Queen's Economics Department Working Paper No. 1174

Nonparametric Cointegration Analysis of Fractional Systems With Unknown Integration Orders

Morten Ørregaard Nielsen
Queen's University and CREATES

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

7-2008

Nonparametric Cointegration Analysis of Fractional Systems With Unknown Integration Orders

Morten Ørregaard Nielsen*
Queen's University and CREATES

October 1, 2009

Abstract

In this paper a nonparametric variance ratio testing approach is proposed for determining the cointegration rank in fractionally integrated systems. The test statistic is easily calculated without prior knowledge of the integration order of the data, the strength of the cointegrating relations, or the cointegration vector(s). The latter property makes it easier to implement than regression-based approaches, especially when examining relationships between several variables with possibly multiple cointegrating vectors. Since the test is nonparametric, it does not require the specification of a particular model and is invariant to short-run dynamics. Nor does it require the choice of any smoothing parameters that change the test statistic without being reflected in the asymptotic distribution. Furthermore, a consistent estimator of the cointegration space can be obtained from the procedure. The asymptotic distribution theory for the proposed test is non-standard but easily tabulated or simulated. Monte Carlo simulations demonstrate excellent finite sample properties, even rivaling those of well-specified parametric tests. The proposed methodology is applied to the term structure of interest rates, where, contrary to both fractional and integer-based parametric approaches, evidence in favor of the expectations hypothesis is found using the nonparametric approach.

JEL Classification: C32.

Keywords: Cointegration rank, cointegration space, fractional integration and cointegration, interest rates, long memory, nonparametric, term structure, variance ratio.

Short title: Nonparametric Fractional Cointegration Analysis.

1 Introduction

Traditionally, cointegration analysis has been developed mostly in the context of processes that are integrated of integer orders, and the most popular case is where observed series are each assumed to have a single unit root. In that case, first differencing an individual series or taking a

*Mailing address: Department of Economics, Dunning Hall Room 307, 94 University Avenue, Queen's University, Kingston, Ontario K7L 3N6, Canada; phone: +1 613 533 2262; fax: +1 613 533 6668; email: mon@econ.queensu.ca

cointegrating linear combination of two or more series yields a weakly dependent stationary and invertible process. A recent direction of this research area both in theoretical and empirical work is to allow the time series in question to be fractionally integrated, i.e. they may be integrated of an unknown, possibly non-integer order. Indeed, the original formulation by Granger (1981) accommodated this more general class of processes.

For economic or financial time series, cointegrating relations define long-run equilibria between the variables in the sense that the time series are tied together by one or more linear combinations that reduce the memory or persistence in the series. Models of cointegration are standard in modern empirical macroeconomics and financial economics since cointegration in a simple and coherent way bridges economic theory (long-run equilibria between variables) and empirical econometric analysis.

To fix ideas, first the concepts of fractional integration and cointegration are defined for a vector time series.

Definition 1 *The n-vector time series Z_t is fractionally integrated of order d, denoted $Z_t \in I(d)$, if*

$$Z_t = \Delta_+^{-d} v_t, \quad t = 1, 2, \ldots, \tag{1}$$

where v_t has continuous spectral density matrix that is bounded, positive semi-definite, and bounded away from the zero matrix at all frequencies.

Definition 2 *The n-vector time series Z_t is cointegrated if $Z_t \in I(d)$, but there exists a full rank $n \times r$ matrix β such that $\beta' Z_t \in I(d-b)$ for $b > 0$. The number r is the cointegration rank and the space spanned by the columns of β is the cointegration space.*

For any (vector or scalar) time series x_t, the operator Δ_+ is defined by

$$\Delta_+^{-d} x_t = (1-L)_+^{-d} x_t = \sum_{j=0}^{t-1} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} x_{t-j}$$

so that only values corresponding to a positive time-index (denoted by the subscripted Δ_+) enters the expression, which is a truncated version of the binomial expansion of $(1-L)^{-d}$. The operator Δ_+ has the advantage that the right-hand side of (1) is well-defined for any value of d, see for instance Marinucci & Robinson (2000) who call this a "type II" process.

The truncation in the Δ_+ operator in (1) implies that a time series that is $I(d)$ according to Definition 1 is not stationary for any value of $d \neq 0$. However, when $d < 1/2$ the time series is asymptotically stationary. The remainder of this paper applies the terms stationary and asymptotically stationary synonymously when this should cause no confusion.

Definition 1 defines an $I(d)$ process to be one whose d'th difference has continuous spectral density matrix that is bounded, positive semi-definite, and bounded away from the zero matrix, i.e. all eigenvalues of the spectral density matrix are non-negative and bounded and at least one eigenvalue is bounded away from zero. Definition 1 applies to both scalar and vector time series Z_t. In the scalar case, an $I(0)$ process is defined as having spectral density function that is continuous, bounded, and bounded away from zero. In the vector case, the assumption of positive semi-definiteness (as opposed to positive definiteness) allows some of the elements of $v_t = \Delta_+^d Z_t$ to be over-differenced. At the same time, the assumption that the spectral density matrix is bounded away from the zero matrix ensures that at least one element of v_t is not over-differenced. Thus, the definition implies that if the components of the vector time series Z_t are integrated of different orders, then $Z_t \in I(d)$, where d is the highest order of integration of the components of Z_t. Note that Definition 1 has the advantage that it is invariant to non-singular linear transformations. If instead, e.g. Engle & Granger (1987), Z_t were defined to be $I(d)$ if and only if all the components of Z_t are themselves $I(d)$, that definition would not be invariant to non-singular linear transformations when the components of Z_t are cointegrated. This is exemplified by Definition 2 where a non-singular linear transformation including β as columns would result in some elements of the transformed vector being $I(d-b)$ and hence the transformed vector would not be $I(d)$ according to the alternative definition. In the integer integration case, Definition 1 has been applied by Johansen (1988, 1991) and Breitung (2002), among others. However, the methodology in this paper also applies with little modification under the alternative definition of fractional integration for vector time series. Indeed, it is no problem to pre-test or pre-estimate the fractional integration orders of the variables, and then choose to include in the analysis only variables with the same integration order. This would have no influence on the asymptotic distribution of the rank test. For a discussion of several different definitions of fractional integration (for vectors) and cointegration, see Robinson & Yajima (2002).

The cointegration concept in Definition 2 is called fractional cointegration if d and/or b are non-integer valued, and the "traditional model" appears as the special case $d = b = 1$. Hence, fractional cointegration generalizes the conventional $I(1) - I(0)$ cointegration framework where $d = b = 1$ by allowing both d and b to be real numbers. The ability to accommodate non-integer values of the fractional differencing parameters, d and b, is very attractive because many economic and financial time series are known to exhibit nonstationarity that may not be exactly $I(1)$, and even more importantly there is no strong a priori reason to assume that the unobservable equilibrium relation is exactly $I(0)$. Thus, fractional cointegration 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.

Recently, fractional integration and cointegration is attracting increasing attention from both theoretical and empirical researchers in economics and finance. For a recent survey, see Henry & Zaffaroni (2003). Although often assumed to be one, a natural starting point in any econometric cointegration analysis is to determine the number of linearly independent cointegrating relations (equilibrium relations), i.e. to determine the cointegration rank. Standard techniques for determining cointegration rank were developed for traditional $I(1) - I(0)$ cointegration, e.g., Stock & Watson (1988) and Johansen (1988, 1991) in a vector autoregressive (VAR) framework, Phillips & Ouliaris (1988) and Harris (1997) using principal components methods, and Bierens (1997) and Breitung (2002) in a nonparametric framework. However, these methods rely on the assumption that $d = b = 1$, and are mostly inapplicable to the more general fractional cointegration considered here. In the fractional cointegration literature the focus has been mostly on estimation of the cointegrating relation, while assuming that there is only one of them. However, there has been some work on determining cointegration rank in fractional systems as discussed in section 3 below.

