Queen's Economics Department Working Paper No. 1029

Determining the Cointegrating Rank in Nonstationary Fractional Systems by the Exact Local Whittle Approach

Katsumi Shimotsu
Department of Economics, Queen's University

Morten Ørregaard Nielsen
Department of Economics, Cornell University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2006

Determining the Cointegrating Rank in Nonstationary Fractional Systems by the Exact Local Whittle Approach

Morten Ørregaard Nielsen*
Department of Economics
Cornell University

Katsumi Shimotsu
Department of Economics
Queen's University

January 15, 2006

Abstract

We propose to extend the cointegration rank determination procedure of Robinson and Yajima (2002) to accommodate both (asymptotically) stationary and nonstationary fractionally integrated processes as the common stochastic trends and cointegrating errors by applying the exact local Whittle analysis of Shimotsu and Phillips (2005). The proposed method estimates the cointegrating rank by examining the rank of the spectral density matrix of the d'th differenced process around the origin, where the fractional integration order, d, is estimated by the exact local Whittle estimator. Similar to other semiparametric methods, the approach advocated here only requires information about the behavior of the spectral density matrix around the origin, but it relies on a choice of (multiple) bandwidth(s) and threshold parameters. It does not require estimating the cointegrating vector(s) and is easier to implement than regression-based approaches, but it only provides a consistent estimate of the cointegration rank, and formal tests of the cointegration rank or levels of confidence are not available except for the special case of no cointegration. We apply the proposed methodology to the analysis of exchange rate dynamics among a system of seven exchange rates. Contrary to both fractional and integer-based parametric approaches, which indicate at most one cointegrating relation, our results suggest three or possibly four cointegrating relations in the data.

JEL Classification: C14, C32.

Key words and phrases: Cointegration rank, fractional cointegration, fractional integration, long memory, nonstationarity, semiparametric estimation, exact local Whittle estimator.

*Corresponding author. Please address correspondence to: Morten Ørregaard Nielsen, Department of Economics, Cornell University, 482 Uris Hall, Ithaca, NY 14853, USA; phone: +1 607 255 6338; fax: +1 607 255 2818; e-mail: mon2@cornell.edu

1 Introduction

The concept of fractional cointegration is attracting increasing attention from both theoretical and empirical researchers in economics and finance. A p-vector time series X_t is said to be cointegrated if each element of X_t is $I(d)$ but there exists a linear combination that is $I(d-b)$ with $b > 0$, where an $I(d)$ time series is defined to be one whose d'th difference is weakly dependent stationary. The concept of cointegration, originally developed by Granger (1981) and Engle and Granger (1987), does not restrict the value of d and b to be integer. However, the estimation methods of cointegration were developed primarily for the so-called $I(0)/I(1)$ cointegration, where it is assumed that $d = b = 1$, i.e. that X_t has a unit root and its linear combination is weakly dependent stationary. Fractional cointegration generalizes the conventional $I(0)/I(1)$ cointegration framework by allowing both d and b to be real numbers. It avoids a knife-edge distinction between $I(1)$ and $I(0)$ processes and enables substantially more flexible modeling of long-run relationships between time series.

Technical difficulties associated with fractional integration have been hindering the development of rigorous analysis of fractional integration and cointegration, but theoretical analyses of fractional cointegration are emerging, e.g. Breitung and Hassler (2002), Chen and Hurvich (2003*a*, 2003*b*, 2004), Robinson and Hualde (2003), Robinson and Marinucci (2003), Velasco (2003*a*, 2003*b*), Dolado and Marmol (2004), Nielsen (2004*b*), Christensen and Nielsen (2004), Marmol and Velasco (2004), Hassler and Breitung (2005), and Hassler, Marmol, and Velasco (2006). Recent applications of fractional cointegration can be found, for example, in Dueker and Startz (1998), Brunetti and Gilbert (2000), Kim and Phillips (2001), Marinucci and Robinson (2001), and Henry and Zaffaroni (2003).

In this paper, we extend the cointegration rank determination procedure of Robinson and Yajima (2002) to accommodate both (asymptotically) stationary and nonstationary fractionally integrated processes for the common stochastic trends and cointegrating errors. This is accomplished by applying the exact local Whittle analysis of Shimotsu and Phillips (2005). The proposed method estimates the cointegrating rank by examining the rank of the spectral density matrix of the d'th differenced process around the origin, using the exact local Whittle estimator to estimate the fractional integration order, d. Similar to other semiparametric methods, the approach advocated here only requires information about the behavior of the spectral density matrix around the origin, but it relies on a choice of (multiple) bandwidth(s) and threshold parameters. Furthermore, it does not require estimating the cointegrating vector(s) and is therefore easier to implement than regression-based approaches which are popular in applied work. However, our approach only provides a consistent estimate of the cointegration rank, and formal tests of the cointegration rank or levels of confidence are not available except for the important special case of testing the null of non-cointegration.

The ability to accommodate both stationary and nonstationary processes follows from applying the exact local Whittle analysis of Shimotsu and Phillips (2005), which generalizes the local Whittle approach of Künsch (1987) and Robinson (1995) to accommodate any value of the fractional differencing parameter, d. This feature is very attractive when analyzing economic data, because many economic time series are known to exhibit (possibly unit root) nonstationarity, and at the same time there is no strong *a priori* reason to assume that the unobservable equilibrium error is $I(0)$. By allowing both stationary and nonstationary fractionally integrated series, the approach advocated here relaxes a limitation of Robinson and Yajima (2002), who admit only stationary data.

Chen and Hurvich (2003a) also examine the rank of an averaged periodogram matrix of tapered, differenced observations, where the number of frequencies used in the periodogram average is held fixed as the sample size grows. Their method accommodates both stationary and nonstationary series and shares a similar advantage with ours, and their assumption that the cointegrating rank r needs to be strictly positive has been relaxed by Chen and Hurvich (2004) to cover the null of no cointegration. In addition, Marmol and Velasco (2004) and Hassler and Breitung (2005) propose residual-based tests of the same null hypothesis.

Similar to other semiparametric methods the exact local Whittle approach advocated here does rely on bandwidth and threshold parameters which have to be chosen in practical applications. Furthermore, formal tests of the cointegration rank or levels of confidence are not available, except for the special case of no cointegration ($r = 0$), where Theorem 6(b) below provides a valid asymptotic test. Hence, Theorem 6(b) can be considered an alternative formal test of the hypothesis of no cointegration also examined by, e.g., Marmol and Velasco (2004) and Hassler and Breitung (2005), using residuals from an estimated cointegration vector.

We apply the proposed methodology to the analysis of exchange rate dynamics following Baillie and Bollerslev (1989, 1994), Nielsen (2004b), and Hassler *et al.* (2006). Previous studies have focused on the estimation of the cointegration vector and the memory parameter of the equilibrium errors, but formal determination of the cointegrating rank has been somewhat neglected, at least in a fractional (co)integration framework. We concentrate on examining the presence of (fractional) cointegration and on determining the cointegrating rank. The data set is a system of exchange rates for seven major currencies against the US Dollar. Applying the parametric approaches of Johansen (1988, 1991) (integer-based) and Breitung and Hassler (2002) (fractional) to the data indicates that at most one cointegrating relation exists among the seven exchange rates. However, using our proposed exact local Whittle methodology we find that three or possibly even four cointegrating relations exist in the data.

The remainder of the paper is organized as follows. Section 2 introduces the model of fractional cointegration. Section 3 analyses the asymptotic behavior of the semiparametric exact local Whittle estimator of d. Section 4 derives the limit distribution of the estimate of the spectral density matrix of the d'th differenced process at the origin and describes the method of determining the cointegrating rank r also presented in Robinson and Yajima (2002). In Section 5 we present the results of a simulation study that demonstrates the finite sample feasibility of our procedure. An empirical application to exchange rate data is presented in Section 6. Proofs are collected in the Appendix in Section 7.

2 A Model of Fractional Cointegration

We consider the p-vector fractional process X_t generated by the model

$$\Delta\left(L; d_1, \ldots, d_p\right) X_t = u_t I\left\{t \geq 1\right\}, \quad t = 1, 2, \ldots, \tag{1}$$

where $I\left\{\cdot\right\}$ is the indicator function, $\Delta\left(L; d_1, \ldots, d_p\right) = diag((1-L)^{d_1}, \ldots, (1-L)^{d_p})$, and $u_t = C\left(L\right)\varepsilon_t$ is a p-vector stationary zero mean process with spectral density matrix $f_u(\lambda)$. The covariance matrix of ε_t has full rank, so without loss of generality we normalize it to I_p (the $p \times p$ identity matrix), see also Assumption 3 below. The rank of $C\left(1\right)$ is $p - r \leq p$.

The rank condition on $C(1)$ determines the cointegrating rank of X_t. As in the standard scenario, this implies that the number of cointegrating vectors is r or equivalently that the system is driven by $p - r$ common stochastic trends. Thus, the system could be generated by a triangular form like the model

$$\begin{aligned} (1-L)^{d-b}\left(X_{1t} - \alpha' X_{2t}\right) &= v_{1t} I\left\{t \geq 1\right\}, \quad t = 0, 1, 2, \ldots, && (2) \\ (1-L)^{d} X_{2t} &= v_{2t} I\left\{t \geq 1\right\}, \quad t = 0, 1, 2, \ldots, && (3) \end{aligned}$$

where X_{1t} is an r-vector, X_{2t} is a $(p-r)$-vector, and α is a $(p-r) \times r$ matrix. For simplicity, the model in (2)-(3) has equal integration orders for all the observed variables (d) and for the cointegrating errors ($d - b$). The triangular form has a straightforward interpretation as equilibrium relations given by (2) and stochastic trends given by (3). Note that in this representation, the cointegrating vectors are the rows of the $r \times p$ matrix $(I_r; -\alpha')$. Also note that (1) is more general than the triangular representation and also incorporates, e.g., the possibility of fractional multicointegration and/or polynomial cointegration which is not present in (2)-(3). However, the triangular system is simple and easy to interpret as a possible generating mechanism for X_t.

