ d\bar{W}(w);$$

(ii) if $\delta_{r_i} - \delta_{r_j} < 1/2$ for all $i < j, j \in \{1, \dots, s\}$,

$$\Lambda_{(n)}^W (T')^{-1} (\widehat{\nu}_m(\widehat{\delta}) - \nu) \rightarrow_d N(0, A^{-1});$$

(iii) if there are $q_S \times q$ and $q_W \times q$ selection matrices S_S and S_W , respectively, such that $q_S + q_W = q$ and all integration order differences in $S_S O$ and $S_W O$ are larger and smaller than $1/2$, respectively,

$$\begin{pmatrix} S_S A_{(n)}^S \\ S_W A_{(n)}^S \end{pmatrix} (T')^{-1} (\widehat{v}_m(\widehat{\delta}) - \nu) \Rightarrow M^{-1}N, \tag{27}$$

where

$$M = \begin{pmatrix} S_S \left(F(0) \circ D \int_0^1 \widetilde{W}(w; \delta) \widetilde{W}'(w; \delta) dw D' \right) S_S' & 0_{q_S \times q_W} \\ 0_{q_W \times q_S} & S_W A_{(n)}^S \end{pmatrix},$$

and

$$N = \begin{pmatrix} S_S \int_0^1 D \widetilde{W}(w; \delta) \circ d\overline{W}(w) \\ X \end{pmatrix},$$

where X is a $q_W \times 1$ Gaussian vector independent of the processes $\widetilde{W}(w; \delta)$, $\overline{W}(w)$, with mean zero and variance covariance matrix $S_W A_{(n)}^S$.

Theorem 2. Let Assumptions 1, 3°, 4, 5° and 6° hold. Then identical results to those in Theorem 1 apply to $\widehat{v}_m^\circ(\widehat{\delta})$.

Finally, we present the useful and interesting global outcome that the limit distributions presented in Theorems 1 and 2 all lead to the Wald test statistics (13), (14) having the same, standard, limit behaviour. Denote by χ_J^2 a chi-square variate with J degrees of freedom.

Corollary 1. Let Assumptions 1–6 hold. Then under H_0 ,

$$W_m(\delta), W_m(\widehat{\delta}) \rightarrow_d \chi_J^2.$$

Corollary 2. Let Assumptions 1, 2, 3°, 4, 5° and 6° hold. Then under H_0 ,

$$W_m^\circ(\delta), W_m^\circ(\widehat{\delta}) \rightarrow_d \chi_J^2.$$

Corollaries 1 and 2 are direct consequences of Theorems 1 and 2, respectively, along with Assumption 2. Note that the standard limit theory of Wald tests, familiar in many classical situations in econometrics and associated with optimal procedures in the $I(1)/I(0)$ cointegration literature (see, e.g., Johansen, 1991; Phillips, 1991a,b), is shown to hold here simultaneously for weak (including stationary) and strong relations, and in the possible presence of unknown integration orders of observables and/or cointegrating errors.

4. Finite sample evidence

A Monte Carlo study of finite sample behaviour was carried out, comparing the performance of our W estimates of the cointegrating parameters with that of the NBS estimates (in terms of Monte Carlo bias and standard deviation), and also the goodness of the χ^2 approximation established in Corollary 1. We consider (2), (15), with $r_0 = 3, A(z) = (1 - \phi z)^{-1} I_3$, for values $\phi = 0, 0.3$, the ε_t being Gaussian with covariance matrix Ω having ij th element ω_{ij} , fixing $\omega_{ii} = 1, i = 1, 2, 3, \omega_{ij} = \rho = 0, 0.5, -0.4$, for $i \neq j$. We also impose $r_1 = 2, r_2 = 1$, which implies $\delta_1 < \delta_2 < \delta_3$ in (2). Thus, we estimate the model

$$z_{1t} + \nu_{12} z_{2t} + \nu_{13} z_{3t} = \Delta^{-\delta_1} u_{1t}^\#, \tag{28}$$

$$z_{2t} + \nu_{23} z_{3t} = \Delta^{-\delta_2} u_{2t}^\#, \tag{29}$$

$$z_{3t} = \Delta^{-\delta_3} u_{3t}^\#,$$

where we set $\nu_{12} = \nu_{13} = \nu_{23} = 1$ and consider combinations $(\delta_1, \delta_2, \delta_3) = (0, 1, 2), (0.4, 0.8, 1.6), (0.2, 1, 1.4), (0, 0.4, 0.8)$, noting that with the parameter values of $\nu = (\nu_{12}, \nu_{13}, \nu_{23})'$, by construction, z_{2t}, z_{3t} , are individually $I(\delta_3)$, whereas z_{1t} is $I(\delta_2)$, so we capture the phenomenon of multicointegration, z_{1t} cointegrating with the cointegrating error arising from the relation between z_{2t} and z_{3t} , producing a new cointegrating error which further reduces the integration order of the observables.

We generated 1000 series of lengths $n = 64, 128, 256$, and for the estimates of ν we chose bandwidths $m = 4, 5, 6$, depending on whether $n = 64, 128, 256$, respectively, whereas for the estimates of δ and f we employed the two possibilities (I, II) = (30, 25), (60, 45), (120, 75) for $n = 64, 128, 256$, respectively. Marinucci and Robinson (2001) justified the appropriateness of the relatively small bandwidths m . We employ a sequential procedure in order to obtain our final estimates of ν . The first step consists of obtaining a preliminary estimate of ν (the NBS estimate), taking the form

$$(\bar{\nu}_{12}, \bar{\nu}_{13})' = - \left(\sum_{j=0}^m s_j \text{Re} \{ I_{z_{-1}}(\lambda_j) \} \right)^{-1} \sum_{j=0}^m s_j \text{Re} \{ I_{z_{-1}z_1}(\lambda_j) \}, \tag{30}$$

$$\bar{\nu}_{23} = - \left(\sum_{j=0}^m s_j \text{Re} \{ I_{z_3}(\lambda_j) \} \right)^{-1} \sum_{j=0}^m s_j \text{Re} \{ I_{z_3z_2}(\lambda_j) \}.$$

Note that (29) is a standard cointegrating relation with cointegrating gap $\delta_3 - \delta_2$, so the asymptotic properties of $\bar{\nu}_{23}$ can be directly derived from Robinson and Marinucci (2001, 2003). The estimation results concerning (28) need to be discussed with caution. The regressors in that equation are $I(\delta_3)$, so given that the cointegrating error is $I(\delta_1)$, one might think that the cointegrating gap $\delta_3 - \delta_1$ should drive the convergence rate of $(\bar{\nu}_{12}, \bar{\nu}_{13})'$. However, this is not the case because the regressors z_{2t}, z_{3t} , in (28) are cointegrated, which invalidates the standard results, although the problem is easily solved by a simple linear transformation. Replacing (28) in (30),

$$\begin{aligned} (\bar{\nu}_{12} - \nu_{12}, \bar{\nu}_{13} - \nu_{13})' &= - \left(\sum_{j=0}^m s_j \text{Re} \{ I_{z_{-1}}(\lambda_j) \} \right)^{-1} \\ &\quad \times \sum_{j=0}^m s_j \text{Re} \left\{ I_{z_{-1} \Delta^{-\delta_1} u_1^\#}(\lambda_j) \right\}, \end{aligned}$$

and defining

$$T = \begin{pmatrix} 1 & \nu_{23} \\ 0 & 1 \end{pmatrix}$$

(which corresponds to the definition of T given in the previous section), clearly

$$\begin{aligned} (T')^{-1} (\bar{\nu}_{12} - \nu_{12}, \bar{\nu}_{13} - \nu_{13})' &= - \left(\sum_{j=0}^m s_j \text{Re} \{ I_R(\lambda_j) \} \right)^{-1} \\ &\quad \times \sum_{j=0}^m s_j \text{Re} \left\{ I_{R \Delta^{-\delta_1} u_1^\#}(\lambda_j) \right\}, \end{aligned}$$

where $I_R(\cdot)$ is the periodogram of the bivariate process $(\Delta^{-\delta_2} u_{2t}^\#, z_{3t})'$, and $I_{R \Delta^{-\delta_1} u_1^\#}(\cdot)$ is the cross-periodogram between $(\Delta^{-\delta_2} u_{2t}^\#, z_{3t})'$ and $\Delta^{-\delta_1} u_{1t}^\#$. Thus, in this particular case, it is just the gap $\delta_2 - \delta_1$ which drives the convergence rate of $(\bar{\nu}_{12}, \bar{\nu}_{13})'$.

In the second step we estimate $\delta_1, \delta_2, \delta_3$, by Shimotsu and Phillips' (2005) exact local Whittle procedure applied to series r_{1t}, r_{2t}, z_{3t} , where $r_{1t} = z_{1t} + \bar{\nu}_{12} z_{2t} + \bar{\nu}_{13} z_{3t}, r_{2t} = z_{2t} + \bar{\nu}_{23} z_{3t}$, where the respective optimization intervals are taken $[\delta_i - 1, \delta_i + 1], i = 1, 2, 3$. Alternative semiparametric estimation procedures

Table 1
Monte Carlo bias of estimates of ν for $\phi = 0, \rho = 0.5$, 1000 replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	-0.004	-0.104	-0.006	-0.080	-0.002	-0.053	-0.003	-0.063	0.000	-0.036	0.000	-0.037
$\hat{\nu}_{12}^{II}$	-0.006	-0.111	-0.008	-0.085	-0.002	-0.059	-0.004	-0.068	-0.001	-0.043	-0.001	-0.044
$\hat{\nu}_{12}^{NBL}$	-0.022	-0.196	-0.047	-0.183	-0.010	-0.144	-0.029	-0.158	-0.005	-0.114	-0.017	-0.123
$\hat{\nu}_{13}^I$	-0.004	-0.101	-0.004	-0.078	-0.002	-0.052	-0.002	-0.057	0.000	-0.036	0.000	-0.034
$\hat{\nu}_{13}^{II}$	-0.005	-0.107	-0.005	-0.085	-0.002	-0.058	-0.003	-0.062	-0.001	-0.042	-0.001	-0.042
$\hat{\nu}_{13}^{NBL}$	-0.021	-0.193	-0.042	-0.193	-0.010	-0.143	-0.026	-0.158	-0.005	-0.113	-0.016	-0.123
$\hat{\nu}_{23}^I$	0.005	0.007	-0.040	-0.057	0.002	0.004	-0.028	-0.032	0.001	0.001	-0.009	-0.019
$\hat{\nu}_{23}^{II}$	0.005	0.006	-0.043	-0.060	0.003	0.004	-0.031	-0.035	0.001	0.001	-0.010	-0.025
$\hat{\nu}_{23}^{NBL}$	-0.010	-0.033	-0.130	-0.147	-0.005	-0.018	-0.112	-0.112	-0.003	-0.011	-0.075	-0.090

Table 2
Monte Carlo bias of estimates of ν for $\phi = 0, \rho = 0$, 1000 replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.000	-0.002	0.001	-0.003	0.000	-0.006	-0.001	-0.002	-0.001	-0.003	0.000	0.010
$\hat{\nu}_{12}^{II}$	0.000	-0.004	0.001	-0.002	0.000	-0.005	0.000	-0.002	-0.001	-0.003	0.000	0.009
$\hat{\nu}_{12}^{NBL}$	0.000	-0.001	-0.001	-0.001	-0.001	-0.005	0.000	-0.001	-0.001	-0.003	0.000	0.008
$\hat{\nu}_{13}^I$	0.000	-0.004	0.000	-0.004	0.000	-0.006	-0.001	-0.002	-0.001	-0.004	0.000	0.010
$\hat{\nu}_{13}^{II}$	0.000	-0.006	0.000	-0.004	0.000	-0.006	-0.001	-0.001	-0.001	-0.004	0.000	0.009
$\hat{\nu}_{13}^{NBL}$	0.000	-0.003	-0.002	-0.004	-0.001	-0.005	0.000	-0.001	-0.001	-0.004	0.000	0.007
$\hat{\nu}_{23}^I$	-0.004	-0.004	0.001	0.003	-0.001	0.003	0.000	0.002	0.000	0.001	0.000	0.001
$\hat{\nu}_{23}^{II}$	-0.004	-0.004	0.000	0.003	-0.001	0.003	0.000	0.002	0.000	0.001	-0.001	-0.001
$\hat{\nu}_{23}^{NBL}$	-0.005	-0.004	-0.003	0.009	-0.002	0.000	0.001	0.004	0.000	0.001	-0.003	-0.001

Table 3
Monte Carlo bias of estimates of ν for $\phi = 0, \rho = -0.4$, 1000 replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	-0.005	0.054	0.001	0.073	-0.002	0.030	-0.002	0.051	0.000	0.021	0.000	0.036
$\hat{\nu}_{12}^{II}$	-0.004	0.061	0.002	0.083	-0.001	0.036	-0.001	0.059	0.000	0.028	0.000	0.041
$\hat{\nu}_{12}^{NBL}$	0.009	0.145	0.035	0.198	0.003	0.102	0.020	0.160	0.001	0.087	0.013	0.126
$\hat{\nu}_{13}^I$	-0.005	0.053	-0.001	0.078	-0.002	0.030	-0.003	0.054	0.000	0.021	0.000	0.037
$\hat{\nu}_{13}^{II}$	-0.004	0.061	0.000	0.091	-0.001	0.037	-0.002	0.063	0.000	0.028	0.000	0.043
$\hat{\nu}_{13}^{NBL}$	0.008	0.151	0.043	0.252	0.003	0.105	0.023	0.194	0.001	0.088	0.013	0.144
$\hat{\nu}_{23}^I$	-0.008	-0.008	0.056	0.055	-0.003	0.000	0.015	0.040	-0.001	0.000	0.012	0.017
$\hat{\nu}_{23}^{II}$	-0.008	-0.007	0.066	0.061	-0.003	0.001	0.021	0.044	-0.001	0.000	0.015	0.020
$\hat{\nu}_{23}^{NBL}$	0.013	0.024	0.133	0.126	0.007	0.018	0.088	0.103	0.002	0.009	0.062	0.069

(e.g., Robinson, 1995a,b) are available, but the exact local Whittle procedure is specifically designed to be applied to fractionally integrated processes, as in our present situation. Then, residuals

$$\tilde{u}_t(\hat{\delta}, \bar{\nu}) = \left(\Delta^{\hat{\delta}_1} z_{1t}^\# + \bar{\nu}_{12} \Delta^{\hat{\delta}_1} z_{2t}^\# + \bar{\nu}_{13} \Delta^{\hat{\delta}_1} z_{3t}^\# + \bar{\nu}_{23} \Delta^{\hat{\delta}_2} z_{3t}^\# + \Delta^{\hat{\delta}_3} z_{3t}^\# \right)$$

are computed, and we estimate $f(\lambda_j)$ by

$$\hat{f}(\lambda_j) = \frac{1}{2m+1} \sum_{k=j-b}^{j+b} I_{\tilde{u}(\hat{\delta}, \bar{\nu})}(\lambda_k), \tag{31}$$

to compute finally the feasible estimate of ν , $\hat{\nu}_m(\hat{\delta})$, as in (11).