In this paper a nonparametric testing approach to determine the cointegration rank is proposed and analyzed. An integral part of this methodology is to allow the possibly fractional integration orders to be unknown. Hence, as argued recently by e.g. Robinson & Hualde (2003), if there is no strong a priori reason to pick any particular integration orders it is important to make only minimal assumptions on their values and possibly allow them to be estimated as part of the procedure. Similar nonparametric cointegration rank testing approaches were suggested by Bierens (1997) and Breitung (2002) in the traditional $I(1) - I(0)$ cointegration model but motivated by nonlinear cointegration alternatives. In the present fractional cointegration framework the proposed nonparametric approach has many additional advantages. For example, the test statistic does not depend on the integration order of the observed variables, d, and neither the test statistic nor its asymptotic distribution depend on the strength of the cointegrating relation(s), b. The latter property is especially useful in fractional cointegration models since b is unobserved and estimating b would require also estimating the cointegrating relations which in turn requires the cointegration rank to be specified. Furthermore, the approach does not require estimating the cointegration vector(s) and is thus easier to implement than regression-based approaches, especially when examining relationships between several variables with possibly multiple cointegrating vectors.

Because the proposed procedure is nonparametric it allows the cointegration rank to be determined without relying on the specification of a particular data generating process or model. This feature in particular distinguishes the approach from fully parametric cointegration rank testing approaches such as Johansen (1988, 1991) for traditional $I(1) - I(0)$ cointegration and Breitung & Hassler (2002) for fractional cointegration. Of course, this aspect is a consequence of the nonparametric nature of the variance ratio test statistic, and is empirically important because misspecified short-run dynamics leads to inconsistent estimation of the remainder of the model and hence to erroneous inferences on the cointegration rank. There is also no need to specify a bandwidth as in the nonparametric approach of Bierens (1997) or the recent semiparametric frequency domain approaches of Robinson & Yajima (2002), Chen & Hurvich (2003), and Nielsen & Shimotsu (2007). The approach proposed here is based on a family of test statistics indexed by a parameter d_1 defined below. Interestingly, this user-chosen parameter appears in the asymptotic distribution of the cointegration rank test statistic (unlike tuning parameters such as lag lengths, bandwidths, etc., in other approaches) which thus reflects the value of d_1.

Furthermore, estimation of the cointegration space is considered. As usual in the cointegration literature, only a basis for the space spanned by the cointegration vectors can be estimated. Since the proposed nonparametric framework is designed for testing cointegration rank, it does not deliver a straightforward distribution theory for the estimated cointegration space. However, it is shown that the estimated cointegration space is consistent in the sense that the angle between the estimated and the true cointegration spaces converges to zero.

To document the finite sample feasibility of the methods proposed in the paper, a simulation study is conducted. The simulations show that the nonparametric cointegration rank test is useful and has non-trivial power and good size in the models considered, even though the sample sizes ($T = 100$ and $T = 250$) would typically be considered quite small for nonparametric inference. In particular, the proposed test seems to be very robust to different specifications of the simulated model, including cointegrated alternatives of both fractional and autoregressive types. This is in contrast to parametric cointegration rank tests, whose performance is typically sensitive to both the underlying model and the lag-augmentation employed.

Finally, the proposed nonparametric methodology is applied to an analysis of the expectations hypothesis for the term structure of interest rates in a system of four interest rates of different maturities. The expectations hypothesis implies that $n-1$ cointegrating relations should exist among n interest rates of different maturities. Therefore, the expectations hypothesis provides a suitable application of the type of test proposed here since it is supported by rejection of a sequence of null hypotheses (that the cointegration rank is zero, then one, etc.), and thus the most powerful test should offer most support, assuming that the expectations hypothesis is true. Applying the nonparametric approach, there is clear evidence in favor of the implications of the expectations hypothesis. This contrasts with parametric approaches, both fractional and integer-based, from which no clear-cut conclusions can be drawn.

The paper proceeds as follows. The next section briefly introduces the nonparametric variance ratio testing approach in a univariate framework and the following section presents the cointegration rank test in a fractional cointegration framework. In section 4 estimation of the cointegration space is discussed and consistency (with rates) of the estimator is shown. Section 5 presents simulation evidence, section 6 presents the empirical application to the term structure of interest rates, and section 7 offers some concluding remarks. All proofs are gathered in the appendix.

The following notation is used throughout the paper. The bracket $\lfloor\cdot\rfloor$ denotes the integer part of the argument, and the norm $\|A\| = (\mathrm{tr}\,(A'A))^{1/2} = (\sum_{i=1}^{l}\sum_{j=1}^{m} A_{ij}^2)^{1/2}$ is the usual Euclidean norm of an $l \times m$ matrix A (if A is a scalar $\|A\|$ is the absolute value). The double arrow $\Longrightarrow$ means weak convergence of a process in $D\,[0,1]$ endowed with the Skorohod topology, whereas $\xrightarrow{D}$ and $\xrightarrow{P}$ denote convergence in distribution and probability, respectively, on $\mathbb{R}^{l\times m}$. The process W_d is type II fractional standard Brownian motion of order d ($> 1/2$), e.g. Marinucci & Robinson (2000), defined as

$$W_d\,(r) = 0, \text{ a.s.}, \; r = 0, \tag{2}$$

$$W_d\,(r) = \frac{1}{\Gamma\,(d)} \int_0^r (r-s)^{d-1}\, \mathsf{d}W_1\,(s)\,,\; r > 0. \tag{3}$$

Note that with this definition W_1 is the standard Brownian motion.

2 Variance Ratio Testing Approach

To illustrate the ideas behind the nonparametric variance ratio approach, first briefly consider a univariate version of the general test, i.e. a univariate test for fractional integration of a particular order such as unity (the unit root hypothesis). The test statistic is constructed as a ratio of variances of the observed series and its fractional partial sum, thereby canceling nuisance parameters from the limiting distribution. Specifically, consider the behavior of the observed univariate time series $\{z_t\}_{t=1}^{T}$ generated according to

$$z_t = \Delta_+^{-d} u_t, \quad d > 1/2, \quad t = 1, 2, \ldots, \tag{4}$$

and consider also its fractional partial sum

$$\tilde{z}_t = \Delta_+^{-d_1} z_t, \quad d_1 > 0, \quad t = 1, 2, \ldots. \tag{5}$$

It is well known that for $d > 1/2$ and under regularity conditions on u_t, a (fractional) functional central limit theorem is obtained for z_t, i.e.

$$T^{1/2-d} z_{\lfloor sT \rfloor} \Longrightarrow \sigma_z W_d(s), \quad 0 < s \leq 1, \tag{6}$$

as $T \to \infty$ for some $\sigma_z > 0$, and similarly for $\tilde{z}_t$.

It follows that the rescaled (uncentered) sample second moments of z_t and $\tilde{z}_t$ satisfy

$$T^{-2d} \sum_{t=1}^{T} z_t^2 \xrightarrow{D} \sigma_z^2 \int_0^1 W_d(s)^2 \, \mathsf{d}s \tag{7}$$

$$T^{-2(d+d_1)} \sum_{t=1}^{T} \tilde{z}_t^2 \xrightarrow{D} \sigma_z^2 \int_0^1 W_{d+d_1}(s)^2 \, \mathsf{d}s \tag{8}$$

as $T \to \infty$. The variance ratio statistic is

$$\rho(d_1) = T^{2d_1} \frac{\sum_{t=1}^{T} z_t^2}{\sum_{t=1}^{T} \tilde{z}_t^2} \xrightarrow{D} \frac{\int_0^1 W_d(s)^2 \, \mathsf{d}s}{\int_0^1 W_{d+d_1}(s)^2 \, \mathsf{d}s} \tag{9}$$

so the nuisance parameter σ_z^2 is eliminated from the asymptotic distribution without the need to estimate any parameters. The statistic $\rho(d_1)$ in (9) is the univariate variance ratio statistic. Note that (9) generalizes the idea of Vogelsang (1998*a*, 1998*b*) and Breitung (2002) who suggested using the ratio of the sample variance of z_t and that of the partial sum of z_t (i.e. $d_1 = 1$) to eliminate the nuisance parameter σ_z^2 and avoid estimation of serial correlation parameters.