There are two main characterizations of fractional integration that have been employed in the literature, see e.g. Marinucci and Robinson (1999) and Robinson (2005). The model (1) is a convenient and unified characterization that applies for both (asymptotically) stationary and nonstationary processes. The generating process for X_{at} is

$$(1-L)^{d_a} X_{at} = u_{at} I\{t \geq 1\}, \quad t = 1, 2, \ldots \tag{4}$$

Expanding the binomial in (4) gives the form

$$\sum_{k=0}^{t} \frac{(-d_a)_k}{k!} X_{a,t-k} = u_{at} I\{t \geq 1\}, \tag{5}$$

where $(d)_k = \Gamma(d+k)/\Gamma(d) = (d)(d+1)\ldots(d+k-1)$ is the forward factorial function and $\Gamma(\cdot)$ is the gamma function. When d_a is a positive integer, the series in (4) terminates, giving the usual formula in terms of the differences and higher order differences of $X_{at.}$ Inverting (4) gives a valid linear representation of X_{at} for all values of d_a,

$$X_{at} = (1-L)^{-d_a} u_{at} I\{t \geq 1\} = \sum_{k=0}^{t-1} \frac{(d_a)_k}{k!} u_{a,t-k}. \tag{6}$$

3 Exact Local Whittle Estimation of d

3.1 Exact Local Whittle Likelihood and Estimator

Since it is not known *a priori* whether there is cointegration ($C(1)$ has reduced rank) or not ($C(1)$ has full rank), it is preferable to employ an estimator of d that makes no assumptions about the presence of cointegration and is consistent in both cases. Furthermore, cointegration is often a property associated with nonstationary time series, especially in empirical applications, so the estimator should be applicable in both the stationary and nonstationary case. Thus, we employ the univariate exact local Whittle (ELW) estimator of Shimotsu and Phillips (2005).

Define the discrete Fourier transform and the periodogram of a generic time series Z_t, $t = 1, \ldots, n$, evaluated at the fundamental frequencies as

$$\begin{aligned} w_z(\lambda_j) &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Z_t e^{it\lambda_j}, \quad \lambda_j = \frac{2\pi j}{n}, \ j = 0, 1, \ldots, n, \\ I_z(\lambda_j) &= w_z(\lambda_j) w_z(\lambda_j)^*, \end{aligned} \tag{7}$$

where the asterisk denotes complex conjugation and transposition.

Let $f_u(0) = G$ and $f_a(\lambda)$ be the a'th diagonal element of $f_u(\lambda)$. For a univariate time series X_{at} generated by (6), Shimotsu and Phillips (2005) propose to estimate (d_a, G_{aa}) by minimizing the objective function

$$Q_m(d, G) = \frac{1}{m} \sum_{j=1}^{m} \left[\log \left(G \lambda_j^{-2d} \right) + \frac{1}{G} I_{\Delta^d x_a}(\lambda_j) \right]. \tag{8}$$

Concentrating $Q_m(d, G)$ with respect to G, Shimotsu and Phillips (2005) define the ELW estimator as

$$\hat{d}_a = \arg \min_{d \in [\Delta_1, \Delta_2]} R_a(d), \tag{9}$$

where

$$R_a(d) = \log \hat{G}_{aa}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \hat{G}_{aa}(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d x_a}(\lambda_j),$$

and $-\infty < \Delta_1 < \Delta_2 < \infty$ are the lower and upper bounds of the admissible values of d. The number $m = m(n)$ is a bandwidth parameter that determines the number of periodogram ordinates used in the estimation.

3.2 Consistency

We introduce the following assumptions on the bandwidth m and the stationary component u_t in (1), which are straightforward multivariate generalizations of the assumptions in Shimotsu and Phillips (2005).

Assumption 1 The spectral density matrix $f_u(\lambda)$ satisfies

$$f_u(\lambda) \sim G \quad \text{as } \lambda \to 0^+,$$

where G is a finite and non-zero matrix with strictly positive diagonal elements.

Assumption 2 In a neighborhood $(0, \delta)$ of the origin, $f_u(\lambda)$ is differentiable and

$$\frac{d}{d\lambda} \log f_u(\lambda) = O\left(\lambda^{-1}\right) \quad \text{as } \lambda \to 0^+.$$

Assumption 3 The errors u_t satisfy

$$u_t = C(L) \varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \operatorname{tr}\left(C_j' C_j\right) < \infty, \tag{10}$$

where $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = I_p$ a.s., $t = 0, \pm 1, \ldots$, $\mathcal{F}_t = \sigma(\{\varepsilon_s, s \leq t\})$, $\operatorname{rank}(C(1)) = p - r \leq p$, and there exists a random variable ε such that $E\varepsilon^2 < \infty$ and for all $\eta > 0$, all non-null p-vectors ζ, and some $K > 0$, $P\left(\left|\zeta' \varepsilon_t\right| > \eta\right) \leq KP(|\varepsilon| > \eta)$.

Assumption 4 As $n \to \infty$,

$$\frac{1}{m} + \frac{m(\log m)^{1/2}}{n} + \frac{\log n}{m^{\gamma}} \to 0 \quad \text{for any } \gamma > 0.$$

Assumption 5

$$\Delta_2 - \Delta_1 \leq 9/2.$$

Assumptions 1-3 are analogous to Assumptions A1-A3 of Robinson (1995) (for scalar X_t) and Assumptions A1-A3 of Lobato (1999), although our assumptions apply to u_t rather than X_t. The condition that $G_{aa} > 0$ is satisfied if $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} C_{a,j} C'_{a,k} > 0$, where G_{ab} is the (a, b)'th element of G and $C_{a,j}$ is the a'th row of C_j, and is relatively innocuous. It is basically to ensure that u_{at} cannot be overdifferenced, i.e. that u_{at} is indeed $I(0)$ and hence that X_{at} through (1) is $I(d_a)$ for $a = 1, ..., p$.

In Assumption 3 the possibility of cointegration is introduced by letting $C(1)$ have possibly reduced rank. If $C(1)$ has rank less than p, i.e. reduced rank, there is cointegration among the elements of X_t, whereas if $C(1)$ has full rank, p, then there is no cointegration. Assumption 4 is slightly stronger than the corresponding assumptions of Robinson (1995) and Lobato (1999). Assumption 5 is identical to Assumption 5 of Shimotsu and Phillips (2005).

Note that the matrix $G = C(1) C(1)' / (2\pi)$ has reduced rank if there is cointegration and full rank otherwise, see e.g. Robinson and Yajima (2002) or Nielsen (2004*a*). Thus, as pointed out by Robinson and Yajima (2002, page 229), the multivariate local Whittle estimation developed by Lobato (1999) and Lobato and Velasco (2000) is not appropriate under cointegration since it assumes full rank of G.

Under these conditions we may now establish the consistency of $\hat{d}_a$ under both the presence and absence of cointegration. In what follows we redefine $d = (d_1, \ldots, d_p)'$ and $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_p)'$.

Theorem 1 *Suppose X_t is generated by (1) and Assumptions 1-5 hold. Then, for $d \in [\Delta_1, \Delta_2]^p$, $\hat{d} \to_p d$ as $n \to \infty$.*

3.3 Asymptotic Normality

We proceed to derive the joint asymptotic distribution of $\hat{d}$, which requires a strengthening of our assumptions as in Shimotsu and Phillips (2005), see also Robinson and Yajima (2002).

Assumption 1$'$ For some $\beta \in (0, 2]$,

$$f_u(0) = G(1 + O(\lambda^{\beta})) \quad \text{as } \lambda \to 0^+,$$

and $G = f_u(0) = C(1) C(1)' / (2\pi)$ is a finite and non-zero matrix with strictly positive diagonal elements.

Assumption $2'$ In a neighborhood $(0, \delta)$ of the origin, $C(e^{i\lambda})$ is differentiable and

$$\frac{d}{d\lambda} C(e^{i\lambda}) = O\left(\lambda^{-1}\right) \qquad \text{as } \lambda \to 0^+.$$

Assumption $3'$ Assumption 3 holds and furthermore the matrices $E\left(\varepsilon_t \otimes \varepsilon_t \varepsilon_t' \middle| \mathcal{F}_{t-1}\right)$ and $E\left(\varepsilon_t \varepsilon_t' \otimes \varepsilon_t \varepsilon_t' \middle| \mathcal{F}_{t-1}\right)$ are nonstochastic, finite, and do not depend on t.

Assumption $4'$ As $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta} (\log m)^2}{n^{2\beta}} + \frac{\log n}{m^\gamma} \to 0 \quad \text{for any } \gamma > 0.$$

Assumption $5'$ Assumption 5 holds.

Assumptions $1'$-$3'$ are multivariate extensions of Assumptions $1'$-$3'$ of Shimotsu and Phillips (2005) and are analogous to Assumptions A1-A3 of Lobato (1999) and Assumptions B-D of Robinson and Yajima (2002). Similarly to Assumptions 1-3 we impose our assumptions on the spectral density of u_t, whereas Lobato (1999) and Robinson and Yajima (2002) impose their assumptions on the spectral density of X_t. Assumption $4'$ is slightly stronger than the comparable Assumption A4 of Lobato (1999) and Assumption E of Robinson and Yajima (2002).

Due to the approximation of the spectral density of (4) near the origin, the value of β is bounded by $\min_{1 \leq a \leq r} b_a$, where b_a is the reduction in the integration order implied by the a'th cointegration vector. For example, if the system (1) is generated by (2)-(3) the approximation of the spectral density of the first element of X_{1t} is $G_{11}\left(1 + O\left(\lambda^{b_1}\right) + O\left(\lambda^{\beta}\right)\right)$ under Assumption $1'$, see Nielsen (2004*a*). A similar condition seems to be missing in Robinson and Yajima (2002). The practical implication of this condition on β is that stronger cointegration allows one to choose a wider bandwidth. In most economic applications with nonstationary data, the cointegrating strength (b_a) will presumably be at least 1/2 which means that at least $\beta = 1/2$ can be used in Assumption $4'$ if also the data is assumed to be generated by certain multivariate ARFIMA models (which imply $\beta = 2$). Thus, Assumption $4'$ essentially reduces to $m = o\left(n^{1/2}\right)$ in that case.

The next theorem establishes the joint asymptotic normality of the univariate ELW estimators when $d \in (\Delta_1, \Delta_2)^p$. We define $D = diag\left(G_{11}, \ldots, G_{pp}\right)$, where G_{ab} is the (a, b)'th element of G, and denote the Hadamard product by $\circ$.