In Tables 1–12 we report Monte Carlo biases and standard deviations (SD) of three different estimates of ν_{ij} , $1 \leq i < j = 2, 3$: $\hat{\nu}_{ij}^I, \hat{\nu}_{ij}^{II}$, which denote our semiparametric estimates where bandwidths I, II, are employed in the estimation of δ, f , and $\bar{\nu}_{ij}^{NBL}$, which is the NBLs estimate, for the different $\rho, \phi, (\delta_1, \delta_2, \delta_3)$,

combinations (noting that the orders' combinations are uniquely identified in the tables by δ_3). The Monte Carlo results should be interpreted in relation to the respective convergence rates of the estimates for the different $(\delta_1, \delta_2, \delta_3)$ combinations: for (0, 1, 2) all estimates are n -consistent; for (0.4, 0.8, 1.6), those of ν_{23} are $n^{0.8}$ -consistent, whereas those of ν_{12}, ν_{13} , are $n^{0.4}m^{0.1}$ -consistent (for I and II estimates) and $n^{0.4}$ -consistent (for NBLs estimates); for (0.2, 1, 1.4), estimates of ν_{23} are $n^{0.4}m^{0.1}$ -consistent (for I and II estimates) and $n^{0.4}$ -consistent (for NBLs estimates), whereas those of ν_{12}, ν_{13} , are $n^{0.8}$ -consistent in all cases; finally, for (0, 0.4, 0.8), the I and II estimates are in all cases $n^{0.4}m^{0.1}$ -consistent, whereas the NBLs estimate of ν_{23} is $n^{0.4}$ -consistent, and those of ν_{12}, ν_{13} , are $n^{0.4}m^{-0.4}$ -consistent (but $n^{0.4}m^{0.1}$ -consistent if $\rho = 0$). Note also that in view of our theoretical results it is expected (and reflected in the Monte Carlo) that the results for the estimates of ν_{12}, ν_{13} will be very similar, because in our present situation the limiting distribution of the vector of estimates of these parameters is singular.

In the white noise situation, detailed results for bias are presented in Tables 1–3. Overall the I estimates dominate, the I

Table 4
Monte Carlo bias of estimates of ν for $\phi = 0.3, \rho = 0.5, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.000	-0.107	-0.006	-0.083	0.000	-0.054	-0.001	-0.065	0.000	-0.036	0.000	-0.040
$\hat{\nu}_{12}^{II}$	-0.003	-0.113	-0.008	-0.088	-0.001	-0.061	-0.003	-0.070	0.000	-0.042	-0.001	-0.046
$\hat{\nu}_{12}^{NBS}$	-0.023	-0.195	-0.047	-0.182	-0.010	-0.143	-0.029	-0.157	-0.005	-0.113	-0.017	-0.123
$\hat{\nu}_{13}^I$	0.001	-0.101	-0.003	-0.078	0.000	-0.053	0.000	-0.058	0.000	-0.035	0.000	-0.036
$\hat{\nu}_{13}^{II}$	-0.002	-0.107	-0.005	-0.085	-0.001	-0.059	-0.002	-0.064	0.000	-0.041	-0.001	-0.043
$\hat{\nu}_{13}^{NBS}$	-0.021	-0.193	-0.042	-0.192	-0.010	-0.142	-0.026	-0.157	-0.005	-0.113	-0.016	-0.123
$\hat{\nu}_{23}^I$	0.006	0.007	-0.049	-0.064	0.002	0.005	-0.031	-0.038	0.001	0.001	-0.013	-0.019
$\hat{\nu}_{23}^{II}$	0.005	0.005	-0.051	-0.067	0.002	0.004	-0.035	-0.041	0.001	0.001	-0.012	-0.025
$\hat{\nu}_{23}^{NBS}$	-0.010	-0.033	-0.130	-0.146	-0.005	-0.018	-0.112	-0.112	-0.003	-0.011	-0.075	-0.090

Table 5
Monte Carlo bias of estimates of ν for $\phi = 0.3, \rho = 0, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.000	-0.002	0.001	-0.001	-0.001	-0.005	-0.001	-0.004	-0.001	-0.003	0.000	0.011
$\hat{\nu}_{12}^{II}$	0.000	-0.003	0.000	0.000	-0.001	-0.005	0.000	-0.003	-0.001	-0.003	0.000	0.010
$\hat{\nu}_{12}^{NBS}$	-0.001	-0.001	-0.001	-0.001	-0.001	-0.006	0.000	-0.002	-0.001	-0.003	0.000	0.008
$\hat{\nu}_{13}^I$	0.000	-0.004	0.000	-0.001	-0.001	-0.006	-0.001	-0.005	-0.001	-0.003	0.000	0.011
$\hat{\nu}_{13}^{II}$	0.000	-0.006	-0.001	-0.001	-0.001	-0.005	-0.001	-0.003	-0.001	-0.003	0.000	0.010
$\hat{\nu}_{13}^{NBS}$	0.000	-0.003	-0.002	-0.004	-0.001	-0.005	0.000	-0.001	-0.001	-0.004	0.000	0.007
$\hat{\nu}_{23}^I$	-0.003	-0.006	-0.001	-0.002	-0.001	0.004	-0.001	-0.001	0.000	0.001	0.001	0.003
$\hat{\nu}_{23}^{II}$	-0.003	-0.006	-0.002	-0.001	-0.001	0.004	-0.002	-0.001	0.000	0.001	0.000	0.000
$\hat{\nu}_{23}^{NBS}$	-0.005	-0.004	-0.004	0.009	-0.002	0.000	0.000	0.004	0.000	0.001	-0.003	-0.001

Table 6
Monte Carlo bias of estimates of ν for $\phi = 0.3, \rho = -0.4, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	-0.007	0.055	0.001	0.076	-0.003	0.033	-0.004	0.053	-0.001	0.020	0.000	0.039
$\hat{\nu}_{12}^{II}$	-0.006	0.062	0.002	0.086	-0.002	0.039	-0.002	0.061	-0.001	0.027	0.000	0.043
$\hat{\nu}_{12}^{NBS}$	0.008	0.145	0.035	0.197	0.003	0.102	0.020	0.160	0.001	0.087	0.013	0.126
$\hat{\nu}_{13}^I$	-0.007	0.054	-0.003	0.078	-0.003	0.033	-0.006	0.054	-0.001	0.020	0.000	0.039
$\hat{\nu}_{13}^{II}$	-0.006	0.061	-0.001	0.092	-0.002	0.039	-0.004	0.064	-0.001	0.027	0.000	0.044
$\hat{\nu}_{13}^{NBS}$	0.008	0.151	0.043	0.251	0.003	0.105	0.023	0.193	0.001	0.088	0.013	0.144
$\hat{\nu}_{23}^I$	-0.010	-0.010	0.062	0.060	-0.005	0.000	0.018	0.043	-0.001	0.001	0.015	0.020
$\hat{\nu}_{23}^{II}$	-0.009	-0.009	0.071	0.065	-0.004	0.001	0.022	0.046	-0.001	0.000	0.016	0.021
$\hat{\nu}_{23}^{NBS}$	0.013	0.024	0.133	0.125	0.007	0.018	0.088	0.103	0.002	0.009	0.062	0.069

and II estimates being clearly superior to NBS, except for some cases when $\rho = 0$, although in this case the three estimates behave very similarly with very small biases. Our results reflect clearly the differences in rates of convergence corresponding to the different $(\delta_1, \delta_2, \delta_3)$ combinations, all the estimates behaving similarly when $(\delta_1, \delta_2, \delta_3) = (0, 1, 2), (0, 0.4, 0.8)$, those of ν_{23} being superior to corresponding estimates of ν_{12}, ν_{13} , when $(\delta_1, \delta_2, \delta_3) = (0.4, 0.8, 1.6)$, and the opposite happening when $(\delta_1, \delta_2, \delta_3) = (0.2, 1, 1.4)$. Biases react in the appropriate direction as n increases, and positive (negative) correlation induces negative (positive) bias in general. Results for bias under the AR structure with $\phi = 0.3$ are reported in Tables 4–6. Results for the NBS estimate (and also those for I, II estimates when $\rho = 0$) are almost unaffected by the AR structure, with minor changes which tend to improve the I, II estimates when $(\delta_1, \delta_2, \delta_3) = (0, 1, 2)$ (and $\rho \neq 0$), worsening them slightly for other combinations.

Results for SD are presented in Tables 7–12. Regarding ν_{12}, ν_{13} , the NBS estimate emerges as competitive relative to the I, II estimates, whereas the NBS estimates of ν_{23} are substantially

worse (except in the $(\delta_1, \delta_2, \delta_3) = (0, 0.4, 0.8)$ case). In fact, when $(\delta_1, \delta_2, \delta_3) = (0, 1, 2), (0.2, 1, 1.4)$, the three estimates of ν_{12}, ν_{13} , perform similarly, NBS being best when $(\delta_1, \delta_2, \delta_3) = (0.4, 0.8, 1.6)$ for small n , but worsening relative to the I, II estimates as n increases. In the $(\delta_1, \delta_2, \delta_3) = (0, 0.4, 0.8)$ case NBS dominates, but again differences shrink noticeably as n increases, so for $n = 256$ all estimates are similar in terms of SD. The II estimate appears to be slightly worse than I, and variations in ρ do not have an important impact on the results, although estimates seem to perform best and worst when $\rho = -0.4$ and $\rho = 0$, respectively. The same qualitative conclusions apply in the presence of AR structure, although the overall results worsen slightly in this case.

We next studied the Wald statistic for testing the (correct) null hypothesis $\nu = (1, 1, 1)'$, which is computed as

$$W = (\hat{\nu}_m(\hat{\delta}) - \nu)' \hat{b}_m(\hat{\delta}) (\hat{\nu}_m(\hat{\delta}) - \nu).$$

Tables 13–15 contain empirical sizes corresponding to nominal $\alpha = 0.01, 0.05, 0.10$. The first main finding of our experiment is

Table 7
Monte Carlo SD of estimates of ν for $\phi = 0, \rho = 0.5, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.058	0.259	0.093	0.270	0.026	0.157	0.048	0.173	0.013	0.112	0.026	0.113
$\hat{\nu}_{12}^H$	0.059	0.263	0.094	0.270	0.026	0.162	0.048	0.175	0.013	0.114	0.026	0.115
$\hat{\nu}_{12}^{NB}$	0.058	0.220	0.094	0.195	0.027	0.143	0.050	0.138	0.014	0.109	0.029	0.099
$\hat{\nu}_{13}^I$	0.058	0.252	0.085	0.245	0.026	0.155	0.045	0.163	0.013	0.112	0.024	0.105
$\hat{\nu}_{13}^H$	0.058	0.255	0.086	0.243	0.025	0.159	0.045	0.164	0.013	0.114	0.024	0.107
$\hat{\nu}_{13}^{NB}$	0.058	0.214	0.082	0.171	0.027	0.141	0.045	0.127	0.014	0.108	0.026	0.091
$\hat{\nu}_{23}^I$	0.038	0.079	0.283	0.221	0.018	0.039	0.194	0.150	0.009	0.021	0.109	0.098
$\hat{\nu}_{23}^H$	0.038	0.079	0.288	0.223	0.018	0.038	0.200	0.154	0.009	0.021	0.119	0.101
$\hat{\nu}_{23}^{NB}$	0.068	0.108	0.336	0.193	0.034	0.059	0.266	0.144	0.018	0.037	0.179	0.100

Table 8
Monte Carlo SD of estimates of ν for $\phi = 0, \rho = 0, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.070	0.282	0.109	0.299	0.030	0.192	0.053	0.190	0.016	0.128	0.029	0.126
$\hat{\nu}_{12}^H$	0.070	0.279	0.113	0.295	0.030	0.192	0.053	0.189	0.016	0.131	0.029	0.128
$\hat{\nu}_{12}^{NB}$	0.064	0.225	0.091	0.201	0.029	0.164	0.051	0.137	0.016	0.117	0.028	0.099
$\hat{\nu}_{13}^I$	0.070	0.287	0.115	0.309	0.030	0.193	0.054	0.196	0.016	0.128	0.029	0.128
$\hat{\nu}_{13}^H$	0.070	0.284	0.118	0.306	0.030	0.194	0.055	0.194	0.016	0.131	0.029	0.130
$\hat{\nu}_{13}^{NB}$	0.064	0.230	0.100	0.216	0.029	0.165	0.052	0.143	0.016	0.117	0.029	0.101
$\hat{\nu}_{23}^I$	0.044	0.077	0.373	0.252	0.020	0.042	0.230	0.164	0.010	0.022	0.132	0.109
$\hat{\nu}_{23}^H$	0.045	0.078	0.392	0.253	0.021	0.042	0.238	0.166	0.010	0.022	0.143	0.112
$\hat{\nu}_{23}^{NB}$	0.075	0.117	0.421	0.218	0.041	0.066	0.288	0.160	0.020	0.037	0.201	0.113

Table 9
Monte Carlo SD of estimates of ν for $\phi = 0, \rho = -0.4, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.060	0.253	0.085	0.232	0.025	0.158	0.048	0.153	0.011	0.094	0.023	0.101
$\hat{\nu}_{12}^H$	0.058	0.255	0.086	0.234	0.025	0.159	0.049	0.154	0.011	0.097	0.024	0.101
$\hat{\nu}_{12}^{NB}$	0.063	0.227	0.088	0.180	0.029	0.154	0.049	0.129	0.014	0.106	0.027	0.094
$\hat{\nu}_{13}^I$	0.060	0.259	0.101	0.266	0.025	0.160	0.053	0.170	0.011	0.095	0.025	0.109
$\hat{\nu}_{13}^H$	0.058	0.260	0.101	0.270	0.025	0.161	0.054	0.171	0.011	0.097	0.026	0.108
$\hat{\nu}_{13}^{NB}$	0.063	0.231	0.099	0.207	0.030	0.156	0.053	0.143	0.014	0.107	0.029	0.100
$\hat{\nu}_{23}^I$	0.045	0.077	0.288	0.213	0.018	0.037	0.201	0.144	0.009	0.019	0.119	0.103
$\hat{\nu}_{23}^H$	0.043	0.077	0.292	0.212	0.018	0.037	0.209	0.145	0.009	0.019	0.132	0.105
$\hat{\nu}_{23}^{NB}$	0.073	0.100	0.338	0.184	0.036	0.056	0.255	0.141	0.018	0.031	0.200	0.107