The idea of using a variance ratio with fractional summation was previously explored by Nielsen (2008) who applied the univariate variance ratio statistic in (9) to test for a unit root in the usual autoregression, i.e. with $d = 1$. This paper further develops the idea and extends the variance ratio testing approach to allow for time series that are fractionally integrated, to testing cointegration rank in multivariate fractional systems, and also demonstrates that a consistent estimator of (a basis for) the associated cointegration space may be obtained from the procedure. These results are obtained without requiring that the fractional integration orders d and b, or the cointegration vectors, are known.

Note that d is the integration order of the observed data, whereas the parameter d_1 indexes the family of tests and is chosen by the econometrician. Since d_1 actually appears in the asymptotic distribution of the variance ratio statistic, unlike the tuning parameters (e.g. lag length or bandwidth parameters) in the usual Dickey-Fuller type unit root tests, it may be possible to tailor the test in such a way that power is maximized against relevant alternatives. Thus, Nielsen (2008) conducts an asymptotic local power analysis and demonstrates that the asymptotic local power of the variance ratio test is monotonic in d_1, and that $d_1 = 0.1$ appears to be a good choice in the sense that it has uniformly (in the local noncentrality parameter) higher power relative to higher values of d_1 and that the gain in power from choosing an even smaller value of d_1 is very minor. It also seems unwise to choose d_1 too small, since then d_1 acts as if it depends inversely on the sample size which may distort the size properties of the test. Another typical choice could be $d_1 = 1$, i.e. partial summation, based on computational simplicity. In the integer integration framework with

$d = 1$, the latter choice would in fact lead to the statistic suggested by Breitung (2002) to test for a unit root against nonlinear alternatives, see also Taylor (2005). However, that test can be improved upon, at least against linear alternatives, even within the usual class of autoregressive models by admitting non-integer values of $d_1 < 1$. In section 5 below, further evidence in favor of choosing $d_1 = 0.1$ relative to $d_1 = 1$ is presented based on simulation evidence from models of fractional cointegration as well as autoregressive models of $I(1) - I(0)$ cointegration.

3 Cointegration Rank Test

Suppose an observed n-vector time series $\{Z_t\}_{t=1}^T$ is cointegrated according to Definition 2. As discussed above, testing of cointegration rank is important in economic and financial applications to determine the number of linearly independent long-run equilibrium relationships, which are given by the columns of β.

The fractional cointegration literature has focused mostly on estimation of cointegration relations, although formal testing of cointegration rank has received some attention recently. On the parametric side, Breitung & Hassler (2002) use a score test motivation to develop a trace test for cointegration rank which is applicable when the time series in question may be fractionally integrated and cointegrated. However, the properties of their parametric fractional trace test depend on the correct specification of the short-run dynamics via a lag-augmentation, as would any parametric test. In addition, their test also depends on the specification of the integration order of the observed variables, d, and, although seemingly innocuous, estimation of this parameter in a preliminary step changes the asymptotic distribution of their test.

An alternative semiparametric frequency domain approach has been studied by Robinson & Yajima (2002), who exploited the fact that if Z_t is cointegrated the spectral density matrix of $v_t = (1 - L)^d Z_t$ has reduced rank at the origin, see also Phillips & Ouliaris (1988). The normalized eigenvalues of the averaged periodogram matrix of the differenced observations are used in a model selection procedure to estimate the cointegration rank. However, no formal testing procedure is available to find the rank (the testing procedure is only valid for $r = 0$), and the model selection procedure requires several bandwidth parameters to be chosen by the researcher which may induce additional uncertainty about the results obtained. Robinson & Yajima (2002) considered stationary variables with $d < 1/2$, and their approach was extended to accommodate both stationary and nonstationary data by Chen & Hurvich (2003, 2006) and Nielsen & Shimotsu (2007).

In this section, an alternative and completely nonparametric approach to cointegration rank testing is analyzed. The precise conditions on Z_t are collected in the following assumption.

Assumption 3 *There exists a full rank orthonormal $n \times n$ matrix $R = [R_{n-r}, R_r]$, where R_{n-r} has $n - r$ columns and R_r has r columns with $0 \leq r \leq n-1$, such that*

$$R' Z_t = \begin{bmatrix} \Delta_+^{-d} I_{n-r} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & \Delta_+^{-(d-b)} I_r \end{bmatrix} u_t, \quad t = 1, 2, \ldots, \tag{10}$$

where I_m is the m-dimensional identity matrix and $d - b < 1/2 < d$. If $r = 0$, $R = R_{n-r}$ and (10) *is* $R'Z_t = \Delta_+^{-d} u_t, t = 1, 2, \ldots$. *Here,* u_t *is generated by the linear process*

$$u_t = \Psi(L)\,\varepsilon_t = \sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-k}, \quad t = 1, 2, \ldots.$$

The $n \times n$ *coefficient matrices* Ψ_k *satisfy* $\sum_{k=0}^{\infty} k^{1/2} \|\Psi_k\| < \infty$, *rank* $(\Psi(1)) = n$ *if* $r = 0$, *rank* $(\Psi(1)) \geq n - r + 1$ *and rank* $(\Psi_{11}(1)) = n - r$ *if* $r \geq 1$, *where* $\Psi_{11}(1)$ *is the upper left* $(n-r) \times (n-r)$ *block of* $\Psi(1) = \sum_{k=0}^{\infty} \Psi_k$*; the remaining blocks being* $\Psi_{12}(1), \Psi_{21}(1)$*, and* $\Psi_{22}(1)$*. The* ε_t *are i.i.d. with* $E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_t' = I_n$, $E\|\varepsilon_t\|^q < \infty$ *for some* $q > \max(4, 2/(2d-1))$.

The representation in Assumption 3 is the assumption of cointegration (when $r \geq 1$). When the vector time series Z_t satisfies Assumption 3, it is fractionally integrated of order d according to Definition 1 and, when $r \geq 1$, also cointegrated according to Definition 2 with cointegration rank equal to $rank(R_r) = r$. That is, in the terminology of Definition 2, for a full-rank $n \times r$ matrix β with cointegration vectors $\beta_k, k = 1, \ldots, r$, as columns, $\beta' Z_t \in I(d-b)$ which implies by (10) that $\beta_k \in \mathcal{C}(R_r)$, where $\mathcal{C}(\cdot)$ denotes the column space of its matrix argument. The subspace spanned by all the linearly independent cointegration vectors, $\text{span}(\{\beta_k, k = 1, \ldots, r\}) = \mathcal{C}(R_r)$, is the cointegration space. Since R_r and R_{n-r} are orthogonal it follows that the (right) null space of R'_{n-r}, denoted $\mathcal{N}(R'_{n-r})$, also spans the cointegration space, i.e. $\mathcal{C}(R_r) = \mathcal{N}(R'_{n-r})$. It follows from Assumption 3 that $\dim(\mathcal{C}(R_r)) = r$, which is the cointegration rank. Note that this can also be seen through the well known equality $\dim(\mathcal{N}(R'_{n-r})) = n - rank(R'_{n-r}) = r$ from the solution of the homogeneous linear system $R'_{n-r}x = 0$ (in a conforming vector x). The formulation in Assumption 3 is quite general and includes several popular generating mechanisms such as the common unobserved components model and the error correction model of Johansen (2008, pp. 652-653). For instance, Z_t could have the common trends or common unobserved components representation $Z_t = Az_{1t} + Bz_{2t}$, where z_{1t} is an $(n-r)$-vector of common stochastic $I(d)$ trends and z_{2t} is an r-vector of common stochastic $I(d-b)$ trends. It then follows that the null spaces of A' and R'_{n-r} coincide, i.e. $\mathcal{N}(A') = \mathcal{N}(R'_{n-r})$.