Theorem 2 *Suppose X_t is generated by (1) and Assumptions $1'$-$5'$ hold. Then, for $d \in (\Delta_1, \Delta_2)^p$,*

$$\begin{aligned} \sqrt{m}(\hat{d} - d) &\rightarrow {}_d N\left(0, \frac{1}{4} D^{-1} (G \circ G) D^{-1}\right) \quad \textit{and} \\ \hat{G}(\hat{d}) &= \frac{1}{m} \sum_{j=1}^{m} \operatorname{Re}\left[I_{\Delta(L;\hat{d})x}(\lambda_j)\right] \rightarrow_p G \quad \textit{as } n \rightarrow \infty. \end{aligned}$$

3.4 Remark

In many economic applications, the mean (initial value) of X_t is unknown and the data generating process is given by

$$\Delta (L; d_1, \ldots d_p) (X_t - \mu) = u_t I \{t \geq 1\}. \quad t = 1, 2, \ldots, \tag{11}$$

where μ is a nonrandom p-vector. Let μ_a be the a'th element of μ. Shimotsu (2004) proposes to estimate μ_a by

$$\widehat{\mu}_a (d) = w(d)\overline{X}_a + (1 - w(d))X_{a1},$$

where $\overline{X}_a = n^{-1} \sum_{t=1}^{n} X_{at,}$ the sample average, and $w(d)$ is a smooth (twice continuously differentiable) weight function such that $w(d) = 1$ for $d \leq 1/2$, $w(d) \in [0, 1]$ for $1/2 \leq d \leq 3/4$, and $w(d) = 0$ for $d \geq 3/4$. With this substitution, the objective function takes the form

$$R_a^{\diamond} (d) = \log \hat{G}_{aa}^{\diamond} (d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \hat{G}_{aa}^{\diamond} (d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d (x_a - \widehat{\mu}_a(d))} (\lambda_j).$$

Shimotsu (2004, Theorem 5) shows that the two-step feasible ELW estimator, which is based on the objective function $R_a^{\diamond} (d)$ and uses a tapered estimator by Velasco (1999) as the first stage estimator, is consistent and has the same $N(0, 1/4)$ limiting distribution as the ELW estimator for $d \in (-1/2, 2)$ under the additional assumption that $f_u(\lambda)$ is bounded for $\lambda \in [0, \pi]$.[1] Therefore, if the data are generated by (11), all the results in this section hold if we assume $f_u(\lambda)$ is bounded and estimate d_a by the two-step feasible ELW estimator. Shimotsu (2004) also shows that the presence of a polynomial time trend can be dealt with simply by prior detrending of the data.

[1] Indeed, Shimotsu (2004) also assumes $f_u(\lambda) = G + E_\beta \lambda^\beta + o(\lambda^\beta)$ with $E_\beta < \infty$ and $\beta \in (1, 2]$, following the assumptions in Velasco (1999). However, in view of the results of Lobato and Velasco (2000, page 415), $f_u(\lambda) = G(1 + O(\lambda^\beta))$ is sufficient.

3.5 Testing Equality of Integration Orders

With the above result of Theorem 2, we are now able to test joint hypotheses on the integration orders, d. For instance, we could test the hypothesis of pairwise equality of the integration orders,

$$H_{ab} : d_a = d_b,$$

or the hypothesis of equality of all the integration orders,

$$H_0 : d_a = d_*, a = 1, ..., p,$$

for some $d_* \in (\Delta_1, \Delta_2)$. The test statistics suggested by Robinson and Yajima (2002) to test H_{ab} and H_0 are

$$\begin{aligned}
\hat{T}_{ab} &= \frac{m^{1/2}(\hat{d}_a - \hat{d}_b)}{\left(\frac{1}{2}\left(1 - \hat{G}_{ab}^2 / \left(\hat{G}_{aa}\hat{G}_{bb}\right)\right)\right)^{1/2} + h\left(n\right)}, \\
\hat{T}_0 &= m\left(S\hat{d}\right)' \left(S\frac{1}{4}\hat{D}^{-1}\left(\hat{G} \circ \hat{G}\right)\hat{D}^{-1}S' + h\left(n\right)^2 I_{p-1}\right)^{-1}\left(S\hat{d}\right),
\end{aligned}$$

where $S = [I_{p-1}; -\iota]$, ι is the $(p-1)$-vector of ones, and $h(n) > 0$ satisfies the following assumption. Note that $h(n) = (\log n)^{-k}$ for any $k > 0$ satisfies Assumption 6 if $(\log m)^2 m^{1+2\beta}/n^{2\beta} = o((\log n)^{-k})$.

Assumption 6 As $n \to \infty$,

$$h\left(n\right) + \frac{\left(\log m\right)^2 m^{1+2\beta}/n^{2\beta} + \left(\log m\right)^2 m^{-1/6}}{h\left(n\right)} \to 0.$$

Theorem 3 *Suppose X_t is generated by (1) and Assumptions 1′-5′ and 6 hold. Then, under H_{ab} and $d \in (\Delta_1, \Delta_2)^p$, as $n \to \infty$,*

$$\begin{aligned}
&(i) \quad \text{If } X_{at} \text{ and } X_{bt} \text{ are not cointegrated, } \hat{T}_{ab} \to_d N\left(0, 1\right), \\
&(ii) \quad \text{If } X_{at} \text{ and } X_{bt} \text{ are cointegrated, } \hat{T}_{ab} \to_p 0,
\end{aligned}$$

and under H_0 and $d_* \in (\Delta_1, \Delta_2)$, as $n \to \infty$,

$$\begin{aligned}
&(iii) \quad \text{If } X_t \text{ is not cointegrated, i.e. } r = 0,\ \hat{T}_0 \to_d \chi^2_{p-1}, \\
&(iv) \quad \text{If } X_t \text{ is cointegrated, i.e. } r \geq 1,\ \hat{T}_0 \to_p 0,
\end{aligned}$$

The proof of the theorem is identical to that of Theorem 2 of Robinson and Yajima (2002) and is omitted. Note that $h(n)$ is included in the definition of $\hat{T}_{ab}$ because $1 - \hat{G}_{ab}^2/(\hat{G}_{aa}\hat{G}_{bb})$ converges to 0 in probability under cointegration and $h(n)^2$ is included in $\hat{T}_0$ for an analogous reason. See Robinson and Yajima (2002, p. 227) for further discussion. In practice, both $\hat{T}_{ab}$ and $\hat{T}_0$ may be sensitive to the choice of $h(n)$. Choosing $h(n)$ too large leads to underrejection of H_0 under non-cointegration, while choosing $h(n)$ too small leads to overrejection of H_0 under cointegration.

It follows from Theorem 3 that tests of equality of the integration orders of the observed variables can be carried out by the approach of Robinson and Yajima (2002, pp. 227-228) even in the present model with potentially nonstationary data. It follows straightforwardly from Theorem 2 that (i) $\hat{T}_{ab}$ diverges to infinity under the alternative where H_{ab} does not hold, and (ii) $\hat{T}_0$ diverges to infinity under the alternative where H_0 does not hold.

4 Exact Local Whittle Estimation of G

Now we consider the estimation of the cointegrating rank of X_t by estimating G and its eigenvalues. For simplicity, we assume in the following that the integration orders are equal for each of the observed variables and denote the common value of $d_1, \dots, d_p$ by d_*. Define

$$\hat{G}(d_*) = \frac{1}{m_1} \sum_{j=1}^{m_1} \mathrm{Re}\left[I_{\Delta(L;d_*,\dots,d_*)x}(\lambda_j) \right], \tag{12}$$

where $I_{\Delta(L;d_*,\dots,d_*)x}(\lambda_j)$ is the periodogram of $(\Delta^{d_*} X_{1t}, \dots, \Delta^{d_*} X_{pt})'$, and let G_a be the a'th column of G. The estimator $\hat{G}(d_*)$ uses a new bandwidth parameter $m_1(n)$ in anticipation of the complications that arise when d_* is estimated.

Lemma 4 *Suppose X_t is generated by (1) and Assumptions $1'$-$5'$ hold with m replaced by m_1. Then, as $n \to \infty$,*

$$m_1^{1/2} vec(\hat{G}(d_*) - G) \to_d N(0, \tfrac{1}{2}(G \otimes G + (G \otimes G_1, \dots, G \otimes G_p)).$$

Because d_* is unknown, we need to substitute it with an estimate. As previously mentioned, we cannot use the multivariate version of the exact local Whittle estimator to estimate d_*, because G does not have full rank when X_t is cointegrated. The estimator also needs to converge to d_* at a faster rate than $m_1^{1/2}$. Therefore, we estimate G by (12) based on m_1 periodogram ordinates and each d_a by $\hat{d}_a$ based on (9) using m ordinates with $m/m_1 \to 0$, and define $\bar{d}_* = p^{-1} \sum_{a=1}^{p} \hat{d}_a$. In particular, we need the following assumption on m and m_1.

Assumption 6$'$ As $n \to \infty$,

$$\frac{1}{m_1} + \frac{\log n}{m_1^{\gamma}} + \frac{m_1^{1/2} (\log n)^2}{m^{1/2}} + \frac{m^{1+2\beta} (\log m)^2}{n^{2\beta}} + \frac{\log n}{m^{\gamma}} \to 0 \quad \text{for any } \gamma > 0.$$

The last three terms of this assumption are analogous to Assumption H of Robinson and Yajima (2002). We conjecture their Proposition 3 still holds if their Assumption H is replaced by the weaker assumption that $m^{-1/2} m_1^{1/2} (\log n)^2 + n^{-2\beta} m^{1+2\beta} (\log m)^2 \to 0$ (in our notation).

Lemma 5 *Suppose X_t is generated by (1) and Assumptions 1$'$-3$'$, 5$'$-6$'$ hold. Then, as $n \to \infty$,*

$$m_1^{1/2} vec(\hat{G}(\bar{d}_*) - G) \to_d N(0, \tfrac{1}{2}(G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p)).$$

When X_t has an unknown mean and is generated by (11), Lemmas 4 and 5 still hold if we use the two-step feasible ELW estimator to estimate d_a and replace X_{at} with $X_{at} - \widehat{\mu}_a(\widehat{d_a})$ in the periodograms in (12).

With Lemmas 4 and 5 in hand, we can use the results in Robinson and Yajima (2002) along with $\hat{G}(\bar{d}_*)$ defined in (12) to estimate the cointegrating rank r and conduct inference. The following proposal to determine the cointegrating rank of r via model selection procedures follows the proposal of Robinson and Yajima (2002, pp. 229-231), and is summarized here for completeness.

First, we state the assumption on the cointegrating rank.

Assumption 7$'$ Rank$(G) = p - r$, for $0 \leq r < p$, and the nonzero eigenvalues of G are distinct.

Let δ_a and $\hat{\delta}_a$ be the a'th eigenvalues of G and $\hat{G}(\bar{d}_*)$, respectively, $a = 1, ..., p$, and ordered descendingly with $\delta_1 > ... > \delta_{p-r} > 0$ and $\delta_{p-r+1} = ... = \delta_p = 0$. Define, for $j = 1, ..., p-1$, the statistics

$$\pi_j = \frac{\sum_{a=p-j+1}^{p} \delta_a}{\sum_{a=1}^{p} \delta_a}, \tag{13}$$

$$\hat{\pi}_j = \frac{\sum_{a=p-j+1}^{p} \hat{\delta}_a}{\sum_{a=1}^{p} \hat{\delta}_a}, \tag{14}$$

$$s_j^2 = \frac{\left(\sum_{a=p-j+1}^{p} \hat{\delta}_a\right)^2 \sum_{a=1}^{p-j} \hat{\delta}_a^2 + \left(\sum_{a=1}^{p-j} \hat{\delta}_a\right)^2 \sum_{a=p-j+1}^{p} \hat{\delta}_a^2}{\left(\sum_{a=1}^{p} \hat{\delta}_a\right)^4}. \tag{15}$$

Then a hypothesis testing procedure based on the $\hat{\pi}_j$ can be employed to determine the cointegrating rank r using the asymptotic theory described below.