Table 10
Monte Carlo SD of estimates of ν for $\phi = 0.3, \rho = 0.5, 1000$ replications.

n	64				128				256			
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
$\hat{\nu}_{12}^I$	0.067	0.271	0.106	0.291	0.030	0.167	0.055	0.192	0.015	0.123	0.030	0.124
$\hat{\nu}_{12}^H$	0.064	0.268	0.102	0.284	0.027	0.165	0.051	0.186	0.013	0.117	0.027	0.119
$\hat{\nu}_{12}^{NB}$	0.059	0.222	0.095	0.195	0.027	0.144	0.050	0.139	0.014	0.110	0.029	0.099
$\hat{\nu}_{13}^I$	0.067	0.263	0.099	0.263	0.030	0.166	0.052	0.182	0.014	0.122	0.028	0.116
$\hat{\nu}_{13}^H$	0.064	0.261	0.095	0.256	0.027	0.164	0.048	0.175	0.013	0.117	0.025	0.110
$\hat{\nu}_{13}^{NB}$	0.058	0.216	0.083	0.172	0.027	0.141	0.046	0.127	0.014	0.108	0.026	0.091
$\hat{\nu}_{23}^I$	0.046	0.094	0.300	0.236	0.022	0.048	0.205	0.165	0.012	0.025	0.118	0.110
$\hat{\nu}_{23}^H$	0.043	0.088	0.296	0.229	0.020	0.043	0.205	0.159	0.010	0.022	0.121	0.103
$\hat{\nu}_{23}^{NB}$	0.067	0.109	0.339	0.195	0.034	0.059	0.267	0.145	0.018	0.037	0.179	0.100

that even if the choice of bandwidth was not affecting dramatically the bias and SD results, it affects heavily sizes, which increase substantially as the bandwidth decreases, so I Wald statistics, W^I (based on I estimates) are in general undersized (most noticeably

when $(\delta_1, \delta_2, \delta_3) = (0, 1, 2)$), whereas Wald statistics W^{II} (based on II estimates) are usually oversized (especially when $(\delta_1, \delta_2, \delta_3) = (0, 0.4, 0.8)$), although in most cases the distance from nominal sizes is not of overriding concern, taking into account

Table 11
Monte Carlo SD of estimates of ν for $\phi = 0.3, \rho = 0$, 1000 replications.

n	64				128				256				
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4	0.8
$\hat{\nu}_{12}^I$	0.082	0.300	0.120	0.324	0.035	0.205	0.060	0.208	0.018	0.139	0.033	0.138	
	$\hat{\nu}_{12}^{II}$	0.077	0.290	0.119	0.312	0.032	0.199	0.057	0.200	0.016	0.135	0.030	0.133
	$\bar{\nu}_{12}^{NB}$	0.064	0.228	0.091	0.200	0.029	0.165	0.051	0.137	0.016	0.117	0.029	0.099
$\hat{\nu}_{13}^I$	0.082	0.305	0.126	0.334	0.035	0.205	0.061	0.216	0.018	0.138	0.033	0.142	
	$\hat{\nu}_{13}^{II}$	0.077	0.295	0.124	0.323	0.032	0.200	0.058	0.206	0.016	0.135	0.030	0.136
	$\bar{\nu}_{13}^{NB}$	0.063	0.233	0.100	0.216	0.029	0.167	0.052	0.143	0.016	0.117	0.029	0.102
$\hat{\nu}_{23}^I$	0.054	0.094	0.399	0.276	0.025	0.051	0.246	0.182	0.012	0.027	0.141	0.122	
	$\hat{\nu}_{23}^{II}$	0.051	0.090	0.403	0.269	0.023	0.046	0.246	0.174	0.010	0.024	0.145	0.115
	$\bar{\nu}_{23}^{NB}$	0.075	0.118	0.424	0.221	0.042	0.066	0.289	0.161	0.020	0.037	0.201	0.113

Table 12
Monte Carlo SD of estimates of ν for $\phi = 0.3, \rho = -0.4$, 1000 replications.

n	64				128				256				
	δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4	0.8
$\hat{\nu}_{12}^I$	0.070	0.262	0.095	0.252	0.028	0.170	0.055	0.168	0.013	0.104	0.027	0.110	
	$\hat{\nu}_{12}^{II}$	0.064	0.259	0.092	0.246	0.027	0.164	0.053	0.161	0.012	0.100	0.025	0.104
	$\bar{\nu}_{12}^{NB}$	0.063	0.230	0.089	0.180	0.029	0.155	0.049	0.129	0.014	0.106	0.027	0.094
$\hat{\nu}_{13}^I$	0.070	0.268	0.113	0.289	0.029	0.172	0.061	0.188	0.013	0.104	0.029	0.121	
	$\hat{\nu}_{13}^{II}$	0.064	0.264	0.108	0.283	0.027	0.166	0.058	0.179	0.012	0.100	0.027	0.112
	$\bar{\nu}_{13}^{NB}$	0.063	0.233	0.100	0.207	0.030	0.157	0.054	0.143	0.014	0.107	0.029	0.100
$\hat{\nu}_{23}^I$	0.054	0.093	0.302	0.231	0.024	0.047	0.210	0.158	0.011	0.025	0.127	0.114	
	$\hat{\nu}_{23}^{II}$	0.050	0.087	0.300	0.223	0.021	0.042	0.213	0.152	0.009	0.021	0.134	0.109
	$\bar{\nu}_{23}^{NB}$	0.073	0.101	0.340	0.185	0.036	0.056	0.256	0.141	0.018	0.031	0.201	0.107

Table 13
Rejection frequencies of W^I, W^{II} for $\rho = 0.5$, 1000 replications.

	α	n	64				128				256			
			δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
WN	0.01	I	0.033	0.082	0.049	0.088	0.009	0.021	0.024	0.034	0.003	0.011	0.010	0.014
	0.01	II	0.062	0.139	0.108	0.174	0.041	0.071	0.093	0.116	0.039	0.082	0.060	0.099
	0.05	I	0.066	0.146	0.103	0.182	0.023	0.055	0.065	0.100	0.009	0.043	0.026	0.046
	0.05	II	0.124	0.234	0.200	0.274	0.098	0.158	0.172	0.237	0.098	0.158	0.126	0.213
	0.10	I	0.092	0.199	0.153	0.244	0.042	0.091	0.096	0.173	0.021	0.067	0.050	0.094
	0.10	II	0.173	0.294	0.266	0.366	0.146	0.231	0.249	0.324	0.134	0.229	0.204	0.297
VAR	0.01	I	0.003	0.016	0.012	0.020	0.000	0.003	0.004	0.003	0.000	0.000	0.000	0.000
	0.01	II	0.014	0.053	0.038	0.068	0.006	0.018	0.022	0.037	0.003	0.016	0.015	0.019
	0.05	I	0.011	0.047	0.034	0.068	0.002	0.010	0.011	0.017	0.000	0.002	0.000	0.002
	0.05	II	0.035	0.113	0.095	0.156	0.009	0.048	0.050	0.096	0.013	0.051	0.035	0.075
	0.10	I	0.017	0.088	0.067	0.115	0.002	0.014	0.024	0.036	0.000	0.007	0.003	0.008
	0.10	II	0.049	0.161	0.140	0.210	0.016	0.078	0.088	0.170	0.027	0.090	0.063	0.124

Table 14
Rejection frequencies of W^I, W^{II} for $\rho = 0$, 1000 replications.

	α	n	64				128				256			
			δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
WN	0.01	I	0.018	0.040	0.048	0.055	0.007	0.014	0.016	0.019	0.001	0.002	0.004	0.006
	0.01	II	0.061	0.074	0.099	0.118	0.037	0.048	0.059	0.078	0.033	0.041	0.040	0.055
	0.05	I	0.043	0.081	0.094	0.124	0.024	0.026	0.039	0.062	0.010	0.014	0.016	0.022
	0.05	II	0.109	0.149	0.175	0.204	0.075	0.105	0.125	0.167	0.064	0.090	0.105	0.127
	0.10	I	0.080	0.118	0.125	0.176	0.037	0.059	0.068	0.100	0.020	0.029	0.033	0.050
	0.10	II	0.161	0.199	0.238	0.286	0.121	0.155	0.197	0.247	0.115	0.153	0.172	0.192
VAR	0.01	I	0.000	0.008	0.010	0.013	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
	0.01	II	0.011	0.024	0.034	0.045	0.001	0.016	0.013	0.019	0.003	0.005	0.006	0.010
	0.05	I	0.004	0.020	0.029	0.041	0.000	0.008	0.006	0.006	0.000	0.000	0.000	0.001
	0.05	II	0.024	0.058	0.072	0.101	0.010	0.030	0.037	0.054	0.009	0.014	0.017	0.034
	0.10	I	0.011	0.038	0.049	0.074	0.000	0.016	0.019	0.022	0.000	0.001	0.001	0.002
	0.10	II	0.038	0.092	0.107	0.149	0.018	0.049	0.064	0.094	0.020	0.036	0.045	0.072

Table 15

Rejection frequencies of W^I, W^{II} for $\rho = -0.4$, 1000 replications.

	α	n	64				128				256			
			δ_3	2	1.6	1.4	0.8	2	1.6	1.4	0.8	2	1.6	1.4
WN	0.01	I	0.044	0.081	0.089	0.119	0.014	0.034	0.038	0.069	0.004	0.012	0.014	0.031
	0.01	II	0.089	0.137	0.163	0.215	0.057	0.108	0.124	0.175	0.043	0.080	0.096	0.141
	0.05	I	0.080	0.158	0.159	0.220	0.038	0.093	0.086	0.153	0.011	0.034	0.044	0.076
	0.05	II	0.162	0.247	0.261	0.324	0.113	0.208	0.217	0.298	0.094	0.155	0.193	0.254
	0.10	I	0.127	0.217	0.227	0.289	0.056	0.142	0.141	0.221	0.033	0.063	0.075	0.126
	0.10	II	0.218	0.323	0.326	0.407	0.160	0.305	0.272	0.369	0.141	0.233	0.262	0.345
VAR	0.01	I	0.007	0.020	0.032	0.051	0.000	0.008	0.010	0.009	0.000	0.001	0.000	0.000
	0.01	II	0.021	0.058	0.070	0.103	0.008	0.030	0.038	0.061	0.004	0.012	0.029	0.044
	0.05	I	0.014	0.062	0.072	0.109	0.001	0.022	0.020	0.044	0.000	0.003	0.003	0.009
	0.05	II	0.053	0.118	0.138	0.197	0.019	0.078	0.089	0.145	0.014	0.056	0.068	0.114
	0.10	I	0.026	0.096	0.109	0.160	0.002	0.039	0.035	0.075	0.000	0.006	0.006	0.022
	0.10	II	0.080	0.170	0.191	0.264	0.036	0.120	0.130	0.209	0.032	0.085	0.105	0.170

our relatively complex semiparametric setting. Sizes decrease substantially for larger cointegrating gaps and when AR structure is present, due in this latter case to overestimation of integration orders. Note that with the bandwidth choice I, we are close to parametric estimation when $\phi = 0$, so I is definitely not a good choice if we suspect that the short memory structure which characterizes our system is of AR form. In this case, the smaller bandwidth alternative II, as expected, is more appropriate, mitigating the undersizing produced by the AR structure. Finally, the largest and smallest sizes occur under negative and no correlation respectively.

5. Empirical example

Numerous studies have been devoted to study the purchasing power parity (PPP) hypothesis by means of cointegration techniques. In the standard $I(1)/I(0)$ framework this includes Corbae and Ouliaris (1988) and Kim (1990), whereas in the fractional setting Cheung and Lai (1993) analysed a fractional version of the PPP model, which has recently been revisited by Gil-Alana and Hualde (2009), who employed the parametric techniques proposed by RH to discuss the plausibility of the absolute or homogeneous version of the PPP. Basically, these authors justified the fractionally cointegrated bivariate model

$$p_t^* - e_t = a + \nu p_t + v_t, \tag{32}$$

with p_t representing the log of the domestic (US) price index, p_t^* being the log of foreign (UK) price index, e_t the log of the exchange rate, the cointegrating error accounting for deviations in the equilibrium between the two observable variables $p_t^* - e_t$ and p_t . They estimated ν parametrically, and tested by means of a parametric Wald test the plausibility of $\nu = 1$, which represents the absolute version of the PPP, providing evidence in favour of this hypothesis in the US/UK data. We extend their analysis in two different directions. First, we analyze the whole cointegrating structure which characterizes the long run relation among the observable variables in $z_t = (e_t, p_t^*, p_t)'$, and discuss whether the restricted parameterization (32) with $\nu = 1$ fits the data adequately. Second, we employ semiparametric techniques instead of parametric ones, and check whether our alternative procedure leads to different conclusions. By using quarterly US and UK CPI for price indexes and British Pound/US dollar (end of period) exchange rate for the period 1957.Q1–2003.Q4 (188 observations) collected from the IFS database, we infer the cointegrating structure of z_t by means of Hualde’s (2008) procedure, which we briefly detail below. As in Hualde (2008), assuming $z_t \sim I(\delta_3)$, we say that a subvector of z_t has a “reduced order linear combination” (ROLC) if the linear combination is $I(\gamma)$ with $\gamma < \delta_3$.

Thus, for $a_t, b_t = e_t, p_t^*, p_t$, we define the null and alternative hypotheses

$$H_{ab} : (a_t, b_t)' \text{ has no ROLC,}$$

$$\bar{H}_{ab} : H_{ab} \text{ is not true,}$$

and also

$$H_1(1) : H_{pp^*} \cup H_{pe} \cup H_{p^*e}, \tag{33}$$

$$\bar{H}_1(1) : \bar{H}_{pp^*} \cap \bar{H}_{pe} \cap \bar{H}_{p^*e},$$

where by Theorem 2 of Hualde (2008), $r_1 < 2$ if and only if $H_1(1)$ holds, whereas $r_1 = 2$ if and only if $\bar{H}_1(1)$ holds. Note that, as mentioned in Hualde (2008), Theorem 1 of Berger (1982) ensures that if the individual tests H_{ab} in (33) have level α , $H_1(1)$ is also α -level. The implementation of H_{ab} is

$$H_{ab} : \{a_t \sim I(\delta_3)\} \cap \{b_t \sim I(\delta_3)\}$$

$$\cap \{a_t \text{ and } b_t \text{ are not cointegrated}\}. \tag{34}$$

This, requires testing three different hypotheses, so to guarantee that the test for H_{ab} has level α , by Bonferroni’s inequality the individual tests must have level $\alpha/3$ (although as will be shown below, this will be slightly modified). The individual tests in (34) will be Robinson and Yajima (2002) and Robinson (2008) procedures, which are based on Gaussian semiparametric estimation of the integration orders, hence, even though in the Monte Carlo experiment we employed the alternative exact local Whittle method, we carried out the whole analysis below based on the usual local Whittle estimate (see e.g. Robinson, 1995b).