Note that because $rank(\Psi_{11}(1)) = n - r$, $R'_{n-r}Z_t$ are exactly those linear transformations of Z_t that are not cointegrating. On the other hand, because $\Psi(1)$ can have reduced rank, no restrictions are placed on the possibility of cointegration among the variables in $R'_r Z_t$, or on the possibility that the variables in $R'_r Z_t$ are (fractionally) integrated of different orders as long as the highest integration order among those variables is $d-b$, i.e. $R'_r Z_t \in I(d-b)$ according to Definition 1. Also note that it is assumed that $d - b < 1/2 < d$, such that Z_t is nonstationary but the cointegrating relations are stationary. This assumption is needed for the proofs as will be explained below. However, while this condition does represent a drawback of the variance ratio approach, Monte Carlo simulation evidence in section 5 shows that the variance ratio tests do have power to detect cointegration even when $d - b \geq 1/2$ (and $b > 0$).

Under Assumption 3 a functional central limit theorem for Z_t is obtained in the non-cointegrating directions given by $\mathcal{C}(R_{n-r})$. Specifically, defining the full rank $(n-r) \times (n-r)$ matrix Φ as $\Phi\Phi' = [\Psi_{11}(1), \Psi_{12}(1)][\Psi_{11}(1), \Psi_{12}(1)]'$, the weak convergence

$$T^{1/2-d} R'_{n-r} Z_{\lfloor sT \rfloor} \Longrightarrow \Phi W_d^{n-r}(s), \quad 0 < s \leq 1, \tag{11}$$

holds, e.g. Corollary 1 of Marinucci & Robinson (2000). Note that $E\varepsilon_t\varepsilon_t' = I_n$ without loss of generality since Ψ_0 is unrestricted. Also note that only weak convergence in the non-cointegrating directions $\mathcal{C}(R_{n-r})$ is needed. In (11), the superscript denotes the dimension of the vector fractional standard Brownian motion given by stacking independent copies of (2)–(3).

Generalizing the variance ratio statistic (9) from the univariate case, consider testing for cointegration rank with a sequence of tests for $r = 0, \ldots, n-1$ based on the statistic

$$R_T(d_1) = A_T B_T^{-1}, \tag{12}$$

where

$$A_T = \sum_{t=1}^{T} Z_t Z_t', \quad B_T = \sum_{t=1}^{T} \tilde{Z}_t \tilde{Z}_t', \tag{13}$$

and

$$\tilde{Z}_t = \Delta_+^{-d_1} Z_t, \quad d_1 > 0, \quad t = 1, 2, \ldots. \tag{14}$$

Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the ordered eigenvalues of $R_T(d_1)$ which are found as the solutions to the eigenproblem

$$|\lambda B_T - A_T| = 0, \tag{15}$$

thus ensuring that both the eigenvalues and corresponding eigenvectors are real (note that the eigenvectors of $R_T(d_1)$ are not necessarily real because $R_T(d_1)$ is not necessarily symmetric).

If η_j denotes the eigenvector associated with λ_j, the eigenvalues are

$$\lambda_j = \frac{\eta_j' A_T \eta_j}{\eta_j' B_T \eta_j}, \quad j = 1, \ldots, n, \tag{16}$$

and the behavior of the eigenvalues depends on the direction of the associated eigenvector. Intuitively, if $\eta_j \in \mathcal{C}(R_{n-r})$, i.e. η_j lies in the non-cointegrating directions, then $T^{-2d}\eta_j' A_T \eta_j = O_P(1)$ and $T^{-2d-2d_1}\eta_j' B_T \eta_j = O_P(1)$ resulting in $T^{2d_1}\lambda_j = O_P(1)$, where the $O_P(1)$ rates are exact. On the other hand, if $\eta_j \in \mathcal{C}(R_r) = \mathcal{N}(R_{n-r}')$ such that η_j lies in the cointegrating directions, then $T^{-2\max(1/2,d-b)}\eta_j' A_T \eta_j = O_P(1)$ and $T^{-2\max(1/2,d-b+d_1)}\eta_j' B_T \eta_j = O_P(1)$, where again the $O_P(1)$ rates are exact. Noting that $d_1 > 0$, it follows that $\lambda_j = O_P(1)$ if $d - b + d_1 < 1/2$, $T^{2(d-b)+2d_1-1}\lambda_j = O_P(1)$ if $d - b < 1/2$ and $d - b + d_1 > 1/2$, and $T^{2d_1}\lambda_j = O_P(1)$ if $d - b > 1/2$. In other words, if the eigenvector η_j is not a cointegration vector, the associated eigenvalue λ_j converges to zero at rate $O_P(T^{-2d_1})$, and if η_j is a cointegration vector, λ_j is $O_P(T^{1-\max(2d-2b+2d_1,1)})$ if $d - b < 1/2$ or $O_P(T^{-2d_1})$ if $d - b > 1/2$.

These arguments illustrate how the order of magnitude of λ_j depends on the direction of η_j, and thus how the difference in convergence rates depending on whether or not η_j is a cointegration vector can be used in testing for cointegration rank, and also illustrates the necessity of the condition $d - b < 1/2$ for the proofs. Thus, the condition appears to be rather specific to the variance ratio cointegration rank testing approach, and does not necessarily apply to other approaches. Indeed, for Gaussian estimation of cointegrating vectors it is whether $b < 1/2$ or not that is more important. For instance Robinson & Hualde (2003) and Hualde & Robinson (2007) show that when $b < 1/2$ their estimator is asymptotically normal and otherwise mixed normality is obtained.

The statistic (12) can be adjusted for a non-zero mean or deterministic time trend in the observed time series by preliminary regression. The following assumption is made to accommodate deterministic terms.

Assumption 4 *The observed time series* $\{Y_t\}_{t=1}^{T}$ *is generated by*

$$Y_t = \alpha' \delta_t + Z_t, \quad t = 1, 2, \ldots, \tag{17}$$

where Z_t *is defined in Assumption 3 and* δ_t *is the deterministic term. Here,* $\delta_t = 0$ *when there are no deterministic terms,* $\delta_t = 1$ *when there is a non-zero mean, and* $\delta_t = [1, t]'$ *when there is correction for a non-zero deterministic linear trend.*

Assumption 4 is an unobserved components representation. In the case with correction for non-zero deterministic terms, the calculation of A_T and B_T and consequently the eigenvalues of (15) are based on least squares residuals $\hat{Z}_t = Y_t - \hat{\alpha}'\delta_t = Z_t - (\hat{\alpha} - \alpha)'\delta_t$ with $\hat{\alpha} = (\sum_{t=1}^{T} \delta_t \delta_t')^{-1} \sum_{t=1}^{T} \delta_t Y_t'$. The variance ratio corrected for deterministic terms is defined as in (12) but with the residuals $\hat{Z}_t$ replacing Z_t in (13) and (14). As expected, the asymptotic distribution changes to reflect the detrending, see Theorem 1 below.

Now introduce the nonparametric variance ratio trace statistic,

$$\Lambda_{n,r}(d_1) = T^{2d_1} \sum_{j=1}^{n-r} \lambda_j, \quad r = 0, \ldots, n-1, \tag{18}$$

where $\lambda_j, j = 1, \ldots, n$, are the eigenvalues of (15) possibly corrected for deterministic terms. Note that (18) defines a family of tests indexed by the fractional integration/summation parameter, d_1. In the integer-based model with $d = 1$, the choice $d_1 = 1$ leads to the test of Breitung (2002) which is therefore a special case of (18). The asymptotic distribution of $\Lambda_{n,r}(d_1)$ is given in the following theorem.