Another possibility is to apply a model selection procedure to determine r. We follow the model selection procedure proposed by Robinson and Yajima (2002) (c.f. Fujikoshi and Veitch, 1979; Fujikoshi, 1985; Gunderson and Muirhead, 1997) and estimate r by

$$\hat{r} = \arg \min_{u=0,\ldots,p-1} L(u), \tag{16}$$

where

$$L(u) = v(n)(p-u) - \sum_{a=1}^{p-u} \hat{\delta}_a,$$

for some $v(n) > 0$ which is assumed to satisfy the following assumption.

Assumption 8$'$ As $n \to \infty$,

$$v(n) + \frac{1}{m_1^{1/2} v(n)} \to 0.$$

Theorem 6 *(a) Suppose X_t is generated by (1) and Assumptions 1$'$-3$'$, 5$'$-7$'$ hold. Then, as $n \to \infty$, $\sqrt{m_1}\left(\hat{\delta}_a - \delta_a\right)$ are asymptotically independent for $a = 1, \ldots, p$, $\sqrt{m_1}\left(\hat{\delta}_a - \delta_a\right) \to_d N\left(0, \delta_a^2\right)$ for $a = 1, ..., p-r$, and $\sqrt{m_1}\left(\hat{\delta}_a - \delta_a\right) \to_p 0$ for $a = p-r+1, ..., p$.*
(b) Suppose X_t is generated by (1), Assumptions 1$'$-3$'$, 5$'$-7$'$ hold, and $r = 0$. Then, as $n \to \infty$,

$$m_1^{1/2}(\hat{\pi}_j - \pi_j)/s_j \to_d N(0,1) \quad \text{for } j = 1, \ldots, p-1,$$

where $\hat{\pi}_j, \pi_j$ and s_j are defined in (13)-(15) and computed using $\hat{G}(\bar{d}_*)$ in (12).
(c) Suppose X_t is generated by (1) and Assumptions 1$'$-3$'$, 5$'$-8$'$ hold. Then

$$\lim_{n \to \infty} \Pr(\hat{r} = r) = 1,$$

where $\hat{r}$ is defined in (16) and computed using $\hat{G}(\bar{d}_*)$ in (12).

The proof of the theorem is identical to that of Theorems 3 and 4 of Robinson and Yajima (2002) and is omitted. Indeed, Robinson and Yajima's model selection procedure estimates r by $\widetilde{r} = \arg\min_{u=1,\ldots,p-1} L(u)$, thereby not allowing for the possibility of $r = 0$, but their proof holds for the important case with $r = 0$, i.e. in the absence of cointegration. As mentioned in Robinson and Yajima (2002), the model selection procedure may be conducted using the correlation matrix $\hat{P}(\bar{d}_*) = \hat{D}(\bar{d}_*)^{-1/2}\hat{G}(\bar{d}_*)\hat{D}(\bar{d}_*)^{-1/2}$, where $\hat{D}(\bar{d}_*)$ is the diagonal matrix whose a'th diagonal element is the same as that of $\hat{G}(\bar{d}_*)$. In simulations we found that the model selection procedure performs substantially better when it is based on $\hat{P}(\bar{d}_*)$ rather than $\hat{G}(\bar{d}_*)$.

As discussed by Robinson and Yajima (2002), part (b) of the theorem could be applied to determine r by hypothesis testing following Phillips and Ouliaris (1988), although this method suffers from the assumption that $r = 0$ in part (b). The suggestion is that there is evidence in favor of the hypothesis that the cointegrating rank is r (against the alternative that the rank is greater than r) when the $100(1-\alpha)\%$ upper confidence interval for π_r,

$$CI(\alpha, r) = \hat{\pi}_r + s_r z_\alpha / m_1^{1/2}, \tag{17}$$

is smaller than some prescribed threshold such as $0.1/p$, where z_α is the $100\,(1-\alpha)\,\%$ point of the standard normal distribution.

5 Simulations

This section reports the results of some simulations that were conducted to examine the finite sample performance of the proposed procedure. The dimension of the system (p) is set to 4, and X_{1t} and X_{2t} are generated from (2) and (3) with $(v_{1t}', v_{2t}')' \sim iidN\,(0, I_4)$. The cointegration coefficient α is set to $\alpha = (1, 1, -1)'$ for $r = 1$, $\alpha = ((1, 0.5)', (0.5, 1)')$ for $r = 2$, and $\alpha = (1, 1, -1)$ for $r = 3$. Note that the cointegrating vectors are the rows of $(I_r; -\alpha')$. The integration orders d_a are estimated by the ELW estimator without allowance for a non-zero mean, where Δ_1 and Δ_2 are set to -1 and 3. Two sample sizes, $n = 128$ and $n = 512$, and 1,000 replications are used. The bandwidth parameters $(m, m_1, v(n))$ are chosen to be $\left(\left[n^{0.65}\right], \left[n^{0.6}\right], m_1^{-0.3}\right)$, where $[x]$ denotes the largest integer smaller than or equal to x. The value of d is chosen to be 1, and the value of b is selected from $\{0.2, 0.4, 0.6, 0.8\}$. Given the range of the data generating mechanism that the model (1) accommodates, this example is far from exhaustive, but an extensive simulation exercise is beyond the scope of this paper.

Tables 1 and 2 report the simulation results with $n = 128$ and $n = 512$, respectively. In both tables, freq($\hat{\pi}_r$) denotes the frequency of $(CI(0.05, 1) < 0.1/p, CI(0.05, 2) < 0.1/p, CI(0.05, 3) < 0.1/p)$, i.e. $\alpha = 0.05$, and freq($\hat{r}$) denotes the frequency of $(\hat{r} = 0, \hat{r} = 1, \hat{r} = 2, \hat{r} = 3)$. The model selection procedure is based on the correlation matrix $\hat{P}(\bar{d}_*)$. Furthermore, rej($\hat{T}_0$) denotes the rejection frequency of the test of the equality of the integration orders, $\hat{T}_0$, with the 5% asymptotic critical value and two choices of $h(n)$, $1/(\log n)^{1/2}$ and $1/\log n$, from left to right.

Tables 1 and 2 about here

The rank determination based on $CI(\alpha, r)$ does not appear to perform very well. Although it never selects $r \geq 1$ when the true r is zero, it tends to choose r too small when $r \geq 1$. This is because the upper confidence interval, $\hat{\pi}_r + s_r z_\alpha / m_1^{1/2}$, does not take a sufficiently small value for $r \geq 2$.

On the other hand, the model selection procedure appears to perform very well, even when $n = 128$. For the small sample size, $n = 128$, it chooses the correct r in many cases except when $r = 3$ and $b \leq 0.4$. For the same value of n, its performance improves as b increases. Furthermore, the accuracy of the procedure increases as n increases except for $b = 0.2$ and $r = 3$.

In the above simulations, $v(n) = m_1^{-0.3}$ is chosen so that $\hat{r} = 0$ is chosen with frequency higher than 95% when $r = 0$ and $n = 128$. Note that a large $\hat{r}$ is more likely to be chosen when a large $v(n)$ is used and a small $v(n)$ leads to a conservative (small) estimate of r. Since the outcome of the model selection procedure may strongly depend on the choice of $v(n)$, it is prudent to compute $\hat{r}$ for different choices of $v(n)$ in practical applications.

The test based on $\hat{T}_0$ works reasonably well with $h(n) = 1/(\log n)^{1/2}$. But the test over-rejects substantially with $h(n) = 1/\log n$. Overall, the test is sensitive to the choice of $h(n)$, but non-rejection of H_0 with small $h(n)$ would strongly suggest the equality of the integration orders.

6 Empirical Application

The analysis of exchange rate dynamics and potential (fractional) cointegrating relations between exchange rates for different currencies has attracted much attention recently. Baillie and Bollerslev (1989) find evidence of a cointegrating relation between seven different (log) spot exchange rates using conventional cointegration methods. This finding is challenged by Diebold, Gardeazabal, and Yilmaz (1994) who show that when including an intercept the conclusion may change for the Baillie and Bollerslev (1989) data set. Diebold *et al.* (1994) find further support of this in an analysis of a different data set covering a longer span of time.

Baillie and Bollerslev (1994) argue that the failure of conventional cointegration tests to find evidence of cointegration in the Baillie and Bollerslev (1989) exchange rate data is due to the presence of fractional cointegration. Thus, they estimate the cointegration vector by OLS following Cheung and Lai (1993) and fit a simple fractionally integrated white noise model to the residuals. They conclude that there exists a cointegrating relationship between the exchange rates with $d = 1$ and $b = 0.11$ (in the notation from the introduction). However, their estimate of the integration order of the equilibrium errors ($d - b = 0.89$) may be upwards biased since relevant short-run dynamics may have been left out. This is indeed what is concluded by Kim and Phillips (2001) who employ their fractional fully modified estimation procedure to a different data set covering a longer time span but the same exchange rates. They find that the equilibrium errors are best described by an ARFIMA(1,d,0) process with $d = 0.33$.

All the above studies concentrate on the estimation of the cointegration vector and/or the estimation of the memory parameter of the equilibrium errors, but no formal testing of the hypothesis of fractional cointegration is attempted. Nielsen (2004*b*) and Hassler *et al.* (2006) take the opposite approach and concentrate on testing for the presence of cointegration against fractional alternatives. In applications to the same exchange rates as in the above studies, Nielsen (2004*b*) finds evidence of cointegration possibly with fractional integration in the cointegrating relation and Hassler *et al.* (2006) find two (polynomial) fractional cointegrating vectors.

We take the same focus as in Nielsen (2004*b*) and Hassler *et al.* (2006), and apply our new procedure to the same data set as in Nielsen (2004*b*) to determine the cointegrating rank. The data set is a system of log exchange rates for the currencies of the following seven countries: (West) Germany, United Kingdom, Japan, Canada, France, Italy, and Switzerland against the US Dollar. The same currencies are examined in the studies cited above. However, where Baillie and Bollerslev (1989, 1994) and Diebold *et al.* (1994) consider approximately 5 years of daily observations and Kim and Phillips (2001) consider 40 years of quarterly observations, our data set is comprised of monthly averages of noon (EST) buying rates and runs from January 1974 through December 2001 for a total of $n = 336$ observations. Thus, our data set, which is extracted from the Federal Reserve Board of Governors G.5 release, covers only the period of the current flexible exchange rate regime, but a much longer span of time than the Baillie and Bollerslev (1989) data set. A long time span has generally been found to be important in detecting long-run relations. Figure 1 shows a time series plot of the seven log-exchange rate data series.