As a preliminary step, we estimate semiparametrically the integration orders of the three observables by local Whittle, using first differences of the series and then adding back 1. For a wide selection of bandwidths the maximum estimated integration order corresponds in all cases to p_t , so we work under the assumption that $p_t \sim I(\delta_3)$. Then we test H_{pp^*} , which, given the assumption that p_t has the maximum integration order, consists of two individual hypotheses ($p_t^* \sim I(\delta_3)$ and p_t and p_t^* are not cointegrated), each of which will be tested with level $\alpha/2$ (to preserve the level α for H_{pp^*}). We test for $p_t^* \sim I(\delta_3)$ by means of Robinson and Yajima’s (2002) statistic \hat{T}_{ab} (comparing the integration orders of p_t and p_t^*), with the trimming sequence $h(n)$ chosen as $b^{-1/(5+2i)}$, $i = 1, \dots, 4$, with b the bandwidth used in the estimation. Choosing as bandwidths all integer values in the interval [55, 65] the hypothesis $p_t^* \sim I(\delta_3)$ is never rejected even at $\alpha = 0.10$. Next, we perform the test for no cointegration of Robinson (2008), which for the same choice of bandwidths the null of no cointegration is always rejected at $\alpha = 0.05$ (and also for 9 out of 11 bandwidths at $\alpha = 0.01$). Thus, there seems to be strong evidence to reject H_{pp^*} . Note that the

Table 16
PPP empirical example: estimates and Wald tests of US/UK data.

(m, b)	n'	183	182	181	180	179	178	177	176	175	174
(6, 40)	$\hat{\nu}_{12}$	-0.709	-0.728	-0.761	-0.823	-0.841	-0.793	-0.774	-0.806	-0.840	-0.788
	$\hat{\nu}_{13}$	0.562	0.602	0.669	0.770	0.808	0.736	0.704	0.762	0.827	0.746
	$\hat{\nu}_{23}$	-1.33	-1.30	-1.24	-1.19	-1.15	-1.12	-1.10	-1.08	-1.05	-1.08
	W	0.343	0.303	0.249	0.150	0.144	0.219	0.253	0.212	0.196	0.269
(6, 85)	$\hat{\nu}_{12}$	-0.613	-0.652	-0.651	-0.693	-0.755	-0.754	-0.766	-0.776	-0.804	-0.779
	$\hat{\nu}_{13}$	0.438	0.482	0.502	0.561	0.662	0.668	0.689	0.706	0.750	0.708
	$\hat{\nu}_{23}$	-1.45	-1.46	-1.47	-1.46	-1.45	-1.46	-1.46	-1.48	-1.46	-1.46
	W	0.669	0.527	0.547	0.411	0.272	0.283	0.269	0.248	0.199	0.241
(10, 40)	$\hat{\nu}_{12}$	-0.860	-0.896	-0.891	-0.985	-1.03	-0.999	-0.987	-1.03	-1.05	-1.02
	$\hat{\nu}_{13}$	0.798	0.861	0.867	1.01	1.09	1.05	1.03	1.10	1.16	1.11
	$\hat{\nu}_{23}$	-1.36	-1.33	-1.28	-1.23	-1.19	-1.16	-1.15	-1.13	-1.11	-1.14
	W	0.117	0.068	0.084	0.025	0.056	0.051	0.047	0.078	0.154	0.141
(10, 85)	$\hat{\nu}_{12}$	-0.841	-0.901	-0.882	-0.947	-1.01	-1.01	-1.02	-1.04	-1.08	-1.06
	$\hat{\nu}_{13}$	0.787	0.863	0.854	0.945	1.04	1.06	1.07	1.11	1.18	1.15
	$\hat{\nu}_{23}$	-1.49	-1.49	-1.51	-1.50	-1.49	-1.49	-1.50	-1.51	-1.51	-1.51
	W	0.172	0.063	0.102	0.027	0.020	0.028	0.039	0.049	0.100	0.082

choice of the relatively large bandwidths is motivated by the poor power results reported by Robinson (2008) for small bandwidths.

Next, H_{pe} is tested, and is strongly rejected because, using again Robinson and Yajima's (2002) procedure, the hypothesis $e_t \sim I(\delta_3)$ is always rejected at $\alpha = 0.01$ for all values of b and $h(n)$. Similarly, H_{p^*e} is also rejected at $\alpha = 0.01$, which implies that $H_1(1)$ is rejected, so there is evidence that $r_1 = 2$, to conclude the first level of the analysis of cointegration.

Next, we search for the possibility of a cointegrating subspace which, following Hualde (2008), amounts to checking the possible cointegration between cointegrating errors arising from the first level analysis. Noting that there is strong statistical evidence that e_t has a different (smaller) order than p_t , e_t could be considered as one of the cointegrating errors, the second arising from the cointegrating relation between p_t and p_t^* . This second cointegrating error, say w_t , is obviously unobservable, but can be approximated by the NBLs residual \hat{w}_t (using bandwidth choices $m = 6, 10$), so we test for

$$H_2(1) \equiv H_{ew} : \{e_t \sim I(\delta_2)\} \cap \{w_t \sim I(\delta_2)\} \\ \cap \{e_t \text{ and } w_t \text{ are not cointegrated}\},$$

where $\delta_2 < \delta_3$, using \hat{w}_t instead of w_t . As before, we commence the analysis by estimating semiparametrically the individual integration orders of e_t and w_t , the second one always being the larger for a wide choice of bandwidths. Thus, the possibility that $w_t \sim I(\delta_2)$ is taken as known, and we proceed with the two individual tests ($e_t \sim I(\delta_2)$ and e_t and w_t are not cointegrated) modifying their respective levels appropriately. By using the previous testing procedures (again for the grid of bandwidths [55, 65], using the two choices $m = 6, 10$), we never reject $e_t \sim I(\delta_2)$ at $\alpha = 0.10$, whereas $\{e_t$ and w_t are not cointegrated $\}$ is always rejected at $\alpha = 0.01$, which implies directly that $H_2(1)$ is rejected, so we conclude that $r_2 = 1$. Thus, the whole cointegrating structure is characterized by the system

$$e_t + \nu_{12}p_t^* + \nu_{13}p_t = \alpha_1 + \Delta^{-\delta_1}u_{1t}^\#, \tag{35}$$

$$p_t^* + \nu_{23}p_t = \alpha_2 + \Delta^{-\delta_2}u_{2t}^\#, \tag{36}$$

$$p_t = \alpha_3 + \Delta^{-\delta_3}u_{3t}^\#, \tag{37}$$

where α_i , $i = 1, 2, 3$, are (possibly nonzero) constants and $\delta_1 < \delta_2 < \delta_3$.

Next we detail the estimation of this system. The first step is to obtain the NBLs estimates of $\nu = (\nu_{12}, \nu_{13}, \nu_{23})'$, which lead to preliminary estimates of $\delta_1, \delta_2, \delta_3$ and f . As before, the estimates of the orders are local Whittle and f is estimated by (31). These estimates of the nuisance parameters lead to our estimate of ν

(noting that due to the filtering applied to the series the presence of $\alpha_i \neq 0$ is eliminated, see Robinson and Iacone, 2005). We report below the NBLs estimates of ν and the local Whittle estimates of the integration orders for the bandwidth combination $m = 6, 10$, and $b = 40, 85$.

1. $m = 6 : \bar{\nu}_{12} = -0.553, \bar{\nu}_{13} = 0.477, \bar{\nu}_{23} = -1.493;$
 - 1a. $b = 40 : \hat{\delta}_3 = 1.779, \hat{\delta}_2 = 1.343, \hat{\delta}_1 = 1.022;$
 - 1b. $b = 85 : \hat{\delta}_3 = 1.566, \hat{\delta}_2 = 1.238, \hat{\delta}_1 = 1.036;$
2. $m = 10 : \bar{\nu}_{12} = -0.584, \bar{\nu}_{13} = 0.525, \bar{\nu}_{23} = -1.491;$
 - 2a. $b = 40 : \hat{\delta}_3 = 1.779, \hat{\delta}_2 = 1.343, \hat{\delta}_1 = 1.021;$
 - 2b. $b = 85 : \hat{\delta}_3 = 1.556, \hat{\delta}_2 = 1.238, \hat{\delta}_1 = 1.036.$

In all cases the signs of the estimates of the cointegrating parameters are consistent with what theory predicts, although they are relatively far from $\nu_{12} = -1, \nu_{13} = 1$, which represents the absolute version of the PPP. These estimates are not very much affected by the choice of m , whereas b has a noticeable effect on the estimates of δ , a small value of b leading to larger estimated cointegrating gaps.

As in Gil-Alana and Hualde (2009), we examine the matter of truncation, which is inherent to the modelling of (35)–(37). This model reflects the time when the data begins, and its misspecification could have an important effect on the degree of filtering applied to the series. This effect is moderated by omitting the possibly badly filtered observations which occur at the beginning of the sample. Thus, we report in Table 16 values of our estimates (for the four different bandwidth combinations) based on the last $n' = n - j$ observations for $j = 5, 6, \dots, 14$, in order to explore sensitivity to starting values. Here, we also report the Wald statistic (W) corresponding to the joint hypothesis $\nu_{12} = -1, \nu_{13} = 1$. Our main findings are as follows. First, the value of b has an important impact on $\hat{\nu}_{23}$, which is close to -1 (when it stabilizes for $n' < 179$) for small b , being substantially larger (in absolute value) when $b = 85$. Alternatively, $\hat{\nu}_{12}, \hat{\nu}_{13}$, are not much affected by b , but are clearly larger in absolute value for larger m . In all cases signs are consistent with theory, with values very close to the absolute version of the PPP for $m = 10$, after the estimates stabilize. Note that if we are willing to assume that e_t has a smaller integration order than p_t or p_t^* , the absolute version of the PPP should be accompanied by the value $\nu_{23} = -1$, so the bandwidth combination $(m, b) = (10, 40)$ is the most favourable to this possibility. However, in any of the four scenarios we presented we never reject this absolute version, which as in Gil-Alana and Hualde (2009) is strongly supported by our data.

6. Final comments

In our semiparametric approach, the choice of bandwidth m is of primary importance. As is typically the case with such asymptotic conditions, (19), (22), (23), (25) and (26) provide no practical guide. Nor, especially in view of the varying convergence rates liable to occur in a multivariate setting, is it likely to be feasible to develop a usable theory of bandwidth choice based on minimizing approximate mean squared error. Rather, as in a number of other complex problems involving smoothing, one might choose to perform the analysis over a grid of m values and assess sensitivity. This can be done without altering the estimation of δ and f . This estimation in turn requires choice of a further bandwidth, and it is the oversmoothing here, relative to m , which is responsible for the property that estimation of δ and f does not affect asymptotic behaviour of the ν estimates. With this in mind, and depending on the choice of m , data-dependent choices of b from the semiparametric integration order estimation, smooth spectral estimation, and autocorrelation-consistent-variance-estimation literatures may be suitable.

It is natural to then ask to what extent there is a parametric counterpart to our results, assuming $f(\lambda) = f(\lambda; \tau)$, $\lambda \in (-\pi, \pi]$, where $f(\lambda; \tau)$ is a known function of λ and an unknown finite-dimensional vector τ (as if, for example, we assume u_t is a VAR of prescribed degree). Given an estimate $\hat{\tau}$ of τ , the $f(\lambda_j; \hat{\tau})$ can be inserted in place of the $\hat{f}(\lambda_j)$ in (9), with $m = [n/2]$. This gives a multivariate extension of one of the ν estimates of RH (the other being an asymptotically-equivalent “time domain” one). In a bivariate setting, under strong cointegration, RH showed precisely the same mixed-normal asymptotics for this estimate as ours, for an arbitrary \sqrt{n} -consistent $\hat{\tau}$, so the present paper has established a kind of adaptivity property. However, under weak cointegration, this parametric estimate of ν is, like ours, asymptotically normal but with a faster, \sqrt{n} , rate, and moreover is not asymptotically independent of $\hat{\tau}$, or of the estimate of δ , which also needs to be \sqrt{n} -consistent. In particular we can obtain optimal estimates of ν , δ and τ by jointly optimizing an approximate form of Gaussian likelihood. It is possible to concentrate out of this only the scale matrix Ω (see Assumption 3), leaving a numerical optimization problem of possibly dauntingly high dimension. With respect to weak cointegration then, our semiparametric estimates of ν are less efficient than optimal ones based on a correctly-specified f , but are likely to be much easier to compute, as well as having the desirable property of providing Wald tests which are valid under both weak and strong cointegration.

We have assumed that the dimensions of the different cointegration subspaces, $r_s, r_{s-1}, \dots, r_2, r_1$, are known. In practice the presence and extent of cointegration has to be determined from the data, a difficult task in itself. As already mentioned, Hualde (2008) proposed an algorithm for choosing these values, which is based on the tests for equality of integration orders and cointegration proposed by Robinson and Yajima (2002), and extended by Nielsen and Shimotsu (2007) to cover nonstationary variables. A serious drawback of these procedures is the need (in addition to a bandwidth like our b) for user-chosen tuning numbers in both steps of the procedure. This is avoided in the algorithm of Robinson (2008), which consists of a sequence of Hausman-type tests using integration order estimates for observables, and is explicitly justified for both stationary and non-stationary observables. Lasak (2005) extended the parametric procedure of Johansen (1991) to fractional series, assuming the (common) integration order of the observables is known and equal to 1, but allowing that of the cointegrating errors (which is also assumed to be constant across all cointegrating relations) to be unknown.

One can include over-identifying restrictions in (8), whence ν represents a vector of unrestricted parameters of dimension less

than q , and C and c are correspondingly re-defined, and used in (9). Though our proofs in the Appendix quickly substitute for C and c their just-identified forms, they can be straightforwardly extended, so long as no constraints link parameters whose estimates have different convergence rates. The same kind of problem arises in other multiparameter problems involving varying convergence rates, such as multiple time series linear regression (see Hannan, 1970, p. 447). Of course this is not an issue when the over-identifying restrictions are all of exclusion type.

Acknowledgements

A simple version of this work formed part of first author’s University of London Ph.D. thesis. The research of the first author was supported by ESRC Grant R000238212 and the Spanish Ministerio de Educación y Ciencia through “Juan de la Cierva” and “Ramón y Cajal” contracts and Ref. no. ECO2008-02641. The research of the second author was supported by ESRC Grants R000239936 and RES-062-23-0036.