Theorem 1 *Suppose Assumptions 3 and 4 hold for the observed time series* Y_t*. Let* $\Lambda_{n,r}(d_1)$ *be given by (18) with residuals* $\hat{Z}_t$ *replacing* Z_t *in (13) and (14), and let* $j = 0$ *when* $\delta_t = 0$*,* $j = 1$ *when* $\delta_t = 1$*, and* $j = 2$ *when* $\delta_t = [1, t]'$*. For* $d_1 > 0$ *and* $r = 0, \ldots, n-1$,

$$\Lambda_{n,r}(d_1) \xrightarrow{D} U_{n-r}(d, d_1) = \mathrm{tr}\left\{ \int_0^1 B_{j,d}^{n-r}(s) B_{j,d}^{n-r}(s)' \mathsf{d}s \left(\int_0^1 \tilde{B}_{j,d,d_1}^{n-r}(s) \tilde{B}_{j,d,d_1}^{n-r}(s)' \mathsf{d}s \right)^{-1} \right\} \tag{19}$$

as $T \to \infty$*, where*

$$B_{j,d}^{n-r}(s) = W_d^{n-r}(s), \quad j = 0,$$

and the demeaned ($j = 1$) *or detrended* ($j = 2$) *standard fractional Brownian motions are defined as*

$$B_{j,d}^{n-r}(s) = W_d^{n-r}(s) - \int_0^1 W_d^{n-r}(r) D_j(r)' \mathsf{d}r \left(\int_0^1 D_j(r) D_j(r)' \mathsf{d}r \right)^{-1} D_j(s), \quad j = 1, 2,$$

Table 1: Simulated critical values $CV_{\xi,n-r}(1,d_1)$ with $\delta_t = 0$

d_1	ξ	$n-r$: 1	2	3	4	5	6	7	8
0.10	0.10	1.54	3.07	4.78	6.60	8.51	10.49	12.54	14.64
	0.05	1.62	3.16	4.86	6.68	8.59	10.57	12.62	14.73
	0.01	1.77	3.33	5.03	6.85	8.75	10.74	12.80	14.90
0.25	0.10	2.78	5.94	9.90	14.50	19.63	25.25	31.35	37.85
	0.05	3.15	6.33	10.34	14.97	20.13	25.78	31.92	38.44
	0.01	3.89	7.13	11.23	15.90	21.12	26.82	33.03	39.61
0.50	0.10	6.77	18.41	35.29	57.27	84.44	116.66	154.17	196.94
	0.05	8.49	20.91	38.59	61.27	89.04	121.99	160.22	203.56
	0.01	12.61	26.32	45.56	69.32	98.44	132.39	171.87	216.87
0.75	0.10	15.36	57.43	129.39	234.47	376.66	560.01	788.39	1064.26
	0.05	21.05	69.48	148.44	260.02	409.23	600.84	836.97	1121.25
	0.01	36.82	99.06	192.12	315.21	482.83	683.28	934.70	1240.08
1.00	0.10	33.69	175.52	473.65	967.52	1702.29	2727.18	4092.73	5849.63
	0.05	49.39	226.69	570.13	1113.51	1906.13	3008.15	4444.85	6282.55
	0.01	99.45	363.74	808.99	1450.00	2398.07	3607.89	5207.49	7229.55

Note: The simulated critical values are based on 100,000 replications and sample size 1,000. The test rejects when the test statistic is larger than the critical values in this table.

with $D_1(s) = 1$, $D_2(s) = [1, s]'$, *and*

$$\tilde{B}^{n-r}_{j,d,d_1}(s) = W^{n-r}_{d+d_1}(s), \quad j = 0,$$

$$\tilde{B}^{n-r}_{j,d,d_1}(s) = W^{n-r}_{d+d_1}(s) - \int_0^1 W^{n-r}_d(r) D_j(r)' \mathsf{d}r \left(\int_0^1 D_j(r) D_j(r)' \mathsf{d}r\right)^{-1} \int_0^s \frac{(s-r)^{d_1-1}}{\Gamma(d_1)} D_j(r) \mathsf{d}r, j = 1, 2.$$

The asymptotic distribution $U_{n-r}(d, d_1)$ of $\Lambda_{n,r}(d_1)$ presented in Theorem 1 depends only on the type of deterministic terms included, i.e. j, the integration order of the observed time series, d, the parameter d_1 indexing the family of tests, and the dimensionality of the problem or number of common stochastic trends, $n-r$. There is no dependence on the parameter b since the asymptotic theory in Theorem 1 is dominated by the asymptotic behavior in the non-cointegrating directions and b is only relevant in the cointegrating directions. Hence, the asymptotic distribution can easily be simulated to obtain quantiles on a case by case basis. For the special case $d = 1$, which is particularly important in empirical economic analyses, quantiles of $U_{n-r}(1, d_1)$ for several values of the parameter d_1 are given in Tables 1-3 for $n - r = 1, 2, \ldots, 8$. The quantiles are simulated based on 100,000 replications and sample size 1,000. Alternatively, quantiles of $U_{n-r}(d, d_1)$ using a consistent estimate of the integration order of the observed data, i.e. quantiles of $U_{n-r}(\hat{d}, d_1)$ for some $\hat{d} \xrightarrow{P} d$, can be simulated to obtain an asymptotic distribution of the variance ratio trace statistic which does not depend on any prespecified fractional integration order. The latter method will be used in the empirical application below.[1]

In the fractional integration case, the nonparametric variance ratio setup offers even greater advantages than in the traditional $I(1) - I(0)$ cointegration model. The reason is that the asymptotic

[1] Note that, for this approach to yield reasonable size properties in practice, the critical values should be a smooth function of d. If the critical values change dramatically for small changes in d, then the actual size may be far from the nominal size whenever d is not estimated with sufficient precision. Unreported simulations confirm that the critical values vary smoothly in d.

Table 2: Simulated critical values $CV_{\xi,n-r}(1, d_1)$ with $\delta_t = 1$

		$n-r$							
d_1	ξ	1	2	3	4	5	6	7	8
0.10	0.10	1.76	3.50	5.32	7.23	9.21	11.26	13.36	15.52
	0.05	1.82	3.57	5.40	7.31	9.29	11.34	13.45	15.60
	0.01	1.94	3.71	5.54	7.46	9.45	11.50	13.61	15.76
0.25	0.10	3.86	7.82	12.33	17.38	22.96	28.96	35.42	42.26
	0.05	4.19	8.22	12.77	17.87	23.47	29.50	35.99	42.85
	0.01	4.89	9.03	13.66	18.81	24.47	30.55	37.10	44.02
0.50	0.10	12.26	27.73	47.92	73.28	104.12	139.88	180.94	227.16
	0.05	14.44	30.74	51.63	77.58	109.17	145.43	187.26	234.16
	0.01	19.50	37.10	59.43	86.50	119.40	156.78	200.05	248.03
0.75	0.10	31.81	88.63	175.62	298.06	461.04	665.50	917.25	1217.34
	0.05	40.90	104.13	198.10	326.10	498.24	709.54	969.15	1278.58
	0.01	64.06	139.60	248.84	389.16	577.09	802.86	1080.02	1403.63
1.00	0.10	69.11	264.93	625.27	1193.84	2032.40	3165.29	4664.48	6563.52
	0.05	97.91	331.65	740.21	1359.66	2264.41	3474.60	5051.41	7038.12
	0.01	179.32	499.54	1018.16	1740.85	2790.42	4149.45	5889.12	8045.26

Note: The simulated critical values are based on 100,000 replications and sample size 1,000. The test rejects when the test statistic is larger than the critical values in this table.

distribution in (19) has the obvious advantage of not depending on the unobservable parameter b, the degree of cointegration. Thus, if a consistent estimate of d can be found by preliminary estimation and subsequently used to find relevant critical values, or if d can be attributed some value from economic theory (e.g. $d = 1$), then this approach enables hypothesis testing on the cointegration rank with no need to specify the unobserved degree of fractional cointegration, b. Indeed, from Assumption 3 and by inspection of the proof of Theorem 1, it is clear that the parameter b does not have to be the same for each cointegrating vector. That is, it can be the case that $\beta_k' Z_t \in I(d - b_k)$ for $k = 1, \ldots, r$ without changing the result in Theorem 1.

The multivariate variance ratio statistic (18) is related other variance ratio type tests of the cointegration rank hypothesis (20) by, e.g., Phillips & Ouliaris (1988), Stock & Watson (1988), and Shintani (2001), in the same way that the univariate variance ratio statistic is related to, e.g., the KPSS statistic of Kwiatkowski, Phillips, Schmidt & Shin (1992). However, these other test statistics are based on (eigenvalues of) the ratio of the sample second moment matrix of $\hat{Z}_t$ and that of $\Delta\hat{Z}_t$, and rely on either parametric or nonparametric estimation of the (one-sided or two-sided) long-run covariance matrix to eliminate serial correlation nuisance parameters from the limiting distribution. On the other hand, the variance ratio statistic (18) relies on the sample second moment matrix of the fractional partial sum of $\hat{Z}_t$ rather than on $\Delta\hat{Z}_t$, which implies that there is no need to estimate serial correlation parameters or long-run covariance matrices. Thus, $\Lambda_{n,r}(d_1)$ is asymptotically invariant to any short-run dynamics in the data generating process for Y_t. As a result, any hypothesis test based on the variance ratio statistic will share this useful property.