Figure 1 about here

Table 3 presents the fractional integration analysis of the data set applying the feasible exact local Whittle estimator of Shimotsu (2004) with allowance for a non-zero mean. In this case, $w(d)$ is chosen to be $(1/2)[1+\cos(4\pi d)]$ for $d \in [1/2, 3/4]$. The two rows are the estimates of the fractional integration orders estimated with bandwidth parameter $m = \left[n^{0.6}\right] = 32$ and $m = \left[n^{0.55}\right] = 18$ (note that $[x]$ denotes the largest integer less than or equal to x). The standard errors reported in parenthesis are calculated as $(4m)^{-1/2}$. When $m = 32$, the values of the $\hat{T}_0$ statistic are 5.31 and 6.80 with $h(n) = 1/(\log n)^{1/2}$ and $1/\log n$, respectively. When $m = 18$, the $\hat{T}_0$ statistic takes the values 3.05 and 5.01 with $h(n) = 1/(\log n)^{1/2}$ and $1/\log n$, respectively. Since the 95% critical value of the $\chi^2(5)$ distribution is 11.01, we easily accept the null of equality of the integration orders. The final column gives estimates of a common integration order $\bar{d}_*$, which we use in our fractional cointegration analysis, computed simply as an average of the estimated integration orders for each exchange rate.

Table 3 about here

From the estimates in Table 3 it is clear that the exchange rates can be well described as $I(1)$ processes. Indeed, none of the estimates are significantly different from unity at conventional significance levels. Hence, the results of Table 3 support the overwhelming evidence in the previous literature that exchange rates are $I(1)$. E.g. Baillie and Bollerslev (1989) conduct unit root tests of the $I(1)$ hypothesis against the $I(0)$ alternative, whereas Baillie (1996) and Nielsen (2004*b*) provide evidence from fractional models. These results in particular support the use of a rank determination procedure that allows for nonstationary data.

Table 4 about here

In Table 4 the estimated eigenvalues of $\hat{G}(\bar{d}_*)$ from (12) with $X_{at} - \widehat{\mu}_a(\widehat{d}_a)$ replacing X_{at} as well as the eigenvalues of the correlation matrix $\hat{P}(\bar{d}_*) = \hat{D}(\bar{d}_*)^{-1/2}\hat{G}(\bar{d}_*)\hat{D}(\bar{d}_*)^{-1/2}$ are displayed for our exchange rate data. The two rows in each panel are the estimated eigenvalues for bandwidth parameters $\left(m_1 = \left[n^{0.55}\right] = 24, m = 32\right)$ and $\left(m_1 = \left[n^{0.45}\right] = 13, m = 18\right)$, respectively. These intermediate results seem to indicate that at least a few of the eigenvalues of G could be zero. Thus, we expect that there will be evidence in favor of cointegration and possibly with more than one cointegrating relation, i.e. we expect that the rank could be greater than unity.

Table 5 about here

Table 5 displays the results of the rank determination analysis applied to the exchange rate data using the model selection procedure with $\hat{P}(\bar{d}_*)$. The results quite clearly indicate the presence of at least three but possibly four cointegrating relations. Indeed, for the case with the largest $v(n)$, i.e. with $v(n) = m_1^{-0.05}$, some evidence that the rank may be as high as five is found. All other choices of bandwidth parameters and $v(n)$ support the finding that the cointegration rank is either three or four.

Table 6 about here

For comparison, we have also computed some parametric rank tests which are shown in Table 6. The first panel of the table shows results from the Johansen (1988, 1991) trace tests with unrestricted constant term and lag augmentations 0, 3, 6, and 12, respectively, as well as the asymptotic 95% critical values for each r. The second panel shows results from applying the Breitung and Hassler (2002) parametric fractional cointegration rank tests (allowing for a nonzero mean in the levels) with the same lag augmentations as for the Johansen trace tests. The final column gives the 95% critical values for the Breitung and Hassler tests for each value of r based on their asymptotic $\chi^2\left((p-r)^2\right)$ distribution.

The Johansen tests in Table 6 all give borderline results between $r = 0$ and $r = 1$, and the Breitung-Hassler tests indicate $r = 1$ (except with no lag augmentation). Intuitively, there may be several reasons for the parametric testing procedures to indicate a lower rank. The Johansen tests may fail to detect some cointegrating relations if the cointegrating strength, b, for those relations is low. In particular, if $d = 1$ and $b < 1/2$, say, the linear combination is nonstationary and may thus not be detected as a cointegrating relation by the $I(0)/I(1)$-motivated Johansen tests. Some evidence that this may in fact be the case is given in Nielsen (2004b). On the other hand, the Johansen tests may have non-trivial power for small b, even if they are designed for stationary alternatives, similar to other alternative procedures such as Dickey-Fuller tests, see Krämer and Marmol (2004). The Breitung-Hassler tests should be able to detect the presence of such "weak" cointegrating relations, requiring only that $d > 1/2$ and $b > 0$. The inability of the Breitung-Hassler tests may instead be due to their parametric nature, i.e. possibly misspecified autocorrelation structure and lag augmentation. Hence, this illustrates the usefulness of our new methodology, and in particular highlights the advantages of its semiparametric nature and its ability to detect fractional cointegration among nonstationary fractionally integrated variables.

7 Appendix: Proofs

7.1 Proof of Theorem 1

We show that the consistency assumptions in Shimotsu and Phillips (2005) are satisfied for each component of X_t, i.e. for each $u_{at}, a = 1, .., p$. Denote by $C_{a,j}$ the a'th row of C_j and write u_{at} as

$$u_{at} = \sum_{j=0}^{\infty} C_{a,j}\varepsilon_{t-j} = \sum_{j=0}^{\infty} \tilde{c}_j \tilde{\varepsilon}_{t-j},$$

defining $\tilde{\varepsilon}_t = \|C_{a,0}\|^{-1} C_{a,0}\varepsilon_t, \tilde{\varepsilon}_{t-1} = \|C_{a,1}\|^{-1} C_{a,1}\varepsilon_{t-1}, ..., \tilde{\varepsilon}_{t-j} = \|C_{a,j}\|^{-1} C_{a,j}\varepsilon_{t-j}, ...,$ and $\tilde{c}_j = \|C_{a,j}\|$, where for any column vector y, $\|y\| = (y'y)^{1/2}$ is the vector Euclidean norm. If $\|C_{a,k}\| = 0$ we set $\tilde{c}_k = \tilde{\varepsilon}_{t-k} = 0$.

To show that Assumption 3 of Shimotsu and Phillips (2005) is satisfied, first note that $\tilde{\varepsilon}_t$ and $\tilde{\varepsilon}_t^2 - 1$ are martingale difference sequences since

$$\begin{aligned} E\left(\tilde{\varepsilon}_t | \mathcal{F}_{t-1}\right) &= \frac{C_{a,0}}{\|C_{a,0}\|} E\left(\varepsilon_t | \mathcal{F}_{t-1}\right) = 0 \\ E\left(\tilde{\varepsilon}_t^2 \middle| \mathcal{F}_{t-1}\right) &= \frac{1}{\|C_{a,0}\|^2} C_{a,0} E\left(\varepsilon_t \varepsilon_t' \middle| \mathcal{F}_{t-1}\right) C_{a,0}' \\ &= \frac{1}{\|C_{a,0}\|^2} C_{a,0} C_{a,0}' = 1. \end{aligned}$$

Second, the coefficients $\tilde{c}_j$ satisfy

$$\sum_{j=0}^{\infty} \tilde{c}_j^2 = \sum_{j=0}^{\infty} C_{a,j} C_{a,j}' < \sum_{j=0}^{\infty} \operatorname{tr}\left(C_j' C_j\right) < \infty.$$

Third, by definition of $\tilde{\varepsilon}_t$ the domination condition in our Assumption 3 with $\zeta' = \|C_{a,0}\|^{-1} C_{a,0}$, implies that the domination condition in Assumption 3 of Shimotsu and Phillips (2005) is satisfied.

In light of our Assumptions 1 and 2, it follows that also Assumptions 1 and 2 of Shimotsu and Phillips (2005) are satisfied. Their Assumptions 4 and 5 are identical to ours, and it thus follows that all their consistency assumptions are satisfied for each component of X_t, i.e. that $\hat{d}_a \to_p d_a$ for $a = 1, ..., p$. ■

7.2 Proof of Theorem 2

By Theorem 1, with probability one as $n \to \infty$, $\hat{d}_a$ satisfies

$$0 = R_a^{(1)}(\hat{d}_a) = R_a^{(1)}(d_a) + R_a^{(2)}(\bar{d}_a)(\hat{d}_a - d_a),$$

where $R_a^{(i)}(\xi) = \frac{\partial^i R_a(\xi)}{\partial \xi^i}$ and $\left|\bar{d}_a - d_a\right| \leq \left|\hat{d}_a - d_a\right|$. We need to show that

$$\begin{aligned} \sqrt{m}\left(R_1^{(1)}(d_1), \ldots, R_p^{(1)}(d_p)\right)' &\to {}_d N\left(0, 4D^{-1}\left(G \circ G\right) D^{-1}\right), & (18) \\ R_a^{(2)}(\bar{d}_a) &\to {}_p 4, \quad a = 1, \ldots p. & (19) \end{aligned}$$

To show (19) we note that, defining $\hat{G}_{aa}^{(i)}(\xi) = \frac{\partial^i \hat{G}_{aa}(\xi)}{\partial \xi^i}$,

$$R_a^{(2)}(\xi) = \frac{\hat{G}_{aa}^{(2)}(\xi) \hat{G}_{aa}(\xi) - \hat{G}_{aa}^{(1)}(\xi)^2}{\hat{G}_{aa}(\xi)^2},$$

and follow the arguments in Shimotsu and Phillips (2005, pp. 1912-1916).