Appendix

Proof of Theorem 1. Denoting $\bar{\nu}_m(\delta)$, $b_m(\delta)$, the equivalent of $\hat{\nu}_m(\delta)$, $\hat{b}_m(\delta)$, with f replacing \hat{f} , respectively, clearly

$$\bar{\nu}_m(\delta) - \nu = -b_m^{-1}(\delta)e_m(\delta),$$

where

$$e_m(\delta) = C' \sum_{j=0}^m \text{Re} \{s_j w_{z^d}(-\lambda_j) f^{-1}(\lambda_j) w_u(\lambda_j)\}.$$

With C as defined it is straightforward to show that

$$b_m(\delta) = \sum_{j=0}^m s_j \text{Re} \{F(\lambda_j) \circ I_{z(\delta)}(-\lambda_j)\}, \tag{38}$$

where $z_t(\delta) = (1_{1,r_s} \otimes z'_{-r_s,t}(\delta_{r_s}), 1_{1,r_{s-1}-r_s} \otimes z'_{-r_{s-1},t}(\delta_{r_{s-1}}), \dots, 1_{1,r_1-r_2} \otimes z'_{-r_1,t}(\delta_{r_1}))'$. Also

$$e_m(\delta) = \sum_{j=0}^m s_j \text{Re} \{w_{z(\delta)}(-\lambda_j) \circ P(\lambda_j) w_u(\lambda_j)\}. \tag{39}$$

To avoid the problem of cointegrated regressors, we will deal with $Tb_m(\delta)T'$ and $Te_m(\delta)$, where the only change with respect to (38), (39), is that the process $z_t(\delta)$ is replaced by $w_t(\delta)$, where

$$w_t(\delta) = \left(1_{1,r_s} \otimes w'_{-r_s,t}(\delta_{r_s}), 1_{1,r_{s-1}-r_s} \otimes w'_{-r_{s-1},t}(\delta_{r_{s-1}}), \dots, 1_{1,r_1-r_2} \otimes w'_{-r_1,t}(\delta_{r_1}) \right)'$$

noting (4), so $w_t(\delta)$ is composed of s subvectors $w_t(\delta) = (w_t^{(s)'}(\delta), \dots, w_t^{(1)'}(\delta))'$, noting that for $l = 1, \dots, s$, an arbitrary element of $w_t^{(l)}(\delta)$ will be $w_{kt}(\delta_l)$, where $k \in \{r_l + 1, \dots, r_0\}$ and we use the simplifying notation $\delta_l = \delta_{r_l}$. Note also that $w_{kt}(\delta_l)$ is $I(\delta_k - \delta_l)$, where for fixed $j = 0, \dots, l - 1$, if $k \in \{r_{j+1} + 1, \dots, r_j\}$, then $\delta_k = \delta_{r_j}$, so in all cases $\delta_k - \delta_l > 0$ in view of Assumption 1.

Parts (i) and (ii) of the theorem are justified on showing

$$\begin{aligned} & A_{(n)}^S (T')^{-1} (\bar{\nu}_m(\delta) - \nu) \\ & \Rightarrow \left(F(0) \circ D \int_0^1 \tilde{W}(w; \delta) \tilde{W}'(w; \delta) dw D' \right)^{-1} \\ & \quad \times \int_0^1 D \tilde{W}(w; \delta) \circ d\bar{W}(w), \end{aligned} \tag{40}$$

under (i), and

$$A_{(n)}^W(T')^{-1}(\bar{v}_m(\delta) - v) \rightarrow_d N(0, A^{-1}), \quad \text{under (ii),} \quad (41)$$

and

$$A_{(n)}^L(\bar{v}_m(\delta) - \widehat{v}_m(\delta)) = o_p(1), \quad (42)$$

$$A_{(n)}^L(\widehat{v}_m(\widehat{\delta}) - \widehat{v}_m(\delta)) = o_p(1), \quad (43)$$

where $L = S$ under (i) and $L = W$ under (ii). First we show (40), dealing initially with the convergence of the elements of $Te_m(\delta)$, and showing first that

$$(A_{(n)}^S)^{-1}Te_m(\delta) = (A_{(n)}^S)^{-1}T\bar{e}_m(\delta) + o_p(1), \quad (44)$$

where

$$T\bar{e}_m(\delta) = \frac{1}{2\pi} \sum_{t=2}^n w_{t-1}(\delta) \circ P(0)A(1)\varepsilon_t.$$

We show that an arbitrary element of $Te_m(\delta)$ has expectation of smaller order. More precisely we show that

$$E \left(\sum_{j=0}^m s_j \text{Re} \left\{ p_a(\lambda_j) I_{UW_k(\delta_l)}(\lambda_j) \right\} \right) = o(n^{\delta_k - \delta_l}), \quad (45)$$

where noting the previous discussion, $a \in \{r_{l+1} + 1, \dots, r_l\}$. We can write the left side of (45) as the real part of

$$\begin{aligned} & \frac{1}{2\pi n} \sum_{j=0}^m s_j p_a(\lambda_j) \int_{-\pi}^{\pi} D_n(\lambda_j - \mu) \\ & \times \sum_{t=1}^n e^{-i(n-t)\lambda_j} D_t(\mu - \lambda_j) f(\mu) a_{n-t}(\delta_k - \delta_l) i_k d\mu, \end{aligned} \quad (46)$$

where $D_t(\lambda) = \sum_{k=1}^t e^{ik\lambda}$ is the Dirichlet kernel, where for $0 < \lambda < \pi$,

$$|D_t(\lambda)| < K \min \{ |\lambda|^{-1}, t \}. \quad (47)$$

Noting that for any $\lambda, k > a$,

$$p_a(\lambda) f(\lambda) i_k = 0, \quad (48)$$

by periodicity we can write (46) as

$$\begin{aligned} & \frac{1}{2\pi n} \sum_{j=0}^m s_j p_a(\lambda_j) \int_{-\pi}^{\pi} D_n(-\mu) \\ & \times \sum_{t=0}^{n-1} e^{-it\lambda_j} D_{n-t}(\mu) [f(\mu + \lambda_j) - f(\lambda_j)] a_t i_k d\mu, \end{aligned} \quad (49)$$

where $a_t = a_t(\delta_k - \delta_l)$. Next, by summation by parts, (49) is

$$\begin{aligned} & \frac{1}{2\pi n} \sum_{j=0}^m s_j p_a(\lambda_j) \int_{-\pi}^{\pi} D_n(-\mu) \\ & \times \left\{ D_1(\mu) [f(\mu + \lambda_j) - f(\lambda_j)] a_{n-1} i_k \sum_{t=0}^{n-1} e^{-it\lambda_j} d\mu \right. \\ & - [f(\mu + \lambda_j) - f(\lambda_j)] \sum_{t=0}^{n-2} (a_{t+1} D_{n-t-1}(\mu) \\ & \left. - a_t D_{n-t}(\mu)) i_k \sum_{h=0}^t e^{-ih\lambda_j} d\mu \right\}. \end{aligned} \quad (50)$$

Because

$$\sum_{t=0}^{n-1} e^{-it\lambda_j} = n, \quad j = 0, \text{ mod } n; = 0, \text{ otherwise,} \quad (51)$$

the contribution of the first term in braces in (50) is bounded in modulus by

$$K |a_{n-1}| \int_{-\pi}^{\pi} |D_n(\mu)| d\mu = O(n^{\delta_k - \delta_l - 1} \log n),$$

since f is boundedly differentiable, $|a_s(c)| \leq K(1 + s)^{c-1}$, for $c > 0, s \geq 0$ by Stirling's approximation, and

$$\int_{-\pi}^{\pi} |D_n(\mu)| d\mu = O(\log n), \quad (52)$$

(see e.g. Zygmund, 1977). Regarding the second term in (50), note that

$$\begin{aligned} & a_{t+1} D_{n-t-1}(\mu) - a_t D_{n-t}(\mu) \\ & = (a_{t+1} - a_t) D_{n-t-1}(\mu) - e^{i(n-t)\mu} a_t. \end{aligned} \quad (53)$$

When $\delta_k - \delta_l = 1$, the contribution of the first term on the right of (53) to the second term of (50) is null, because $a_{t+1}(1) = a_t(1), t = 0, \dots, n - 2$. For $\delta_k - \delta_l \neq 1$, this contribution is bounded in modulus by

$$\begin{aligned} & K n^{-1} \left\{ \sum_{j=0}^m \int_{-\pi}^{\pi} |D_n(\mu)|^2 \|f(\mu + \lambda_j) - f(\lambda_j)\| d\mu \right\}^{\frac{1}{2}} \\ & \times \left\{ \sum_{j=0}^m \int_{-\pi}^{\pi} \left| \sum_{t=0}^{n-2} (a_{t+1} - a_t) D_{n-t-1}(\mu) (D_t(-\lambda_j) + 1) \right|^2 \right. \\ & \left. \times \|f(\mu + \lambda_j) - f(\lambda_j)\| d\mu \right\}^{\frac{1}{2}}. \end{aligned} \quad (54)$$

The term in the first braces is bounded by

$$K m \int_{-\pi}^{\pi} |\mu| |D_n(\mu)|^2 d\mu = O(m \log n),$$

by (47) and (52), since f is boundedly differentiable. The term in the second braces is bounded by

$$\begin{aligned} & K \sum_{j=0}^m \int_{-\pi}^{\pi} |\mu| \sum_{t=0}^{n-2} \sum_{s=0}^{n-2} (a_{t+1} - a_t) D_{n-t-1}(\mu) \\ & \times (D_t(-\lambda_j) + 1) (a_{s+1} - a_s) D_{n-s-1}(-\mu) (D_s(\lambda_j) + 1) d\mu \\ & = O \left(n^2 \log n \sum_{j=1}^m j^{-2} \left(\sum_{t=1}^n t^{\delta_k - \delta_l - 2} \right)^2 \right), \end{aligned}$$

by Lemma C.1 of RH and (47), which is $O(n^2 \log n)$ if $\delta_k - \delta_l < 1$, or $O(n^{2(\delta_k - \delta_l)} \log n)$ if $\delta_k - \delta_l > 1$, implying that (54) is $O(m^{1/2} \log n)$ if $\delta_k - \delta_l < 1$, or $O(n^{\delta_k - \delta_l - 1} m^{1/2} \log n)$ if $\delta_k - \delta_l > 1$. Finally, the contribution of the second term on the right of (53) to the second term of (50) is bounded in modulus by

$$\begin{aligned} & K n^{-1} \sum_{j=0}^m \left\{ \int_{-\pi}^{\pi} |\mu D_n(\mu)|^2 d\mu \right. \\ & \left. \times \int_{-\pi}^{\pi} \left| \sum_{t=0}^{n-2} e^{i(n-t)\mu} a_t (D_t(-\lambda_j) + 1) \right|^2 d\mu \right\}^{\frac{1}{2}}. \end{aligned} \quad (55)$$

The first integral is $O(1)$ by (47), whereas the second is bounded by $K \sum_{t=1}^n a_t^2 |D_t(\lambda_j)|^2$, since

$$\int_{-\pi}^{\pi} e^{-i(s-t)\mu} d\mu = 2\pi, \quad s = t; = 0, \text{ otherwise.}$$

Thus (55) is bounded by $Kn^{-1} \sum_{j=1}^m \{n^{2(\delta_k - \delta_l) + 1} j^{-2}\}^{1/2} = O(n^{\delta_k - \delta_l - 1/2} \log m)$, to conclude the proof of (45).

Next, in view of Propositions 1, 2 of RH,

$$\sum_{j=0}^m \operatorname{Re} \{s_j p_a(\lambda_j) (I_{u w_k(\delta_l)}(\lambda_j) - E I_{u w_k(\delta_l)}(\lambda_j))\} = \frac{p_a(0)}{2\pi} \sum_{t=1}^n w_{k,t-1}(\delta_l) A(1) \varepsilon_t + o_p(n^{\delta_k - \delta_l})$$

holds on showing $\operatorname{Var}\{\operatorname{Re}\{\sum_{j=m+1}^{\lfloor n/2 \rfloor} s_j p_a(\lambda_j) I_{u w_k(\delta_l)}(\lambda_j)\}\} = o(n^{2(\delta_k - \delta_l)})$, but, as mentioned in Robinson and Marinucci (2001), this follows by a simple modification of their Theorem 5.1, as $p_a(\lambda)$ is a well-behaved function without poles. This completes the proof of (44).

Next, $w_t(\delta) = D \bar{w}_t(\delta)$, where $\bar{w}_t(\delta) = (w'_{-r_s, t}(\delta_{r_s}), w'_{-r_{s-1}, t}(\delta_{r_{s-1}}), \dots, w'_{-r_1, t}(\delta_{r_1}))'$, and set $\bar{c}(v) = (\bar{c}'_1(v), \bar{c}'_2(v))'$, where

$$\begin{aligned} \bar{c}_1(v) &= n^{-\frac{1}{2}} (\bar{\Lambda}_{(n)}^S)^{-1} \\ &\quad \times \left(w'_{-r_s, [nv]}(\delta_{r_s}), w'_{-r_{s-1}, [nv]}(\delta_{r_{s-1}}), \dots, w'_{-r_1, [nv]}(\delta_{r_1}) \right)', \\ \bar{c}_2(v) &= n^{-\frac{1}{2}} \sum_{t=1}^{\lfloor nv \rfloor} P(0) A(1) \varepsilon_t, \end{aligned}$$

where $\bar{\Lambda}_{(n)}^S$ is the $(sr_0 - \sum_{j=1}^s r_j) \times (sr_0 - \sum_{j=1}^s r_j)$ block diagonal matrix with j th block given by $\Lambda_{r_s - j + 1}^S, j = 1, \dots, s$. By Marinucci and Robinson (2000) and a standard functional central limit theorem for martingales (see e.g. Brown, 1971)

$$\bar{c}(v) \Rightarrow (\tilde{W}'(v; \delta), \bar{W}'(v))', \tag{56}$$

to conclude that

$$(\Lambda_{(n)}^S)^{-1} T e_m(\delta) \Rightarrow \int_0^1 D \tilde{W}(v; \delta) \circ d \bar{W}(v),$$

by Theorem 2.2 of Kurtz and Protter (1991).