Furthermore, there is no need to estimate the cointegration vectors or the subspace spanned by the cointegration vectors in order to implement the variance ratio approach as in, e.g., Harris (1997) and Snell (1999), who require estimation of the $I(1)$ and $I(0)$ subsystems employing a principal components approach. There is also no need to specify a bandwidth as in the nonparametric

Table 3: Simulated critical values $CV_{\xi,n-r}(1,d_1)$ with $\delta_t = [1,t]'$

		$n-r$							
d_1	ξ	1	2	3	4	5	6	7	8
0.10	0.10	1.93	3.81	5.75	7.74	9.80	11.90	14.06	16.26
	0.05	1.98	3.88	5.82	7.82	9.88	11.99	14.15	16.35
	0.01	2.08	4.01	5.97	7.97	10.04	12.15	14.31	16.51
0.25	0.10	4.84	9.57	14.73	20.29	26.31	32.74	39.58	46.77
	0.05	5.18	9.98	15.20	20.80	26.86	33.30	40.17	47.39
	0.01	5.83	10.79	16.13	21.77	27.92	34.41	41.36	48.64
0.50	0.10	19.70	40.68	66.24	96.36	131.88	172.50	218.55	269.51
	0.05	22.46	44.18	70.64	101.23	137.67	178.89	225.42	277.17
	0.01	28.29	51.54	79.32	111.23	149.17	191.36	239.51	292.54
0.75	0.10	70.31	159.00	281.48	439.52	643.94	892.65	1192.40	1542.26
	0.05	85.11	180.11	310.84	475.95	688.77	945.67	1252.74	1612.75
	0.01	119.35	228.93	372.68	551.65	781.30	1052.08	1378.49	1757.86
1.00	0.10	228.18	586.52	1157.34	1970.49	3113.59	4589.24	6484.16	8829.19
	0.05	291.93	697.41	1325.41	2202.48	3418.60	4973.74	6957.87	9404.43
	0.01	457.46	971.59	1706.81	2709.98	4074.66	5797.09	7956.28	10613.21

Note: The simulated critical values are based on 100,000 replications and sample size 1,000. The test rejects when the test statistic is larger than the critical values in this table.

approach of Bierens (1997) or the semiparametric frequency domain approaches of Robinson & Yajima (2002), Chen & Hurvich (2003), and Nielsen & Shimotsu (2007). All these parameters are tuning parameters which change the test statistics and possibly also the inference drawn from the data since the asymptotic distribution theory does not reflect the particular choice of tuning parameters. The present nonparametric approach does not depend on any such tuning parameters although the parameter d_1 does need to be specified and may be considered a tuning parameter in finite samples. However, an important difference between d_1 and tuning parameters is that d_1 appears in the asymptotic distribution for $\Lambda_{n,r}(d_1)$. This opens up the possibility of finding a value of d_1 for which the corresponding test has desirable properties, and Nielsen (2008) argues in favor of $d_1 = 0.1$ in an asymptotic local power analysis of the univariate variance ratio test. Simulation evidence in section 5 below also supports the choice of $d_1 = 0.1$ compared to $d_1 = 1$. The value $d_1 = 0.1$ is used in the simulations and in the empirical application below, although comparisons with results for $d_1 = 1$ are also made.

Next, consider using the statistic $\Lambda_{n,r}(d_1)$ to construct a test for cointegration rank, where large values of $\Lambda_{n,r_0}(d_1)$ are associated with rejection of

$$H_0 : r = r_0 \text{ vs } H_1 : r > r_0. \tag{20}$$

The properties of this test are given in the next theorem.

Theorem 2 *Under the assumptions of Theorem 1, the test that rejects the null hypothesis H_0 in (20) when $\Lambda_{n,r_0}(d_1) > CV_{\xi,n-r_0}(d,d_1)$, where $CV_{\xi,n-r_0}(d,d_1)$ is found from*

$$P\left(U_{n-r_0}(d,d_1) > CV_{\xi,n-r_0}(d,d_1)\right) = \xi,$$

has asymptotic size ξ and is consistent against $H_1 : r > r_0$.

Armed with the results of Theorems 1 and 2, the following straightforward testing strategies can be suggested for determining the cointegration rank based on the variance ratio rank test. First of all, if r_0 in (20) is specified based on some a priori knowledge of the model, simply calculate the test statistic (18) and compare it with the relevant critical value $CV_{\xi,n-r_0}(d, d_1)$. One problem with this approach, however, is that the test is one-sided since it is consistent only against $r > r_0$, and so will not be able to detect deviations from the null in the direction $r < r_0$. Secondly, it is most often the case that there is no a priori information about r, or even if there is a priori information it is often desirable to estimate r based on the observed data. Thus, consider the following sequence of tests. First, compare $\Lambda_{n,0}(d_1)$ with the relevant critical value $CV_{\xi,n}(d, d_1)$. If $\Lambda_{n,0}(d_1) < CV_{\xi,n}(d, d_1)$, i.e. the hypothesis that $r = 0$ is not rejected, then $\hat{r} = 0$. If $\Lambda_{n,0}(d_1) > CV_{\xi,n}(d, d_1)$, the hypothesis that $r = 0$ is rejected and $\Lambda_{n,1}(d_1)$ is then compared with the critical value $CV_{\xi,n-1}(d, d_1)$. If $\Lambda_{n,1}(d_1) < CV_{\xi,n-1}(d, d_1)$, then $\hat{r} = 1$ and if $\Lambda_{n,1}(d_1) > CV_{\xi,n-1}(d, d_1)$ the hypothesis that $r = 1$ is rejected and $\Lambda_{n,2}(d_1)$ is compared with the critical value $CV_{\xi,n-2}(d, d_1)$, etc. This testing strategy defines an estimate of the cointegration rank which, based on Theorems 1 and 2, is equal to the true cointegration rank with probability $1 - \xi$ as $T \to \infty$ and converges in probability to the true cointegration rank if also $\xi \to 0$ as $T \to \infty$.

4 Estimation of the Cointegration Space

As usual in the cointegration literature, without identifying restrictions it is only possible to estimate a basis for the space spanned by the cointegration vectors, i.e. $\text{span}(\{\beta_k, k = 1, \ldots, r\}) = \mathcal{C}(R_r) = \mathcal{N}(R'_{n-r})$, see e.g. Johansen (1988, 1991), Bierens (1997), and more recently Chen & Hurvich (2003, 2006). It is thus not constructive in the present framework to define particular eigenvectors of the problem (15) as estimators of the cointegration vectors, since their consistency would not be well defined. Moreover, since the nonparametric variance ratio framework is a testing strategy for cointegration rank, it does not deliver a straightforward distribution theory for the estimated cointegration space (the space spanned by a subset of the eigenvectors of (15)). It will be shown, however, that the estimated cointegration space is consistent in the sense that the angle between the estimated and the true cointegration spaces converges to zero as $T \to \infty$.

The cointegration rank or equivalently the dimension of the column space of R_r was determined in section 3 above and is, in this section, assumed known. For the purpose of testing cointegration rank, the $n - r$ smallest eigenvalues of (15) were used to find the dimension of $\mathcal{C}(R_r) = \mathcal{N}(R'_{n-r})$. However, for estimation of the cointegration space $\mathcal{C}(R_r)$, the eigenvectors η_j of (15) corresponding to the r largest eigenvalues are used. Denote the $n \times r$ matrix with these eigenvectors as columns by $\eta(r)$. The estimator of the cointegration space is given by $\eta(r)$, the columns of which are linearly independent and thus form a basis for $\mathcal{C}(\eta(r))$.