For (18) we apply the Cramér-Wold device and examine

$$\sum_{a=1}^{p} \eta_a \sqrt{m} R_a^{(1)}(d_a) = \sum_{a=1}^{p} \eta_a \sqrt{m} \left[\frac{\hat{G}_{aa}^{(1)}(d_a)}{\hat{G}_{aa}(d_a)} - \frac{2}{m} \sum_{j=1}^{m} \log \lambda_j \right].$$

Following Shimotsu and Phillips (2005, pp. 1916-1918) this expression is

$$\sum_{a=1}^{p} \eta_a \frac{2m^{-1/2} \sum_1^m \nu_j I_{u_a}(\lambda_j) + o_p(1)}{G_{aa} + o_p(1)}, \tag{20}$$

where $\nu_j = \log j - m^{-1} \sum_1^m \log j$. Since $\sum_1^m \nu_j = 0$, (20) is, apart from $o_p(1)$ terms,

$$\begin{aligned}
& \sum_{a=1}^{p} \eta_a \frac{2m^{-1/2}}{G_{aa}} \sum_{j=1}^{m} \nu_j \left(I_{u_a}(\lambda_j) - G_{aa} \right) \\
= & \sum_{a=1}^{p} \eta_a \frac{2}{m^{1/2}} \sum_{j=1}^{m} \frac{\nu_j}{G_{aa}} \left(C_a(\lambda_j) I_\varepsilon(\lambda_j) C_a(\lambda_j)^* - G_{aa} \right) \\
= & \sum_{a=1}^{p} \eta_a \frac{2}{m^{1/2}} \sum_{j=1}^{m} \frac{\nu_j}{G_{aa}} \left(C_a(\lambda_j) \frac{1}{2\pi n} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t' C_a(\lambda_j)^* - G_{aa} \right) \qquad (21) \\
& + \sum_{a=1}^{p} \eta_a \frac{2}{m^{1/2}} \sum_{j=1}^{m} \frac{\nu_j}{G_{aa}} C_a(\lambda_j) \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s \neq t} \varepsilon_t \varepsilon_s' e^{-i(t-s)\lambda_j} C_a(\lambda_j)^*, \qquad (22)
\end{aligned}$$

where $C_a(\lambda)$ denotes the a'th row of $C(\lambda) = \sum_{j=1}^{\infty} C_j e^{i\lambda_j}$. Rewrite (21) as

$$\begin{aligned}
& \sum_{a=1}^{p} \eta_a \frac{2}{m^{1/2}} \sum_{j=1}^{m} \frac{\nu_j}{G_{aa}} \left(\frac{1}{2\pi} C_a(\lambda_j) C_a(\lambda_j)^* - G_{aa} \right) \qquad (23) \\
& + \sum_{a=1}^{p} \eta_a \frac{2}{m^{1/2}} \sum_{j=1}^{m} \frac{\nu_j}{G_{aa}} \left(\frac{1}{2\pi} C_a(\lambda_j) \left(\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t' - I_p \right) C_a(\lambda_j)^* \right). \qquad (24)
\end{aligned}$$

By Assumption 1′ (b), (23) is $O\left(m^{-1/2} \sum_1^m \nu_j \lambda_j^\beta\right) = O\left(\frac{m^{1+2\beta}}{n^{2\beta}} (\log m)^2\right)$ and, since $\varepsilon_t \varepsilon_t' - I_p$ is a martingale difference sequence, (24) is $o_p\left(m^{-1/2} \sum_1^m \nu_j (1 + O(\lambda_j^\beta))\right) = o_p\left(\frac{m^{1+2\beta}}{n^{2\beta}} (\log m)^2\right)$.

Now, equation (22) can be written as $\sum_{t=1}^{n} z_{tn}$, where

$$\begin{aligned}
z_{tn} &= \varepsilon_t' \sum_{s=1}^{t-1} c_{t-s,n} \varepsilon_s, \\
c_{tn} &= \frac{1}{\pi n \sqrt{m}} \sum_{j=1}^{m} \nu_j \theta_j \cos(t\lambda_j), \\
\theta_j &= 2 \sum_{a=1}^{p} \eta_a G_{aa}^{-1} \operatorname{Re}\left(C_a(\lambda_j)' \bar{C}_a(\lambda_j) \right),
\end{aligned}$$

$\bar{C}_a(\lambda_j)$ is conjugate of $C_a(\lambda_j)$, and z_{tn} is a martingale difference sequence. Thus, we can apply the central limit theorem of Brown (1971) if

$$\sum_{t=1}^{n} E\left(z_{tn}^2 \middle| \mathcal{F}_{t-1}\right) - \sum_{a=1}^{p}\sum_{b=1}^{p} \eta_a \eta_b \frac{4G_{ab}^2}{G_{aa}G_{bb}} \rightarrow {}_p 0, \tag{25}$$

$$\sum_{t=1}^{n} E\left(z_{tn}^4\right) \rightarrow 0. \tag{26}$$

The proofs of (25) and (26) follow those in Lobato (1999, pp. 142-143).

The second statement of the theorem follows from a simple multivariate extension of the arguments in Shimotsu and Phillips (2005, pp. 1912-1916). ■

7.3 Proof of Lemma 4

Observe that

$$m_1^{1/2}(\hat{G}(d_*) - G) = \frac{1}{\sqrt{m_1}} \sum_{j=1}^{m_1} [\operatorname{Re} I_u(\lambda_j) - G] = I + II + III,$$

where

$$\begin{aligned} I &= \frac{1}{\sqrt{m_1}} \sum_{j=1}^{m_1} \operatorname{Re}\left[I_u(\lambda_j) - C(e^{i\lambda_j}) I_\varepsilon(\lambda_j) C(e^{i\lambda_j})^*\right], \\ II &= \frac{1}{\sqrt{m_1}} \sum_{j=1}^{m_1} \operatorname{Re}\left[C(e^{i\lambda_j}) \left(I_\varepsilon(\lambda_j) - I_p/2\pi\right) C(e^{i\lambda_j})^*\right], \\ III &= \frac{1}{\sqrt{m_1}} \sum_{j=1}^{m_1} \operatorname{Re}\left[f_u(\lambda_j) - G\right]. \end{aligned}$$

III is $O(m_1^{\beta+1/2} n^{-\beta})$ by Assumption $1'$. For I and II, Robinson and Yajima (2002, pp. 237-238) show that

$$I = o_p(1), \quad \operatorname{vec}(II) \rightarrow_d N(0, \tfrac{1}{2}(G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p)),$$

giving the required result. ■

7.4 Proof of Lemma 5

From Lemma 4, if suffices to show $\hat{G}(\bar{d}_*) - \hat{G}(d_*) = o_p(m_1^{-1/2})$. With a slight abuse of notation, for a scalar variable d define $M = \{d : m^{1/2}|d - d_*| \leq \log n\}$. From Theorem 2, $\Pr(\bar{d}_* \notin M) \rightarrow 0$ as $n \rightarrow \infty$. Therefore, for any $\varepsilon > 0$,

$$\Pr(m_1^{1/2}||\hat{G}(\bar{d}_*) - \hat{G}(d_*)|| > \varepsilon) = \Pr(m_1^{1/2}||\hat{G}(\bar{d}_*) - \hat{G}(d_*)|| > \varepsilon, \bar{d}_* \in M) + o(1).$$

Thus we assume $\bar{d}_* \in M$ in the following. Define $\theta = d - d_*$, then we may rewrite M in terms of θ as $M = \{\theta : m^{1/2}|\theta| \le \log n\}$. Because $m_1^{1/2}[\hat{G}(\bar{d}_*) - \hat{G}(d_*)] = m_1^{-1/2}\sum_{j=1}^{m_1}\mathrm{Re}[I_{\Delta(L;\theta,\ldots,\theta)u}(\lambda_j) - I_u(\lambda_j)]$, the desired result follows if, for $a, b = 1, \ldots, p$,

$$\sup_{\theta\in M}\left|R_n^{ab}(\theta)\right| \to_p 0, \quad \text{as } n \to \infty,$$

where $R_n^{ab}(\theta) = m_1^{-1/2}\sum_{j=1}^{m_1}\mathrm{Re}[w_{\Delta^\theta u_a}(\lambda_j)w^*_{\Delta^\theta u_b}(\lambda_j) - w_{u_a}(\lambda_j)w^*_{u_b}(\lambda_j)]$. Applying Lemma 5.1 (a) of Shimotsu and Phillips (2005) to $(\Delta^\theta u_{at}, u_{at})$ and reversing the role of X_t and u_t, we obtain

$$w_{\Delta^\theta u_a}(\lambda_j) = D_n(e^{i\lambda_j};\theta)w_{u_a}(\lambda_j) - (2\pi n)^{-1/2}\tilde{U}_{a,\lambda_j n}(\theta),$$

where $D_n(e^{i\lambda};\theta) = \sum_{k=0}^{n}(-\theta)_k e^{ik\lambda}/k!$, $\tilde{U}_{a,\lambda n}(\theta) = \sum_{p=0}^{n-1}\tilde{\theta}_{\lambda p}e^{-ip\lambda}u_{a,n-p}$, and $\tilde{\theta}_{\lambda p} = \sum_{k=p+1}^{n}(-\theta)_k e^{ik\lambda}/k!$. As shown in the proof of Theorem 1, each component of u_t satisfies the consistency assumptions in Shimotsu and Phillips (2005), and we have $E\sup_{\theta\in M}|n^{\theta-1/2}j^{1/2-\theta}\tilde{U}_{a,\lambda_j n}(\theta)|^2 = O((\log n)^2)$. Since

$$|n^\theta - 1| \le n^{|\theta|}(\log n)|\theta| = O(m^{-1/2}(\log n)^2), \quad \theta \in M, \tag{27}$$

by the mean value theorem, it follows that $E\sup_{\theta\in M}|(2\pi n)^{-1/2}\tilde{U}_{a,\lambda_j n}(\theta)|^2 = O(j^{-1}(\log n)^2)$ for $j = 1, \ldots, m_1$. Therefore,

$$\begin{aligned}
R_n^{ab}(\theta) \quad = \quad & \frac{1}{\sqrt{m_1}}\sum_{j=1}^{m_1}\mathrm{Re}\left[\left(|D_n(e^{i\lambda_j};\theta)|^2 - 1\right)w_{u_a}(\lambda_j)w^*_{u_b}(\lambda_j)\right] \\
& -\frac{1}{\sqrt{m_1}}\sum_{j=1}^{m_1}\mathrm{Re}\left[D_n(e^{i\lambda_j};\theta)w_{u_a}(\lambda_j)\,(2\pi n)^{-1/2}\,\tilde{U}^*_{b,\lambda_j n}(\theta)\right] \\
& -\frac{1}{\sqrt{m_1}}\sum_{j=1}^{m_1}\mathrm{Re}\left[D_n(e^{i\lambda_j};\theta)^* w^*_{u_b}(\lambda_j)\,(2\pi n)^{-1/2}\,\tilde{U}_{a,\lambda_j n}(\theta)\right] + o_p(1),
\end{aligned} \tag{28}$$

where the $o_p(1)$ term is uniform in $\theta \in M$. Lemma 5.2 of Shimotsu and Phillips (2005) gives, uniformly in $\theta \in M$,

$$\lambda_j^{-2\theta}|D_n(e^{i\lambda_j};\theta)|^2 - 1 = O(\lambda_j^2) + O(j^{-1-\theta}). \tag{29}$$