Next, noting that by Theorem 4.4 and a simple modification of Theorem 5.1 of Robinson and Marinucci (2001) and Assumption 3

$$(\Lambda_{(n)}^S)^{-1} \sum_{j=m+1}^{\lfloor n/2 \rfloor} s_j \operatorname{Re} \{F(\lambda_j) \circ I_{w(\delta)}(-\lambda_j)\} (\Lambda_{(n)}^S)^{-1} = o_p(1),$$

then

$$\begin{aligned} (\Lambda_{(n)}^S)^{-1} T b_m(\delta) T' (\Lambda_{(n)}^S)^{-1} \\ \Rightarrow F(0) \circ D \int_0^1 \tilde{W}(v; \delta) \tilde{W}'(v; \delta) dv D', \end{aligned}$$

by (56) and a simple multivariate extension of Propositions 4, 5, 6 of RH, to conclude the proof of (40).

Next, we show (41), for which we initially justify that

$$(\Lambda_{(n)}^W)^{-1} T e_m(\delta) = R_m + o_p(1), \tag{57}$$

where $R_m = (R_m^{(s)'}, \dots, R_m^{(1)'})'$, with $R_m^{(l)} = (R_m^{(l, r_{l+1}+1)'}, \dots, R_m^{(l, r_l)'})'$ for $l = 1, \dots, s, R_m^{(l, a)} = (R_m^{(l, a, r_{l+1})}, \dots, R_m^{(l, a, r_0)})', a \in \{r_{l+1} + 1, \dots, r_l\}$, where

$$R_m^{(l, a, k)} = \sum_{t=2}^n \varepsilon_t' \sum_{v=1}^{t-1} c_{t-v}^{(l, a, k)} \varepsilon_v,$$

for $k \in \{r_l + 1, \dots, r_0\}$, and

$$c_t^{(l, a, k)} = \frac{\lambda_m^{\delta_k - \delta_l}}{2\pi n m^{1/2}} \sum_{j=1}^m \operatorname{Re} (E^{(l, a, k)}(\lambda_j) e^{it\lambda_j}),$$

where $E^{(l, a, k)}(\lambda_j)$ is a Hermitian matrix given by

$$\begin{aligned} E^{(l, a, k)}(\lambda_j) &= (1 - e^{-i\lambda})^{-(\delta_k - \delta_l)} A' (e^{i\lambda}) p'_a(\lambda) i'_k A (e^{-i\lambda}) \\ &\quad + (1 - e^{i\lambda})^{-(\delta_k - \delta_l)} A' (e^{i\lambda}) i_k p_a(-\lambda) A (e^{-i\lambda}), \end{aligned}$$

using again the notation $\delta_l = \delta_{r_l}$, and noting that for any fixed $j = 0, \dots, l-1$, if $k \in \{r_{j+1} + 1, \dots, r_j\}$, $\delta_k = \delta_{r_j}$.

First, define

$$\tilde{w}_{kt}(\delta_l) = \Delta^{\delta_l - \delta_k} u_{kt},$$

which is the covariance stationary version of $w_{kt}(\delta_l)$, noting the absence of truncation in the right hand side. We show first that

$$\sum_{j=0}^m \operatorname{Re} \{s_j p_a(\lambda_j) I_{u w_k(\delta_l)}(\lambda_j)\} = 2 \sum_{j=1}^m \operatorname{Re} \{p_a(\lambda_j) I_{u \tilde{w}_k(\delta_l)}(\lambda_j)\} + o_p(n^{\delta_k - \delta_l} m^{\frac{1}{2} - (\delta_k - \delta_l)}). \tag{58}$$

The left side of (58) is

$$\begin{aligned} 2 \sum_{j=1}^m \operatorname{Re} \{p_a(\lambda_j) I_{u \tilde{w}_k(\delta_l)}(\lambda_j)\} \\ + p_a(0) I_{u \tilde{w}_k(\delta_l)}(0) + p_a(0) (I_{u w_k(\delta_l)}(0) - I_{u \tilde{w}_k(\delta_l)}(0)) \\ + 2 \sum_{j=1}^m \operatorname{Re} \{p_a(\lambda_j) (I_{u w_k(\delta_l)}(\lambda_j) - I_{u \tilde{w}_k(\delta_l)}(\lambda_j))\}. \end{aligned} \tag{59}$$

Clearly, the second term in (59) is $O_p(n^{\delta_k - \delta_l}) = o_p(n^{\delta_k - \delta_l} m^{1/2 - (\delta_k - \delta_l)})$, because under Assumption 3, $\sum_{t=1}^n u_t = O_p(n^{1/2})$, $\sum_{t=1}^n \tilde{w}_{kt}(\delta_l) = O_p(n^{1/2 + \delta_k - \delta_l})$ (see e.g. Robinson, 1994b). The third term in (59) is

$$\frac{p_a(0)}{2\pi n} \sum_{t=1}^n u_t \sum_{s=1}^n (w_{kt}(\delta_l) - \tilde{w}_{kt}(\delta_l)), \tag{60}$$

where the second summation in (60) has mean 0 and variance

$$i'_k \sum_{s=1}^n \sum_{t=1}^n \sum_{r=0}^{\infty} \sum_{v=0}^{\infty} \int_{-\pi}^{\pi} a_{s+r} f(\mu) e^{i(r-v)\mu} a_{t+v} i_k d\mu. \tag{61}$$

Clearly, (61) is bounded by

$$\begin{aligned} K \int_{-\pi}^{\pi} \left| \sum_{s=1}^n \sum_{r=0}^{\infty} a_{s+r} e^{ir\mu} \right|^2 d\mu \\ \leq K \sum_{t=1}^n \sum_{s=1}^n \sum_{r=0}^{\infty} (t+r)^{\delta_k - \delta_l - 1} (s+r)^{\delta_k - \delta_l - 1} \\ \leq K \sum_{t=1}^n \sum_{r=0}^{\infty} (t+r)^{2(\delta_k - \delta_l) - 2} \\ + K \sum_{t=2}^n \sum_{s=1}^{t-1} \sum_{r=0}^{\infty} (t+r)^{\delta_k - \delta_l - 1} (s+r)^{\delta_k - \delta_l - 1} \\ \leq K \sum_{t=1}^n \sum_{r=t}^{\infty} r^{2(\delta_k - \delta_l) - 2} + K \sum_{t=2}^n \sum_{s=1}^{t-1} \sum_{r=0}^{\infty} (s+r)^{2(\delta_k - \delta_l) - 2} \\ \leq Kn^{2(\delta_k - \delta_l) + 1}. \end{aligned}$$

Thus $\sum_{s=1}^n (w_{ks}(\delta_l) - \tilde{w}_{ks}(\delta_l)) = O_p(n^{(\delta_k - \delta_l) + 1/2})$, with the same conclusion as for the second term. The expectation of the fourth term is the real part of

$$\begin{aligned} \frac{1}{\pi n} \int_{-\pi}^{\pi} \sum_{j=1}^m p_a(\lambda_j) D_n(\lambda_j - \mu) \\ \times \sum_{s=0}^{\infty} \sum_{r=1}^n e^{-ir\lambda_j} f(\mu) a_{r+s} i_k e^{-is\mu} d\mu, \end{aligned}$$

which by (48) and periodicity has modulus

$$\begin{aligned} & \left| \frac{1}{\pi n} \int_{-\pi}^{\pi} \sum_{j=1}^m p_a(\lambda_j) D_n(-\mu) \sum_{s=0}^{\infty} \sum_{r=s+1}^{n+s} e^{-ir\lambda_j} \right. \\ & \quad \times \left. (f(\mu + \lambda_j) - f(\lambda_j)) a_r i_k e^{-is(\mu-\lambda_j)} d\mu \right| \\ & \leq \frac{K}{n} \left\{ \int_{-\pi}^{\pi} \sum_{j=1}^m |D_n(-\mu)| \|f(\mu + \lambda_j) - f(\lambda_j)\|^2 d\mu \right. \\ & \quad \times \left. \int_{-\pi}^{\pi} \sum_{j=1}^m \left| \sum_{s=0}^{\infty} \sum_{r=s+1}^{n+s} a_r e^{-ir\lambda_j} e^{-is(\theta-\lambda_j)} \right|^2 d\theta \right\}^{\frac{1}{2}}. \end{aligned}$$

By Assumption 3 and (47) the first integral is $O(m)$. The second integral is

$$\begin{aligned} & 2\pi \sum_{j=1}^m \sum_{s=0}^{\infty} \sum_{k=s+1}^{n+s} \sum_{r=s+1}^{n+s} a_k a_r e^{i(r-k)\lambda_j} \\ & \leq K \sum_{j=1}^m \sum_{s=0}^{\infty} \frac{(s+1)^{2(\delta_k-\delta_l)-2}}{|\lambda_j|^2} = O(n^2), \end{aligned}$$

by Lemma 3.2 of Robinson and Marinucci (2001), to conclude that the expectation of the fourth term of (59) is $O(m^{1/2})$. The variance of the fourth term in (59) is bounded by the real part of

$$\begin{aligned} & \frac{1}{\pi^2 n^2} \sum_{j=1}^m \sum_{v=1}^m \sum_{t=1}^n \sum_{r=1}^n \sum_{s=1}^n \sum_{q=1}^n \sum_{w=0}^{\infty} \sum_{p=0}^{\infty} e^{i\lambda_j(t-s) - i\lambda_v(r-q)} \\ & \quad \times \left[p_a(\lambda_j) E(u_t u'_r) p'_a(-\lambda_v) i'_k a_{s+w} E(u_{-w} u'_{-p}) a_{q+p} i_k + p_a(\lambda_j) \right. \\ & \quad \times \left. E(u_t u'_{-p}) a_{q+p} i_k i'_k a_{s+w} E(u_{-w} u'_r) p'_a(-\lambda_v) + \kappa \right], \end{aligned} \tag{62}$$

where κ is the fourth cumulant matrix of $p_a(\lambda_j)u_t$, $p_a(-\lambda_v)u_r$, $i'_k a_{s+w}u_{-w}$, $i'_k a_{q+p}u_{-p}$. We just detail the contribution to the variance of the first term in braces in (62). By simple application of the Cauchy inequality the second and third terms have the same order as the first term. This contribution is bounded by

$$\begin{aligned} & Kn^{-2} \left| \sum_{j=1}^m \sum_{k=1}^m \sum_{v=0}^{\infty} \sum_{s=1}^n a_{s+v} e^{-i\lambda_j s} \sum_{q=1}^n a_{q+v} e^{i\lambda_k q} \sum_{t=1}^n e^{it(\lambda_j-\lambda_k)} \right| \\ & \leq Kn^{-1} \left| \sum_{j=1}^m \sum_{v=0}^{\infty} \sum_{s=1}^n a_{s+v} e^{-i\lambda_j s} \sum_{q=1}^n a_{q+v} e^{i\lambda_j q} \right|, \end{aligned} \tag{63}$$

by (51), while (63) is bounded by

$$\begin{aligned} & Kn^{-1} m \sum_{v=0}^{\infty} \sum_{s=1}^n a_{s+v}^2 + Kn^{-1} \\ & \quad \times \left| \sum_{j=1}^m \sum_{v=0}^{\infty} \sum_{s \neq q}^n a_{s+v} a_{q+v} e^{i\lambda_j(q-s)} \right|. \end{aligned} \tag{64}$$

Clearly, the first term in (64) is $O(mn^{2(\delta_k-\delta_l)-1})$, and by (47) the second is bounded by

$$\begin{aligned} & Kn^{-1} \sum_{v=0}^{\infty} \sum_{s \neq q}^n \sum_{q=2}^{q-1} |a_{s+v} a_{q+v}| \frac{1}{|\lambda_{q-s}|} \\ & \leq K \sum_{v=0}^{\infty} \sum_{q=2}^n \sum_{s=1}^{q-1} \frac{(s+v)^{\delta_k-\delta_l-1} (q+v)^{\delta_k-\delta_l-1}}{q-s} \end{aligned}$$

$$\begin{aligned} & \leq K \sum_{q=2}^n \sum_{s=1}^{q-1} \frac{1}{q-s} \sum_{v=s}^{\infty} v^{2(\delta_k-\delta_l)-2} \leq K \sum_{q=2}^n \sum_{s=1}^{q-1} \frac{s^{2(\delta_k-\delta_l)-1}}{q-s} \\ & = K \sum_{q=1}^{n-1} q^{-1} \sum_{s=1}^{n-q} s^{2(\delta_k-\delta_l)-1} \leq Kn^{2(\delta_k-\delta_l)} \log n. \end{aligned}$$

Thus, the fourth term in (59) is $O_p(m^{1/2} + n^{\delta_k-\delta_l} \log^{1/2} n)$, which is $o_p(n^{\delta_k-\delta_l} m^{1/2-(\delta_k-\delta_l)})$, by (25), to conclude the proof of (58). Now (57) follows straightforwardly. Next, defining $R_m(v)$ as R_m but with

$$R_m^{(l,a,k)}(v) = \sum_{t=2}^{[nv]} \varepsilon_t' \sum_{v=1}^{t-1} c_{t-v}^{(l,a,k)} \varepsilon_v,$$

where $[\cdot]$ denotes integer part, by a standard functional central limit theorem for martingales (see e.g. Brown, 1971), $R_m(v) \Rightarrow \tilde{W}(v)$, where $\tilde{W}(v)$ is a Brownian motion with covariance matrix given by A . This result is straightforwardly obtained following the steps of Lobato (1999), Christensen and Nielsen (2006) and Nielsen and Frederiksen (2008), noting that some cancellations take place due to (48). Thus, we conclude

$$(\Lambda_{(n)}^W)^{-1} T e_m(\delta) \rightarrow_d N(0, A).$$

Finally, the proof of (41) is completed on showing

$$(\Lambda_{(n)}^W)^{-1} T b_m(\delta) T' (\Lambda_{(n)}^W)^{-1} \rightarrow_p A. \tag{65}$$

We omit most details because the proof is similar to that for $T e_m(\delta)$. First, it is possible to show that $T b_m(\delta) T'$ can be approximated by

$$2 \sum_{j=1}^m \text{Re} \{ F(\lambda_j) \circ I_{\tilde{w}(\delta)}(-\lambda_j) \},$$

where $\tilde{w}_t(\delta)$ is the covariance stationary version of $w_t(\delta)$, by following similar steps to those in the proof of (58). Then (65) follows straightforwardly, noting again the cancellations due to (48).