The next theorem shows that the angle between the estimated basis for the cointegration space, $\eta(r)$, and the basis for the true cointegration space, R_r, vanishes as $T \to \infty$, and also gives the rate at which the angle disappears. Denote the angle between $\eta(r)$ and R_r by θ. It is well known from linear algebra, e.g. Harville (1997, pp. 60-61), that the cosine of the angle is

$$\cos\theta = \frac{\text{tr}(R'_r \eta(r))}{\|R_r\| \, \|\eta(r)\|}. \tag{21}$$

If the angle θ between the estimated and true bases for the cointegration space converges to zero, then $\cos\theta$ should converge to unity as $T \to \infty$. On the other hand, if the angle between $\eta(r)$ and R_r is θ, then the angle between $\eta(r)$ and R_{n-r} is $\theta - \pi/2$ because the columns of R are all orthogonal and in particular R_r and R_{n-r} are orthogonal. Thus, as an alternative to (21), the angle θ can equivalently be described in terms of the sine function and the matrix R_{n-r} as

$$\sin\theta = \cos\left(\theta - \frac{\pi}{2}\right) = \frac{\operatorname{tr}(R'_{n-r}\eta(r))}{\|R_{n-r}\| \, \|\eta(r)\|}. \tag{22}$$

A related interpretation of $\sin\theta$, see Chen & Hurvich (2003), is that it is the square-root of the sum of the squared lengths of the residuals from the orthogonal projections of $\{\eta_{n-r+1}, \ldots, \eta_n\}$ onto $\mathcal{C}(R_r)$, i.e. onto the space spanned by the true cointegrating vectors. The formulation (22) leads to a simple proof of the following theorem.

Theorem 3 *Let the $n \times r$ matrix $\eta(r)$ denote the matrix of eigenvectors of (15) corresponding to the r largest eigenvalues and let $R = [R_{n-r}, R_r]$ be defined as in (10). Under the assumptions of Theorem 1, the angle θ between $\eta(r)$ and R_r satisfies, as $T \to \infty$,*

$$\sin\theta = \frac{\operatorname{tr}(R'_{n-r}\eta(r))}{\|R_{n-r}\| \, \|\eta(r)\|} = O_P\left(T^{-d-d_1+\max(d-b+d_1,1/2)}\right).$$

The rate of convergence of $\sin\theta$ for $d-b+d_1 > 1/2$, i.e. $O_P\left(T^{-b}\right)$, is the same as that obtained by Chen & Hurvich (2003) in their semiparametric frequency domain framework. However, the proof of the $\sin\theta$ result in Theorem 3 seems more direct than that in Chen & Hurvich (2003), who rely on results for invariant subspaces of perturbed matrices. Finally, note that in the integer integration case with $d = b = d_1 = 1$, the rate of convergence of $\sin\theta$ in Theorem 3 is the same as that obtained by Bierens (1997) for his nonparametric estimator in the $I(1) - I(0)$ cointegration model.

5 Finite Sample Performance

In this section simulation evidence is provided to evaluate the finite sample performance of the proposed nonparametric variance ratio test compared to other procedures. The bivariate vector $Y_t = (y_{1t}, y_{2t})'$ is simulated according to the simple triangular model

$$y_{1t} = \Delta_+^{-d} u_{1t}, \quad t = 1, \ldots, T, \tag{23}$$

$$y_{2t} = y_{1t} + u_{2t}, \quad t = 1, \ldots, T, \tag{24}$$

where the properties of $u_t = (u_{1t}, u_{2t})'$ determine the cointegration properties (and in particular the cointegration rank) of the model. Three different generating mechanisms are considered for u_t. First, Model A is the fractional system,

$$\text{Model A} : u_{1t} = \varepsilon_{1t}, u_{2t} = \Delta_+^{-(d-b)}\varepsilon_{2t}, \quad t = 1, \ldots, T. \tag{25}$$

Table 4: Simulated size and size-corrected power for Model A

Panel A: Uncorrelated errors ($\rho = 0$)																
	$T = 100$								$T = 250$							
b	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4
0.0	0.04	0.04	0.05	0.06	0.09	0.05	0.06	0.08	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.06
0.2	0.13	0.08	0.37	0.15	0.06	0.20	0.11	0.06	0.15	0.09	0.85	0.42	0.13	0.36	0.21	0.10
0.4	0.38	0.18	0.93	0.51	0.12	0.76	0.42	0.13	0.51	0.20	1.00	0.97	0.45	0.97	0.82	0.40
0.6	0.83	0.39	1.00	0.87	0.24	1.00	0.90	0.33	0.96	0.49	1.00	1.00	0.82	1.00	1.00	0.90
0.8	1.00	0.71	1.00	0.98	0.40	1.00	1.00	0.68	1.00	0.85	1.00	1.00	0.97	1.00	1.00	1.00
1.0	1.00	0.95	1.00	1.00	0.59	1.00	1.00	0.94	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Correlated errors ($\rho = 0.5$)																
	$T = 100$								$T = 250$							
b	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4
0.0	0.04	0.04	0.05	0.06	0.09	0.05	0.06	0.08	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.06
0.2	0.14	0.09	0.49	0.19	0.07	0.26	0.14	0.07	0.15	0.09	0.95	0.55	0.16	0.48	0.28	0.13
0.4	0.41	0.19	0.98	0.65	0.15	0.87	0.54	0.17	0.55	0.21	1.00	0.99	0.56	0.99	0.91	0.51
0.6	0.84	0.40	1.00	0.94	0.30	1.00	0.96	0.40	0.97	0.50	1.00	1.00	0.89	1.00	1.00	0.95
0.8	1.00	0.72	1.00	0.99	0.47	1.00	1.00	0.75	1.00	0.86	1.00	1.00	0.98	1.00	1.00	1.00
1.0	1.00	0.96	1.00	1.00	0.64	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: The simulated size and size-corrected rejection frequencies for Model A in (23)–(25) are reported for the uncorrelated error case with $\rho = 0$ (Panel A) and the correlated error case with $\rho = 0.5$ (Panel B). For each test, entries in the rows labeled $b = 0.0$ are simulated rejection frequencies under the null, i.e. the size of the test, and all other entries denote size-corrected power. $\Lambda_{n,r}(d_1)$ is the variance ratio cointegration rank test in (18), BH_p is the Breitung & Hassler (2002) cointegration rank test with lag-augmentation p, and JT_p is the Johansen (1988) trace test with lag-augmentation p. The simulations are based on 100,000 replications and use critical values from Table 1, the $\chi^2(4)$ distribution, and Johansen (1995, Table 15.1), respectively.

In Model B short-run dynamics is added to the system,

$$\text{Model B} \quad : \quad u_{1t} = e_{1t}, u_{2t} = \Delta_+^{-(d-b)} e_{2t}, \quad t = 1, \ldots, T, \tag{26}$$

$$e_{jt} = 0.5 e_{j,t-1} + \varepsilon_{jt}, j = 1, 2, \quad t = 1, \ldots, T, \tag{27}$$

and Model C is based on the traditional autoregressive alternatives,

$$\text{Model C} : u_{1t} = \varepsilon_{1t}, u_{2t} = a u_{2,t-1} + \varepsilon_{2t}, \quad t = 1, \ldots, T. \tag{28}$$

In all three models, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is i.i.d. normal with mean zero, unit variances, and contemporaneous correlation coefficient ρ.

Throughout, the focus is on the case where $d = 1$, i.e. the observed bivariate vector Y_t is $I(1)$. Thus, Models A and B are fractionally cointegrated (rank is one) when $b > 0$, and when $b = 0$ there is no cointegration (rank is zero). In Model C there is cointegration of the traditional $I(1) - I(0)$ kind with rank one when $a < 1$, and when $a = 1$ there is a unit root in u_{2t} in (24) and no cointegration. The sample sizes considered are $T = 100$ and $T = 250$, significance level $\xi = 5\%$ was employed throughout, and $100,000$ replications were used. All calculations were made in Ox, see Doornik (2006).