In view of (27), (29), and $E|I_u(\lambda_j)| < \infty$ for $j = 1, \ldots, m_1$, the first term on the right hand side of (28) is $O_p(m_1^{5/2}n^{-2} + m_1^{-1/4} + m_1^{1/2}m^{-1/2}(\log n)^2) = o_p(1)$ uniformly in $\theta \in M$. The proof completes if we show that the second and third terms on the right hand side of (28) are $o_p(1)$ uniformly in $\theta \in M$. Equation (67) on page 1920 and equation (72) on page 1921 of Shimotsu and Phillips (2005) give the decomposition of $\tilde{U}_{a,\lambda_j n}(\theta)$ as

$$\begin{aligned}\tilde{U}_{a,\lambda_j n}(\theta) &= \sum_{p=0}^{n-2} b_{np}(\theta) \sum_{q=0}^{p} u_{a,n-q} \\ &\quad + \frac{(-\theta)_n}{n!} \left[\frac{e^{i\lambda_j}}{1-e^{i\lambda_j}} \sum_{k=1}^{n} u_k - \frac{e^{i\lambda_j}}{1-e^{i\lambda_j}} (2\pi n)^{1/2} w_{u_a}(\lambda_j) \right],\end{aligned}$$

where, as in Shimotsu and Phillips (2005),

$$b_{np}(\theta) = \sum_{k=p+1}^{n-1} \frac{(1+\theta)\Gamma(k-\theta)}{\Gamma(-\theta)\Gamma(k+2)} e^{i(k-p)\lambda_j}.$$

Using the product definition of the gamma function by Weierstrass, we can show $|z\Gamma(z)| \in (0,\infty)$ for $|z| \leq 1/2$. See, c.f., Chapter 12 of Whittaker and Watson (1927). Consequently, multiplying $b_{np}(\theta)$ by $1/\theta$ does not change the bound of $b_{np}(\theta)$ shown in equation (69) on page 1920 of Shimotsu and Phillips (2005). Furthermore, $(1/\theta)[(-\theta)_n/n!] = -\Gamma(n-\theta)/[\Gamma(1-\theta)\Gamma(n+1)] = O(n^{-\theta-1})$. Therefore, the argument on pages 1920-1921 of Shimotsu and Phillips (2005) carries through even if we multiply $\tilde{U}_{a,\lambda_j n}(\theta)$ by $1/\theta$, and it follows that

$$E \sup_{\theta \in M} |\theta^{-1} n^{-1/2} \tilde{U}_{a,\lambda_j n}(\theta)|^2 = O(j^{-1}(\log n)^2).$$

In conjunction with (27) and (29), we find that the second and third terms on the right hand side of (28) are, uniformly in $\theta \in M$,

$$O_p\left(\frac{1}{\sqrt{m_1}} \sum_{j=1}^{m_1} |\theta|\, j^{-1/2} \log n \right) = O_p(m^{-1/2}(\log n)^2) = o_p(1),$$

and we complete the proof. ■

Acknowledgements

We are grateful to P. M. Robinson and two anonymous referees for constructive and very useful suggestions and comments. This work was initiated in the spring of 2002 when both authors were visiting Yale University and the Cowles Foundation. We are very grateful for their hospitality.

References

[1] Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73** 5-59.

[2] Baillie, R. T. and Bollerslev, T. (1989). Common stochastic trends in a system of exchange rates. *Journal of Finance* **44** 167-181.

[3] Baillie, R. T. and Bollerslev, T. (1994). Cointegration, fractional cointegration, and exchange rate dynamics. *Journal of Finance* **49** 737-745.

[4] Breitung, J. and Hassler, U. (2002). Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics* **110** 167-185.

[5] Brown, B. M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics* **42** 59-66.

[6] Brunetti, C. and Gilbert, C. L. (2000). Bivariate FIGARCH and fractional cointegration. *Journal of Empirical Finance* **7** 509-530.

[7] Chen, W. W. and Hurvich, C. M. (2003*a*). Semiparametric estimation of multivariate fractional cointegration. *Journal of the American Statistical Association* **98** 629-642.

[8] Chen, W. W. and Hurvich, C. M. (2003*b). Estimating fractional cointegration in the presence of polynomial trends. *Journal of Econometrics* **117** 95-121.

[9] Chen, W. W. and Hurvich, C. M. (2004). Semiparametric estimation of fractional cointegrating subspaces. Working paper, New York University.

[10] Cheung, Y. E. and Lai, K. S. (1993). A fractional cointegration analysis of purchasing power parity. *Journal of Business & Economic Statistics* **11** 103-122.

[11] Christensen, B. J. and Nielsen, M. Ø. (2004). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. Available at http://www.people.cornell.edu/pages/mon2/. Forthcoming in *Journal of Econometrics.*

[12] Davidson, J. (2002). A model of fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics* **110** 187-212.

[13] Diebold, F. X., Gardeazabal, J., and Yilmaz, K. (1994). On cointegration and exchange rate dynamics. *Journal of Finance* **49** 727-735.

[14] Dolado, J. J. and Marmol, F. (2004). Asymptotic inference results for multivariate long memory processes. *Econometrics Journal* **7** 168-190.

[15] Dueker, M. and Startz, R. (1998). Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. *Review of Economics & Statistics* **83** 420-426.

[16] Engle, R. and Granger, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* **55** 251-276.

[17] Fujikoshi, Y. (1985). Two methods for estimation of dimensionality in canonical correlation analysis and the multivariate linear model. In *Statistical Theory and Data Analysis* (L. Matusita, ed.) 233-240. Elsevier, Amsterdam.

[18] Fujikoshi, Y. and Veitch, L. G. (1979). Estimation of dimensionality in canonical correlation analysis. *Biometrika* **66** 345-351.

[19] Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16** 121-130.

[20] Gunderson, B. K. and Muirhead, R. J. (1997). On estimating the dimensionality in canonical correlation analysis. *Journal of Multivariate Analysis* **62** 121-136.

[21] Henry, M. and Zaffaroni, P. (2003). The long range dependence paradigm for macroeconomics and finance. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Oppenheim, and M. S. Taqqu, eds.) 417-438. Birkhäuser, Boston.

[22] Hassler, U. and Breitung, J. (2005). A residual-based LM type test against fractional cointegration. Forthcoming in *Econometric Theory.*

[23] Hassler, U., Marmol, F., and Velasco, C. (2006). Residual log-periodogram inference for long-run relationships. *Journal of Econometrics* **130** 165-207.

[24] Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics & Control* **12** 231-254.

[25] Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59** 1551-1580.

[26] Kim, C. S. and Phillips, P. C. B. (2001). Fully modified estimation of fractional cointegration models. Working paper, Yale University.

[27] Krämer, W. and Marmol, F. (2004). The power of residual-based tests for cointegration when residuals are fractionally integrated. *Economics Letters* **82** 63-69.

[28] Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society* (Y. Prokhorov and V. V. Sazanov, eds.) **1** 67-74. VNU Science Press, Utrecht.

[29] Lobato, I. N. (1999). A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics* **90** 129-153.

[30] Lobato, I. N. and C. Velasco (2000). Long memory in stock-market trading volume. *Journal of Business & Economic Statistics* **18** 410-427.

[31] Marinucci, D. and Robinson, P. M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning & Inference* **80** 111-122.

[32] Marinucci, D. and Robinson, P. M. (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics* **105** 225-247.

[33] Marmol, F. and Velasco, C. (2004). Consistent testing of cointegrating relationships. *Econometrica* **72** 1809-1844.

[34] Nielsen, M. Ø. (2004*a*). Spectral analysis of fractionally cointegrated systems. *Economics Letters* **83** 225-231.

[35] Nielsen, M. Ø. (2004*b*). Optimal residual based tests for fractional cointegration and exchange rate dynamics. *Journal of Business & Economic Statistics* **22** 331-345.

[36] Phillips, P. C. B. and Ouliaris, S. (1988). Testing for cointegration using principal components methods. *Journal of Economic Dynamics & Control* **12** 205-230.

[37] Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **23** 1630-1661.

[38] Robinson, P. M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71** 1727-1766.

[39] Robinson, P. M. and Marinucci, D. (2003). Semiparametric frequency domain analysis of fractional cointegration. In *Time Series With Long Memory* (P. M. Robinson, ed.) 334-373. Oxford University Press, Oxford.

[40] Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics* **106** 217-242.

[41] Robinson, P. M. (2005). The distance between rival nonstationary fractional processes. *Journal of Econometrics* **128** 195-236.

[42] Shimotsu, K. (2004). Exact local Whittle estimation of fractional integration with unknown mean and time trend. Working paper, Queen's University. Available at http://qed.econ.queensu.ca/pub/faculty/shimotsu/.

[43] Shimotsu, K. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics* **33** 1890-1933.

[44] Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20** 87-127.

[45] Velasco, C. (2003*a*). Gaussian semi-parametric estimation of fractional cointegration. *Journal of Time Series Analysis* **24** 345-378.

[46] Velasco, C. (2003*b*). Nonparametric frequency domain analysis of nonstationary multivariate time series. *Journal of Statistical Planning and Inference* **116** 209-247.

[47] Whittaker, E. T. and Watson, G. N. (1927). *A Course of Modern Analyis*, 4th edition, Cambridge University Press.

Table 1. Simulation results for $n = 128$

	$r=0$	$r=1$	$r=2$	$r=3$
			$b = 0.2$	
freq($\hat{\pi}_r$)	(0,0,0)	(132,0,0)	(38,0,0)	(181,0,0)
freq($\hat{r}$)	(981,19,0,0)	(0,997,3,0)	(0,120,880,0)	(0,170,617,213)
rej($\hat{T}_0$)	0.078, 0.155	0.063, 0.163	0.065, 0.170	0.048, 0.155
			$b = 0.4$	
freq($\hat{\pi}_r$)	(0,0,0)	(304,0,0)	(123,0,0)	(362,0,0)
freq($\hat{r}$)	(981,19,0,0)	(0,998,2,0)	(0,29,971,0)	(0,38,446,516)
rej($\hat{T}_0$)	0.078, 0.155	0.067, 0.173	0.074, 0.210	0.072, 0.222
			$b = 0.6$	
freq($\hat{\pi}_r$)	(0,0,0)	(456,0,0)	(240,0,0)	(485,0,0)
freq($\hat{r}$)	(981,19,0,0)	(0,998,2,0)	(0,7,993,0)	(0,9,214,777)
rej($\hat{T}_0$)	0.078, 0.155	0.070, 0.192	0.088, 0.230	0.087, 0.309
			$b = 0.8$	
freq($\hat{\pi}_r$)	(0,0,0)	(577,0,0)	(407,0,0)	(619,0,0)
freq($\hat{r}$)	(981,19,0,0)	(0,997,3,0)	(0,3,997,0)	(0,1,83,916)
rej($\hat{T}_0$)	0.078, 0.155	0.070, 0.191	0.096, 0.243	0.092, 0.363

Note: freq($\hat{r}$) denotes frequency of ($\hat{r} = 0, \hat{r} = 1, \hat{r} = 2, \hat{r} = 3$), freq($\hat{\pi}_r$) denotes frequency of $CI(0.05, r) < 0.1/p$ for ($r = 1, r = 2, r = 3$), and rej($\hat{T}_0$) denotes the rejection frequency of $\hat{T}_0$ with the 5% asymptotic critical value and $h(n) = 1/(\log n)^{1/2}$ and $h(n) = 1/\log n$, resp.