Next, we show (42), which follows on showing

$$(\Lambda_{(n)}^L)^{-1} T (\hat{e}_m(\delta) - e_m(\delta)) = o_p(1), \tag{66}$$

$$(\Lambda_{(n)}^L)^{-1} T (\hat{b}_m(\delta) - b_m(\delta)) T' (\Lambda_{(n)}^L)^{-1} = o_p(1), \tag{67}$$

where $L = S$ under (i) and $L = W$ under (ii). We just prove (66), the proof for (67) being significantly simpler. Taking again an arbitrary element, this result holds if

$$\begin{aligned} & \sum_{j=0}^m \text{Re} \{ s_j (\hat{p}_a(\lambda_j) - p_a(\lambda_j)) I_{uw_k(\delta_l)}(\lambda_j) \} \\ & = o_p(n^{\delta_k-\delta_l} m^{1/2-\min\{1/2, \delta_k-\delta_l\}}). \end{aligned} \tag{68}$$

Noting that

$$\hat{p}_a(\lambda) - p_a(\lambda) = i'_a f(\lambda)^{-1} [f(\lambda) - \hat{f}(\lambda)] \hat{f}(\lambda)^{-1},$$

the two possible terms for which $s_j = 1$ are $O_p(n^{\delta_k-\delta_l-x}) = o_p(n^{\delta_k-\delta_l} m^{1/2-\min\{1/2, \delta_k-\delta_l\}})$ by (22), because by Assumption 3 $\sum_{t=1}^n u_t = O_p(n^{1/2})$, and by results in Robinson and Marinucci (2001) and previous arguments, $\sum_{t=1}^n w_{kt}(\delta_l) = O_p(n^{\delta_k-\delta_l+1/2})$. By summation by parts, the remaining terms in (68) equal

$$\begin{aligned} & 2\text{Re} \left\{ (\hat{p}_a(\lambda_{m^*}) - p_a(\lambda_{m^*})) \sum_{j=1}^{m^*} I_{uw_k(\delta_l)}(\lambda_j) - \sum_{j=1}^{m^*-1} (\hat{p}_a(\lambda_{j+1}) \right. \\ & \quad \left. - p_a(\lambda_{j+1}) - (\hat{p}_a(\lambda_j) - p_a(\lambda_j))) \sum_{h=1}^j I_{uw_k(\delta_l)}(\lambda_h) \right\}, \end{aligned} \tag{69}$$

where $m^* = m - 1$ if $m = n/2$ or $m^* = m$, otherwise. Using techniques in Robinson and Marinucci (2003) it is not difficult to show that

$$\sum_{h=1}^j I_{uw_k(\delta_l)}(\lambda_h) = O_p \left(n^{\delta_k - \delta_l} m^{1 - (\delta_k - \delta_l)} 1(\delta_k - \delta_l < 1) + n^{\delta_k - \delta_l} 1(\delta_k - \delta_l \geq 1) \right), \tag{70}$$

uniformly in $j \in \{1, \dots, m\}$. Thus, by Assumption 5, the first term of (69) is

$$O_p \left(n^{\delta_k - \delta_l - \kappa} m^{1 - (\delta_k - \delta_l)} 1(\delta_k - \delta_l < 1) + n^{\delta_k - \delta_l - \kappa} 1(\delta_k - \delta_l \geq 1) \right),$$

which is $O_p(n^{\delta_k - \delta_l} m^{1/2 - \min\{1/2, \delta_k - \delta_l\}})$ noting (22). Similarly, by (70) and Assumptions 3 and 5, the second term of (69) is

$$O_p \left((n^{\delta_k - \delta_l - 1 - \kappa} + n^{\delta_k - \delta_l - \phi}) m^{2 - (\delta_k - \delta_l)} \right), \quad \text{if } \delta_k - \delta_l < 1, \\ O_p \left((n^{\delta_k - \delta_l - 1 - \kappa} + n^{\delta_k - \delta_l - \phi}) m \right), \quad \text{if } \delta_k - \delta_l \geq 1,$$

which is $O_p(n^{\delta_k - \delta_l} m^{1/2 - \min\{1/2, \delta_k - \delta_l\}})$ by (23), to conclude the proof of (66).

Finally, noting that $\widehat{v}_m(\widehat{\delta}) - \nu = -\widehat{b}_m^{-1}(\widehat{\delta})\widehat{e}_m(\widehat{\delta})$, where

$$\widehat{e}_m(\widehat{\delta}) = \sum_{j=0}^m s_j \text{Re} \{ w_{z(\widehat{\delta})}(-\lambda_j) \circ \widehat{P}(\lambda_j) w_{\widehat{u}}(\lambda_j) \},$$

where $z_t(\widehat{\delta}) = (1_{1,r_5} \otimes z'_{-r_5,t}(\widehat{\delta}_{r_5}), 1_{1,r_5-1-r_5} \otimes z'_{-r_5-1,t}(\widehat{\delta}_{r_5-1}), \dots, 1_{1,r_1-r_2} \otimes z'_{-r_1,t}(\widehat{\delta}_{r_1}))'$, $\widehat{P}(\lambda)$ is as $P(\lambda)$, replacing f by \widehat{f} , and $\widehat{u}_t = \Delta(\widehat{\delta})\Delta^{-1}(\delta)u_t^\#$. Then, (43) follows on establishing

$$(\Lambda_{(n)}^L)^{-1} T(e_m(\widehat{\delta}) - e_m(\delta)) = o_p(1), \tag{71}$$

$$(\Lambda_{(n)}^L)^{-1} T(\widehat{e}_m(\widehat{\delta}) - e_m(\widehat{\delta}) - \widehat{e}_m(\delta) + e_m(\delta)) = o_p(1), \tag{72}$$

$$(\Lambda_{(n)}^L)^{-1} T(b_m(\widehat{\delta}) - b_m(\delta))T'(\Lambda_{(n)}^L)^{-1} = o_p(1), \tag{73}$$

$$(\Lambda_{(n)}^L)^{-1} T(\widehat{b}_m(\widehat{\delta}) - b_m(\widehat{\delta}) - \widehat{b}_m(\delta) + b_m(\delta))T'(\Lambda_{(n)}^L)^{-1} = o_p(1), \tag{74}$$

where $e_m(\widehat{\delta})$ is like $\widehat{e}_m(\widehat{\delta})$ but with $f(\lambda)$ replacing $\widehat{f}(\lambda)$. We just prove (71), (72), the proofs for (73), (74) being similar but simpler.

The left side of an arbitrary element of (71) is the real part of

$$\sum_{j=0}^m s_j p_a(\lambda_j) [w_{w_k(\widehat{\delta}_l)}(-\lambda_j) - w_{w_k(\delta_l)}(-\lambda_j)] \\ \times [w_{\widehat{u}}(\lambda_j) - w_u(\lambda_j)] + \sum_{j=0}^m s_j p_a(\lambda_j) w_{w_k(\delta_l)}(-\lambda_j) \\ \times [w_{\widehat{u}}(\lambda_j) - w_u(\lambda_j)] + \sum_{j=0}^m s_j p_a(\lambda_j) \\ \times [w_{w_k(\widehat{\delta}_l)}(-\lambda_j) - w_{w_k(\delta_l)}(-\lambda_j)] w_u(\lambda_j). \tag{75}$$

We just consider the third term of (75), as, following similar techniques to those of RH, one could easily show that the same order of magnitude obtained applies also to the whole of (75). By Taylor's theorem, the third term of (75) is the real part of

$$\sum_{r=1}^{R-1} \frac{(\delta_l - \widehat{\delta}_l)^r}{r!} \sum_{j=0}^m s_j p_a(\lambda_j) w_{w_k(\delta_l)}^{(r)}(-\lambda_j) w_u(\lambda_j) \\ + \frac{(\delta_l - \widehat{\delta}_l)^R}{R!} \sum_{j=0}^m s_j p_a(\lambda_j) w_{w_k(\delta_l)}^{(R)}(-\lambda_j) w_u(\lambda_j), \tag{76}$$

where for a scalar b , $w_{w_k(b)}^{(r)}(\lambda) = (2\pi n)^{-1/2} \sum_{t=2}^n \sum_{s=1}^{t-1} a_s^{(r)}(b) u_{t-s} e^{it\lambda}$, $a_s^{(r)}(b) = d^r a_s(b)/db^r$, and $|\widehat{\delta}_l - \delta_l| \leq |\widehat{\delta}_l - \delta_l|$. By a straightforward extension of results in Robinson and Marinucci (2001, 2003)

$$\sum_{j=0}^m s_j p_a(\lambda_j) w_{w_k(\delta_l)}^{(r)}(-\lambda_j) w_u(\lambda_j) \\ = \begin{cases} O_p(n^{\delta_k - \delta_l} (\log m)^r m^{1 - (\delta_k - \delta_l)}) & \text{if } \delta_k - \delta_l < 1 \\ O_p(n^{\delta_k - \delta_l} (\log m)^r) & \text{if } \delta_k - \delta_l \geq 1, \end{cases}$$

the only differences being that the weights $a_s^{(r)}(\delta_k - \delta_l)$ that are involved (see Lemma C.1 of RH), are not covered by the weights of Robinson and Marinucci (2001) (but it can be easily shown that they just contribute the $(\log m)^r$ factors), and the smooth weighting factor $s_j p_a(\lambda_j)$, which, as mentioned before, can be handled by simple modification of the proofs of Robinson and Marinucci (2001, 2003). Next, the summation in the second term of (76) is bounded by

$$K \sum_{j=0}^m \left| w_{w_k(\widehat{\delta}_l)}^{(R)}(-\lambda_j) \right| \|w_u(\lambda_j)\| \\ \leq K n^2 \sum_{j=1}^m \left| a_j^{(R)}(\delta_k - \widehat{\delta}_l) \right| = O_p(n^{\delta_k - \delta_l + \epsilon + 2}),$$

for any $\epsilon > 0$ in view of Lemma C.5 of RH. Thus, by Assumption 4, choosing $R > (\kappa + 2)/\kappa$, the third term of (75) is

$$O_p(n^{\delta_k - \delta_l - \kappa} \log m (m^{1 - (\delta_k - \delta_l)} 1(\delta_k - \delta_l < 1) + 1(\delta_k - \delta_l \geq 1))) \\ = o_p(n^{\delta_k - \delta_l} m^{1/2 - \min\{1/2, \delta_k - \delta_l\}}),$$

in view of (19).

Next, noting that the left side of an arbitrary element of (72) is the real part of

$$\sum_{j=0}^m s_j (\widehat{p}_a(\lambda_j) - p_a(\lambda_j)) w_{w_k(\widehat{\delta}_l)}(-\lambda_j) [w_{\widehat{u}}(\lambda_j) - w_u(\lambda_j)] \\ + \sum_{j=0}^m s_j (\widehat{p}_a(\lambda_j) - p_a(\lambda_j)) \\ \times [w_{w_k(\widehat{\delta}_l)}(-\lambda_j) - w_{w_k(\delta_l)}(-\lambda_j)] w_u(\lambda_j), \tag{77}$$

by summation by parts, similar analysis to that of (75) and a straightforward extension of (70), it can be easily shown that by Assumptions 4 and 5, (77) is

$$O_p(n^{\delta_k - \delta_l - \kappa} \log m (n^{-\kappa} + n^{-\phi} m) (m^{1 - (\delta_k - \delta_l)} 1(\delta_k - \delta_l < 1) + 1(\delta_k - \delta_l \geq 1))),$$

which is $O_p(n^{\delta_k - \delta_l} m^{1/2 - \min\{1/2, \delta_k - \delta_l\}})$ by (19), (22), (23), to conclude the proof of (43).

Finally we show (iii) which, from previous results, will be established on justifying (27) for $\bar{v}_m(\delta)$. Setting $S = (S'_S, S'_W)'$, note that the first q_S components of $S w_t(\delta)$ are purely nonstationary, whereas the last q_W are asymptotically stationary. Also, noting that $(S')^{-1} = S$, it is evident that

$$S(T')^{-1}(\bar{v}_m(\delta) - \nu) = -(STb_m(\delta)T'S')^{-1}STe_m(\delta). \tag{78}$$

In view of the proofs of (i), (ii), it can be shown that the proper normalization for the left of (78) is given by $\Lambda_{(n)}^M = (\Lambda_{(n)}^S S'_S, \Lambda_{(n)}^W S'_W)'$, noting that $(\Lambda_{(n)}^M)^{-1} = ((\Lambda_{(n)}^S)^{-1} S'_S, (\Lambda_{(n)}^W)^{-1} S'_W)'$. Thus,

$$\Lambda_{(n)}^M S(T')^{-1}(\bar{v}_m(\delta) - \nu) = -((\Lambda_{(n)}^M)^{-1} STb_m(\delta)T'S'(\Lambda_{(n)}^M)^{-1})^{-1} \\ \times (\Lambda_{(n)}^M)^{-1} STe_m(\delta),$$

$$\begin{pmatrix} S_S(\Lambda_{(n)}^S)^{-1}Tb_m(\delta)T'(\Lambda_{(n)}^S)^{-1}S'_S & S_S(\Lambda_{(n)}^S)^{-1}Tb_m(\delta)T'(\Lambda_{(n)}^W)^{-1}S'_W \\ S_W(\Lambda_{(n)}^W)^{-1}Tb_m(\delta)T'(\Lambda_{(n)}^S)^{-1}S'_S & S_W(\Lambda_{(n)}^W)^{-1}Tb_m(\delta)T'(\Lambda_{(n)}^W)^{-1}S'_W \end{pmatrix}^{-1} \begin{pmatrix} S_S(\Lambda_{(n)}^S)^{-1}Te_m(\delta) \\ S_W(\Lambda_{(n)}^W)^{-1}Te_m(\delta) \end{pmatrix}.$$

Box I.

so that $(\Lambda_{(n)}^S S'_S, \Lambda_{(n)}^W S'_W)'(T')^{-1}(\bar{v}_m(\delta) - v)$ equals the expression in **Box I**.

Next, setting $c_1(v) = D\bar{c}_1(v)$ and $c(v) = (c'_1(v)S'_S, c'_2(v), R'_m(v)S'_W)'$, as in the proofs of (i), (ii), by **Marinucci and Robinson (2000)** and **Brown (1971)**

$$c(v) \Rightarrow \left(\tilde{W}'(v; \delta) D'S'_S, \bar{W}'(v), \widehat{W}'(v)S'_W \right)', \tag{79}$$

the Gaussian processes on the right of (79) are mutually independent because $S_S c_1(w)$, $c_2(w)$, $S_W R_m(w)$, are mutually uncorrelated processes. Then (27) holds by **Theorem 2.2 of Kurtz and Protter (1991)**, noting that by results in **Robinson and Marinucci (2001)**

$$S_S(\Lambda_{(n)}^S)^{-1}Tb_m(\delta)T'(\Lambda_{(n)}^W)^{-1}S'_W = o_p(1). \quad \square$$

Proof of Theorem 2. **Theorem 2** is a direct consequence of obtaining for $\widehat{v}_m^\circ(\delta)$ and $\widehat{v}_m^\circ(\delta)$ results identical to those derived for $\widehat{v}_m(\delta)$ and $\widehat{v}_m(\delta)$. Because (42), (43) imply that the result for $\bar{v}_m^\circ(\delta)$ applies also to $\widehat{v}_m^\circ(\delta)$ and $\widehat{v}_m^\circ(\delta)$, we only give the proof for the infeasible estimate $\bar{v}_m^\circ(\delta) = b_m^\circ(\delta)^{-1}a_m^\circ(\delta)$, where $b_m^\circ(\delta)$ and $a_m^\circ(\delta)$ are defined like $\widehat{b}_m^\circ(\delta)$ and $\widehat{a}_m^\circ(\delta)$ respectively, with f replacing \widehat{f} . Clearly

$$\begin{aligned} \bar{v}_m^\circ(\delta) - v &= -b_m^\circ(\delta)^{-1}e_m^\circ(\delta), \\ e_m^\circ(\delta) &= C' \sum_{j=0}^m \text{Re} \{s_j w_{2d}(-\lambda_j) \widehat{f}^{-1}(0) w_u(\lambda_j)\}. \end{aligned}$$

Based on the proof of **Theorem 1** the required result holds on showing

$$(\Lambda_{(n)}^S)^{-1}T(e_m^\circ(\delta) - \bar{e}_m(\delta)) = o_p(1), \quad \text{under (i),} \tag{80}$$

$$(\Lambda_{(n)}^W)^{-1}T(e_m^\circ(\delta) - e_m(\delta)) = o_p(1), \quad \text{under (ii),} \tag{81}$$

$$(\Lambda_{(n)}^S)^{-1}T(b_m^\circ(\delta) - b_m(\delta))T'(\Lambda_{(n)}^S)^{-1} = o_p(1), \quad \text{under (i),} \tag{82}$$

$$(\Lambda_{(n)}^W)^{-1}T(b_m^\circ(\delta) - b_m(\delta))T'(\Lambda_{(n)}^W)^{-1} = o_p(1), \quad \text{under (ii),} \tag{83}$$

$$S_S(\Lambda_{(n)}^S)^{-1}T(b_m^\circ(\delta) - b_m(\delta))T'(\Lambda_{(n)}^W)^{-1}S'_W = o_p(1), \tag{84}$$

under (iii).

First, in order to prove (80), we show that

$$\Lambda^S(n)^{-1}E(Te_m^\circ(\delta)) = o(1),$$

by proving this for an arbitrary element of $Te_m^\circ(\delta)$, namely $\sum_{j=0}^m \text{Re}\{s_j p_a(0) I_{u w_k(\delta_l)}(\lambda_j)\}$. If $\delta_k - \delta_l > 1$ and $m = [n/2]$, it can be shown (see e.g. **Robinson and Marinucci, 2001**) that $E(\sum_{t=1}^n w_{kt}(\delta_l) u_t) = o(n^{\delta_k - \delta_l})$. For $m < [n/2]$,

$$\sum_{j=0}^m \text{Re} \{s_j I_{u w_k(\delta_l)}(\lambda_j)\} = \sum_{j=1}^n I_{u w_k(\delta_l)}(\lambda_j) + o_p(n^{\delta_k - \delta_l}),$$

by **Propositions 4.1, 4.2 of Robinson and Marinucci (2003)**; we then conclude as in the case $m = [n/2]$. For $\delta_k - \delta_l = 1$, noting (26), the result follows by (4.21) in **Robinson and Marinucci (2001)** and (26). For $\delta_k - \delta_l < 1$, the expectation can be written as the real part of

$$\begin{aligned} &\frac{1}{2\pi n} \sum_{j=0}^m s_j p_a(0) \int_{-\pi}^{\pi} D_n(\lambda_j - \mu) \\ &\times \sum_{t=1}^n e^{-i(n-t)\lambda_j} D_t(\mu - \lambda_j) (f(\mu) - f(\lambda_j)) a_{n-t} i_k d\mu \end{aligned} \tag{85}$$

$$\begin{aligned} &+ \frac{1}{2\pi n} \sum_{j=0}^m s_j p_a(0) \int_{-\pi}^{\pi} D_n(\lambda_j - \mu) \\ &\times \sum_{t=1}^n e^{-i(n-t)\lambda_j} D_t(\mu - \lambda_j) (f(\lambda_j) - f(0)) a_{n-t} i_k d\mu. \end{aligned} \tag{86}$$

First, the proof of (45) readily implies that (85) is $o(n^{\delta_k - \delta_l})$. Next, (86) is equal to

$$n^{-1} \sum_{j=1}^m p_a(0) (f(\lambda_j) - f(0)) \sum_{t=0}^{n-1} a_t i_k (n-t) e^{-it\lambda_j}. \tag{87}$$

By summation by parts, (87) has modulus

$$\begin{aligned} &\left| -n^{-1} \sum_{j=1}^m p_a(0) (f(\lambda_j) - f(0)) \right. \\ &\times \left. \sum_{t=0}^{n-2} [(a_{t+1} - a_t)(n-t) - a_{t+1}] i_k (D_t(-\lambda_j) + 1) \right| \\ &\leq Km \sum_{t=1}^n t^{\delta_k - \delta_l - 2} + Kn^{-1} m \sum_{t=1}^n t^{\delta_k - \delta_l - 1} = O(m) = o(n^{\delta_k - \delta_l}), \end{aligned}$$

by (26) and (47).

Next, as in the proof of (i) of **Theorem 1** it is not difficult to show that

$$\begin{aligned} &\sum_{j=0}^m \text{Re} \{s_j p_a(0) [I_{u w_k(\delta_l)}(\lambda_j) - E I_{u w_k(\delta_l)}(\lambda_j)]\} \\ &= \frac{p_a(0)}{2\pi} \sum_{t=1}^n w_{k,t-1}(\delta_l) A(1) \varepsilon_t + o_p(n^{\delta_k - \delta_l}), \end{aligned}$$

so we conclude (80).

Next, working again with an arbitrary element, (81) holds on showing

$$\begin{aligned} &\sum_{j=0}^m \text{Re} \{s_j (p_a(\lambda_j) - p_a(0)) I_{u w_k(\delta_l)}(\lambda_j)\} \\ &= o_p(n^{\delta_k - \delta_l} m^{1/2 - (\delta_k - \delta_l)}). \end{aligned} \tag{88}$$

By the bounds for periodogram expectations given in **Robinson (1995a, 2005)** and **Assumption 3**, the expectation of the modulus of the left side of (88) is bounded by

$$\begin{aligned} &K \left\{ \sum_{j=1}^m \|p_a(\lambda_j) - p_a(0)\| E \|I_u(\lambda_j)\| \right. \\ &\times \left. \sum_{v=1}^m \|p_a(\lambda_v) - p_a(0)\| E I_{w_k(\delta_l)}(\lambda_v) \right\}^{\frac{1}{2}} \\ &\leq K \left\{ n^{2(\delta_k - \delta_l) - 2 - 2\eta} \sum_{j=1}^m j^{1+\eta} \sum_{v=1}^m v^{1+\eta - 2(\delta_k - \delta_l)} \right\}^{\frac{1}{2}} \\ &\leq Kn^{\delta_k - \delta_l - 1 - \eta} m^{2+\eta - (\delta_k - \delta_l)}, \end{aligned}$$

so that (88) holds as $m^{3/2+\eta}/n^{1+\eta} \rightarrow 0$ as $n \rightarrow \infty$, by (25).

Next, (82), (84) hold by similar arguments to those for (80), (81), and finally (83) holds if

$$\sum_{j=0}^m \operatorname{Re} \left\{ s_j (f^{ab}(\lambda_j) - f^{ab}(0)) I_{w_k(\delta_l) w_i(\delta_v)}(\lambda_j) \right\} = o_p \left(n^{\delta_k + \delta_i - \delta_l - \delta_v} m^{1 - (\delta_k + \delta_i - \delta_l - \delta_v)} \right), \quad (89)$$

$k > l, i > v$, where if $a, b \in \{r_{j+1} + 1, \dots, r_j\}$, necessarily $k, i \in \{r_j + 1, \dots, r_0\}$. By similar arguments to those in the proof of (88), the expectation of the modulus of the left side of (89) is bounded by

$$K \sum_{j=1}^m \lambda_j^{1+\eta} E \left| I_{w_k(\delta_l) w_i(\delta_v)}(\lambda_j) \right| \leq K n^{\delta_k + \delta_i - \delta_l - \delta_v - 1} m^{2+\eta - (\delta_k + \delta_i - \delta_l - \delta_v)},$$

so that (89) holds as $m^{1+\eta}/n^{1+\eta} \rightarrow 0$ as $n \rightarrow \infty$, again by (25), to conclude the proof. \square

References

- Andrews, D.W.K., Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* 72, 569–614.
- Bandi, F., Perron, B., 2004. Long memory and the relation between implied and realized volatility. Preprint.
- Berger, R.L., 1982. Multiparameter hypothesis testing and acceptance sampling. *Technometrics* 24, 295–300.
- Brown, B.M., 1971. Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.
- Chen, W., Hurvich, C., 2003. Semiparametric estimation of multivariate fractional cointegration. *Journal of the American Statistical Association* 98, 629–642.
- Chen, W., Hurvich, C., 2006. Semiparametric estimation of fractional cointegrating subspaces. *Annals of Statistics* 27, 2939–2979.
- Cheung, Y.-W., Lai, K., 1993. A fractional cointegration analysis of purchasing power parity. *Journal of Business and Economics Statistics* 11, 93–101.
- Christensen, B.J., Nielsen, N.O., 2006. Asymptotic normality of narrowed band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics* 133, 343–371.
- Corbae, D., Ouliaris, S., 1988. Cointegration and tests of purchasing power parity. *Review of Economics and Statistics* 70, 508–511.
- Dolado, J., Marmol, F., 1996. Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes. Banco de España-Servicio de Estudios, Madrid. Preprint.
- Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction model. Representation, estimation and testing. *Econometrica* 55, 251–276.
- Flores, R., Szafarz, A., 1996. An enlarged definition of cointegration. *Economics Letters* 50, 193–195.
- Gil-Alana, L.A., Hualde, J., 2009. Fractional integration and cointegration: an overview and an empirical application. In: Patterson, K., Mills, T.C. (Eds.), *Palgrave Handbook of Econometrics*, vol. II. MacMillan, Palgrave, pp. 434–469.
- Hannan, E.J., 1970. *Multiple Time Series*. Wiley, New York.
- Hualde, J., 2008. Consistent estimation of cointegrating subspaces. Universidad Pública de Navarra. Preprint.
- Hualde, J., Robinson, P.M., 2006. Semiparametric estimation of fractional cointegration. STICERD—Econometrics Paper Series No. EM/2006/502.
- Hualde, J., Robinson, P.M., 2007. Root-n-consistent estimation of weak fractional cointegration. *Journal of Econometrics* 140, 450–484.
- Hurvich, C.M., Brodsky, J., 2001. Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. *Journal of Time Series Analysis* 22, 221–249.
- Jacobson, T., Vredin, A., Warne, A., 1998. Are real wages and unemployment related? *Economica* 65, 69–96.
- Jeganathan, P., 1999. On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* 15, 583–621.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Johansen, S., 1996. *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- Johansen, S., 2008. A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* 24, 651–676.
- Johansen, S., Juselius, K., 1992. Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics* 53, 211–244.
- Johansen, S., Juselius, K., 1994. Identification of the long-run and the short-run structure. An application to the ISLM model. *Journal of Econometrics* 63, 7–36.
- Kim, Y., 1990. Purchasing power parity in the long run. A cointegration approach. *Journal of Money, Credit and Banking* 22, 491–503.
- Kim, C.S., Phillips, P.C.B., 2000. Fully modified estimation of fractional cointegration models. Yale University, New Haven. Preprint.
- Kurtz, T.G., Protter, P., 1991. Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.
- Lasak, K., 2005. Likelihood based testing for fractional cointegration. Universitat Autònoma de Barcelona. Preprint.
- Lobato, I.N., 1999. A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics* 90, 129–153.
- MacDonald, R., Marsh, I.W., 1997. On fundamentals and exchange rates: a Casselian perspective. *The Review of Economics and Statistics* 79, 655–664.
- Marinucci, D., Robinson, P.M., 2000. Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, 103–120.
- Marinucci, D., Robinson, P.M., 2001. Semiparametric fractional cointegration analysis. *Journal of Econometrics* 105, 225–247.
- Moulines, E., Soulier, P., 1999. Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics* 27, 1415–1439.
- Nielsen, M.O., Frederiksen, P., 2008. Fully modified narrow-band least squares estimation of stationary fractional cointegration. Preprint.
- Nielsen, M.O., Shimotsu, K., 2007. Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. *Journal of Econometrics* 141, 574–596.
- Parzen, E., 1957. On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.
- Phillips, P.C.B., 1991a. Spectral regression for cointegrated time series. In: Barnett, W.A., Powell, J., Tauchen, G. (Eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press, Cambridge.
- Phillips, P.C.B., 1991b. Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
- Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
- Robinson, P.M., 1994a. Semiparametric analysis of long memory time series. *Annals of Statistics* 22, 515–539.
- Robinson, P.M., 1994b. Time series with strong dependence. In: Sims, C.A. (Ed.), *Advances in Econometrics*. Sixth World Congress. Cambridge University Press, Cambridge, pp. 47–95.
- Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048–1072.
- Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630–1661.
- Robinson, P.M., 2005. The distance between rival nonstationary fractional processes. *Journal of Econometrics* 128, 283–300.
- Robinson, P.M., 2008. Diagnostic testing for cointegration. *Journal of Econometrics* 143, 206–225.
- Robinson, P.M., Gerolimetto, M., 2006. Instrumental variables estimation of stationary and non-stationary cointegrating regressions. *Econometrics Journal* 9, 291–306.
- Robinson, P.M., Henry, M., 2003. Higher-order kernel semiparametric M-estimation of long memory. *Journal of Econometrics* 114, 1–27.
- Robinson, P.M., Hualde, J., 2003. Cointegration in fractional systems with unknown integration orders. *Econometrica* 71, 1727–1766.
- Robinson, P.M., Iacone, F., 2005. Cointegration in fractional systems with deterministic trends. *Journal of Econometrics* 129, 263–298.
- Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes. *Annals of Statistics* 29, 947–986.
- Robinson, P.M., Marinucci, D., 2003. Semiparametric frequency-domain analysis of fractional cointegration. In: Robinson, P.M. (Ed.), *Time Series with Long Memory*. Oxford University Press, Oxford, pp. 334–373.
- Robinson, P.M., Yajima, Y., 2002. Determination of cointegrating rank in fractional systems. *Journal of Econometrics* 106, 217–241.
- Shimotsu, K., Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33, 1890–1933.
- Velasco, C., 1999a. Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325–371.
- Velasco, C., 1999b. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.
- Zygmund, A., 1977. *Trigonometric Series*. Cambridge University Press, Cambridge.