Table 2. Simulation results for $n = 512$

	$r = 0$	$r = 1$	$r = 2$	$r = 3$
			$b = 0.2$	
freq($\hat{\pi}_r$)	(0,0,0)	(475,0,0)	(93,0,0)	(443,0,0)
freq($\hat{r}$)	(1000,0,0,0)	(0,1000,0,0)	(0,105,895,0)	(0,356,555,89)
rej($\hat{T}_0$)	0.046, 0.099	0.036, 0.110	0.052, 0.147	0.068, 0.222
			$b = 0.4$	
freq($\hat{\pi}_r$)	(0,0,0)	(956,0,0)	(658,0,0)	(919,0,0)
freq($\hat{r}$)	(1000,0,0,0)	(0,1000,0,0)	(0,1,999,0)	(0,1,155,844)
rej($\hat{T}_0$)	0.046, 0.099	0.041, 0.120	0.058, 0.170	0.116, 0.409
			$b = 0.6$	
freq($\hat{\pi}_r$)	(0,0,0)	(995,0,0)	(970,0,0)	(998,0,0)
freq($\hat{r}$)	(1000,0,0,0)	(0,1000,0,0)	(0,0,1000,0)	(0,0,1,999)
rej($\hat{T}_0$)	0.046, 0.099	0.043, 0.119	0.050, 0.165	0.114, 0.517
			$b = 0.8$	
freq($\hat{\pi}_r$)	(0,0,0)	(998,0,0)	(998,4,0)	(1000,2,0)
freq($\hat{r}$)	(1000,0,0,0)	(0,1000,0,0)	(0,0,1000,0)	(0,0,0,1000)
rej($\hat{T}_0$)	0.046, 0.099	0.039, 0.110	0.041, 0.151	0.072, 0.534

Note: freq($\hat{r}$) denotes frequency of ($\hat{r} = 0, \hat{r} = 1, \hat{r} = 2, \hat{r} = 3$), freq($\hat{\pi}_r$) denotes frequency of $CI(0.05, r) < 0.1/p$ for ($r = 1, r = 2, r = 3$), and rej($\hat{T}_0$) denotes the rejection frequency of $\hat{T}_0$ with the 5% asymptotic critical value and $h(n) = 1/(\log n)^{1/2}$ and $h(n) = 1/\log n$, resp.

Table 3. Feasible ELW estimates of fractional integration orders for log exchange rates

Bandwidth	CAN	SWE	FRA	GER	ITA	JPN	UK	$\bar{d}_*$
$m = \left[n^{0.6}\right] = 32$	1.1614 (0.0884)	1.0064 (0.0884)	1.1080 (0.0884)	1.0392 (0.0884)	1.0759 (0.0884)	0.9621 (0.0884)	0.9847 (0.0884)	1.0482
$m = \left[n^{0.5}\right] = 18$	1.2055 (0.1179)	1.1138 (0.1179)	1.2145 (0.1179)	1.2076 (0.1179)	1.1429 (0.1179)	1.0996 (0.1179)	0.9098 (0.1179)	1.1277

Note: Standard errors are given in parenthesis, see Shimotsu and Phillips (2005) and Shimotsu (2004). A nonzero mean was allowed in the estimation, c.f. (11).

Table 4. Estimated eigenvalues of $10,000 \times \hat{G}(\bar{d}_*)$ and $\hat{P}(\bar{d}_*)$ for log exchange rates

Bandwidth	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$	$\hat{\delta}_5$	$\hat{\delta}_6$	$\hat{\delta}_7$
	Eigenvalues of $10,000 \times \hat{G}(\bar{d}_*)$						
$m_1 = \left[n^{0.55}\right] = 24, m = 32$	7.0356	1.7711	0.9248	0.3588	0.1339	0.1086	0.0697
$m_1 = \left[n^{0.45}\right] = 13, m = 18$	6.1708	1.5718	1.3095	0.2424	0.1080	0.0746	0.0389
	Eigenvalues of $\hat{P}(\bar{d}_*)$						
$m_1 = \left[n^{0.55}\right] = 24, m = 32$	4.2937	1.0300	0.8235	0.5004	0.2134	0.0835	0.0554
$m_1 = \left[n^{0.45}\right] = 13, m = 18$	4.2012	1.2079	0.8669	0.4497	0.1633	0.0639	0.0470

Note: The estimation allowed for a nonzero mean as in Section 3.4 and Table 3.

Table 5. Rank estimates for log exchange rates using the model selection procedure with $\hat{P}(\bar{d}_{*})$

$L(u)$	$v(n) = m_1^{-0.45}$	$v(n) = m_1^{-0.35}$	$v(n) = m_1^{-0.25}$	$v(n) = m_1^{-0.15}$	$v(n) = m_1^{-0.05}$
			$m_1 = 24, m = 32$		
$L(0)$	-5.3250	-4.6984	-3.8374	-2.6542	-1.0284
$L(1)$	-5.5089	-4.9718	-4.2338	-3.2196	-1.8261
$L(2)$	-5.6647	-5.2171	-4.6021	-3.7570	-2.5957
$L(3)$	-5.6905	-5.3325	-4.8405	-4.1644	-3.2353
$L(4)$	-5.4294	-5.1608	-4.7918	-4.2848	-3.5880
$L(5)$	-4.8452	-4.6661	-4.4201	-4.0821	-3.6176
$L(6)$	-4.0545	-3.9649	-3.8419	-3.6729	-3.4407
$\hat{r}$	3	3	3	4	5
			$m_1 = 13, m = 18$		
$L(0)$	-4.7929	-4.1476	-3.3135	-2.2356	-0.8426
$L(1)$	-5.0612	-4.5080	-3.7931	-2.8692	-1.6752
$L(2)$	-5.3126	-4.8516	-4.2559	-3.4859	-2.4909
$L(3)$	-5.4645	-5.0958	-4.6192	-4.0032	-3.2072
$L(4)$	-5.3301	-5.0536	-4.6961	-4.2342	-3.6371
$L(5)$	-4.7785	-4.5941	-4.3558	-4.0478	-3.6498
$L(6)$	-3.8859	-3.7937	-3.6745	-3.5205	-3.3215
$\hat{r}$	3	3	4	4	5

Note: The model selection procedure determines $\hat{r}$ as the arg min of $L(u)$, and the calculation of $L(u)$ allowed for a nonzero mean as in Tables 3 and 4.

Table 6. Parametric rank tests for log exchange rates

	Johansen (1988, 1991) trace tests					Breitung and Hassler (2002) tests				
r	no lags	3 lags	6 lags	12 lags	95% C.V.	no lags	3 lags	6 lags	12 lags	95% C.V.
0	112.02	108.18	114.52	128.19	123.04	288.26	142.39	90.34	83.91	66.34
1	65.88	69.58	73.55	86.71	93.92	111.77	43.25	47.88	46.01	51.00
2	33.92	41.16	47.69	53.57	68.68	80.86	25.19	24.81	18.82	37.65
3	14.82	18.14	23.66	29.71	47.21	54.28	15.44	11.59	9.58	26.30
4	8.07	9.28	8.41	13.64	29.38	31.29	6.89	4.41	4.92	16.92
5	2.76	2.53	3.38	6.32	15.34	12.83	1.75	1.01	.54	9.49
6	0.44	0.43	0.33	0.22	3.84	3.36	0.05	0.11	0.15	3.84
Rank	0	0	0	1	–	6	1	1	1	–

Note: The Johansen (1988, 1991) trace tests were calculated with an unrestricted constant term, and the Breitung and Hassler (2002) tests allowed for a nonzero mean in the levels of the time series.

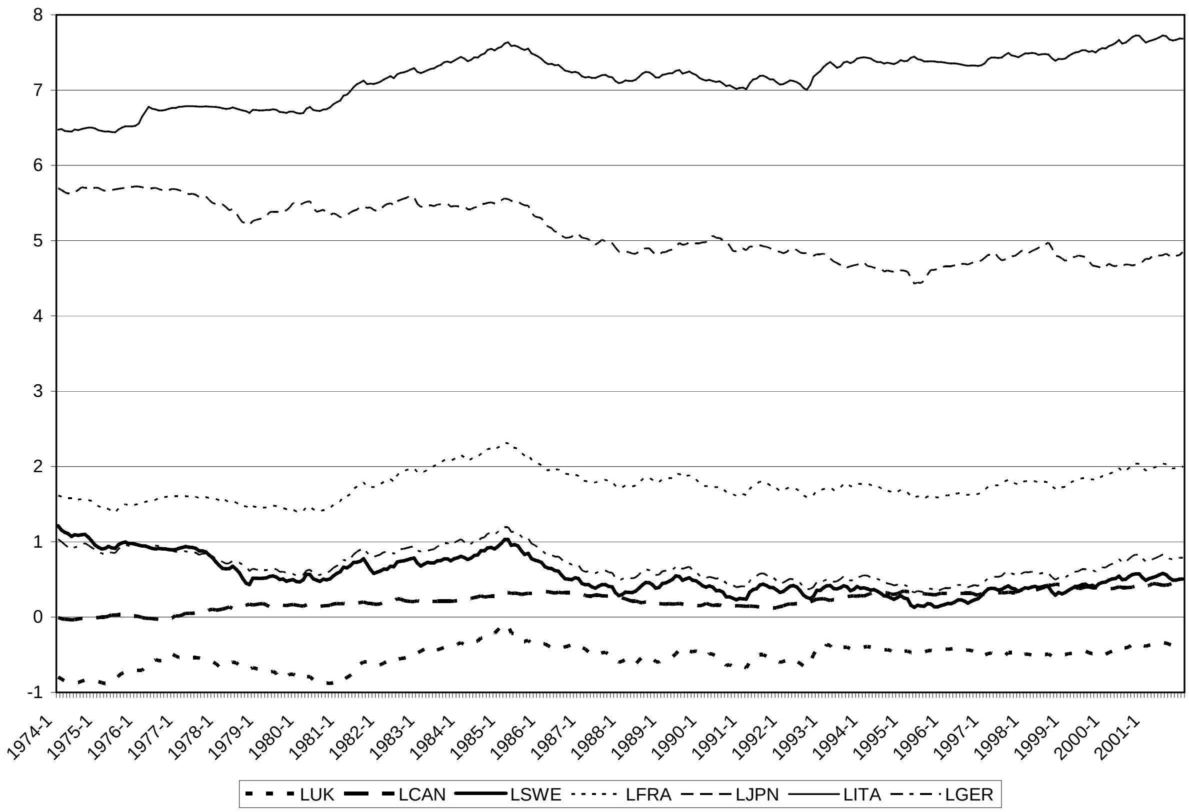

Figure 1: Time series plot of log exchange rate data