Table 4 presents simulated rejection frequencies for Model A with $b = 0$ (the null of no cointegration), i.e. the simulated size of the tests, and with $b = 0.2, 0.4, \ldots, 1.0$ (fractional cointegration). In the latter case, the simulated rejection frequencies correspond to finite sample power of the tests,

Table 5: Simulated size and size-corrected power for Model B

Panel A: Uncorrelated errors ($\rho = 0$)																	
	$T = 100$								$T = 250$								
b	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4
0.0	0.02	0.03	1.00	0.07	0.10	0.28	0.07	0.09	0.03	0.04	1.00	0.06	0.07	0.28	0.06	0.07
0.2	0.10	0.07	0.00	0.07	0.05	0.02	0.06	0.05	0.11	0.08	0.00	0.13	0.09	0.02	0.10	0.08
0.4	0.23	0.12	0.00	0.14	0.08	0.02	0.15	0.08	0.33	0.15	0.00	0.46	0.27	0.04	0.45	0.26
0.6	0.53	0.24	0.00	0.33	0.14	0.06	0.43	0.18	0.77	0.33	0.00	0.87	0.57	0.40	0.95	0.72
0.8	0.90	0.47	0.00	0.59	0.23	0.32	0.86	0.40	1.00	0.65	0.00	0.99	0.84	1.00	1.00	0.99
1.0	1.00	0.77	0.03	0.83	0.35	0.92	1.00	0.72	1.00	0.94	0.03	1.00	0.96	1.00	1.00	1.00
Panel B: Correlated errors ($\rho = 0.5$)																	
	$T = 100$								$T = 250$								
b	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4
0.0	0.02	0.03	1.00	0.06	0.10	0.28	0.07	0.09	0.03	0.04	1.00	0.06	0.07	0.27	0.06	0.06
0.2	0.10	0.07	0.01	0.07	0.06	0.04	0.07	0.06	0.12	0.08	0.00	0.16	0.11	0.06	0.13	0.09
0.4	0.24	0.13	0.01	0.18	0.09	0.14	0.19	0.10	0.35	0.16	0.00	0.59	0.34	0.39	0.57	0.34
0.6	0.55	0.24	0.03	0.40	0.16	0.47	0.53	0.22	0.79	0.34	0.03	0.94	0.65	0.96	0.98	0.80
0.8	0.90	0.47	0.12	0.66	0.26	0.91	0.91	0.46	1.00	0.66	0.25	1.00	0.87	1.00	1.00	1.00
1.0	1.00	0.77	0.36	0.85	0.38	1.00	1.00	0.75	1.00	0.94	0.75	1.00	0.97	1.00	1.00	1.00

Note: The simulated size and size-corrected rejection frequencies for Model B in (23)–(24) and (26)–(27) are reported for the uncorrelated error case with $\rho = 0$ (Panel A) and the correlated error case with $\rho = 0.5$ (Panel B). For each test, entries in the rows labeled $b = 0.0$ are simulated rejection frequencies under the null, i.e. the size of the test, and all other entries denote size-corrected power. $\Lambda_{n,r}(d_1)$ is the variance ratio cointegration rank test in (18), BH_p is the Breitung & Hassler (2002) cointegration rank test with lag-augmentation p, and JT_p is the Johansen (1988) trace test with lag-augmentation p. The simulations are based on 100,000 replications and use critical values from Table 1, the $\chi^2(4)$ distribution, and Johansen (1995, Table 15.1), respectively.

which are size-corrected.[2] Panel A reports the results for the uncorrelated error case with $\rho = 0$ and Panel B for the correlated error case with $\rho = 0.5$. The reported tests are the variance ratio cointegration rank test $\Lambda_{2,0}(d_1)$, the Breitung & Hassler (2002) fractional trace test (denoted BH), and the Johansen (1988) trace test (denoted JT). All the tests take the null to be non-cointegration. The variance ratio test is implemented with $d_1 = 0.1$ and $d_1 = 1$ with critical values from Table 1. The BH and JT tests are implemented with lag-augmentations 0, 1, and 4, denoted by subscripts, with critical values from the $\chi^2((n-r)^2)$ distribution and Johansen (1995, Table 15.1), respectively.

The results of Table 4 show that all tests have very good size properties in Model A, except BH_4 and JT_4, which are both slightly oversized when $T = 100$. In the smaller sample, the size-corrected power of the variance ratio test with $d_1 = 0.1$ is close to that of the parametric BH and JT tests with lag-augmentation 1, and the variance ratio test has much higher power than the BH and JT tests with lag-augmentation 4. Even for the larger sample size, the $\Lambda_{2,0}(0.1)$ test has size-corrected finite sample rejection frequencies that are comparable to those of the parametric tests with lag-augmentation between 1 and 4. These conclusions hold even though the samples are quite small for nonparametric inference. Not surprisingly, the BH_0 (which is a correctly specified parametric test in Model A) and the JT_0 tests have higher power than the variance ratio test in this model.

In Table 5, laid out as the previous table, simulated rejection frequencies for Model B are reported. Here, short-run dynamics distort the performance of the BH and JT tests in the smaller

[2] Uncorrected power simulations are available from the author upon request.

Table 6: Simulated size and size-corrected power for Model C

Panel A: Uncorrelated errors ($\rho = 0$)																	
	$T = 100$								$T = 250$								
a	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4
1.0	0.04	0.05	0.05	0.06	0.09	0.05	0.06	0.08	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.06
0.9	0.27	0.15	0.11	0.13	0.09	0.20	0.18	0.13	0.68	0.37	0.25	0.34	0.31	0.84	0.79	0.63
0.8	0.57	0.32	0.28	0.30	0.16	0.65	0.55	0.30	0.98	0.66	0.74	0.81	0.65	1.00	1.00	0.98
0.7	0.83	0.50	0.55	0.53	0.24	0.96	0.86	0.48	1.00	0.80	0.98	0.98	0.84	1.00	1.00	1.00
0.6	0.96	0.63	0.81	0.73	0.32	1.00	0.98	0.64	1.00	0.88	1.00	1.00	0.93	1.00	1.00	1.00
0.5	0.99	0.73	0.95	0.86	0.38	1.00	1.00	0.74	1.00	0.93	1.00	1.00	0.96	1.00	1.00	1.00
Panel B: Correlated errors ($\rho = 0.5$)																	
	$T = 100$								$T = 250$								
a	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4	$\Lambda_{2,0}(0.1)$	$\Lambda_{2,0}(1)$	BH_0	BH_1	BH_4	JT_0	JT_1	JT_4
1.0	0.04	0.05	0.05	0.06	0.09	0.05	0.06	0.08	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.06
0.9	0.26	0.14	0.15	0.18	0.12	0.32	0.28	0.18	0.67	0.37	0.41	0.51	0.45	0.96	0.94	0.81
0.8	0.56	0.32	0.44	0.44	0.22	0.86	0.74	0.42	0.98	0.65	0.92	0.94	0.79	1.00	1.00	1.00
0.7	0.82	0.49	0.76	0.70	0.32	1.00	0.96	0.62	1.00	0.81	1.00	1.00	0.92	1.00	1.00	1.00
0.6	0.95	0.63	0.95	0.87	0.40	1.00	1.00	0.75	1.00	0.89	1.00	1.00	0.97	1.00	1.00	1.00
0.5	0.99	0.73	0.99	0.95	0.46	1.00	1.00	0.83	1.00	0.94	1.00	1.00	0.98	1.00	1.00	1.00

Note: The simulated size and size-corrected rejection frequencies for Model C in (23)–(24) and (28) are reported for the uncorrelated error case with $\rho = 0$ (Panel A) and the correlated error case with $\rho = 0.5$ (Panel B). For each test, entries in the rows labeled $a = 1.0$ are simulated rejection frequencies under the null, i.e. the size of the test, and all other entries denote size-corrected power. $\Lambda_{n,r}(d_1)$ is the variance ratio cointegration rank test in (18), BH_p is the Breitung & Hassler (2002) cointegration rank test with lag-augmentation p, and JT_p is the Johansen (1988) trace test with lag-augmentation p. The simulations are based on 100,000 replications and use critical values from Table 1, the $\chi^2(4)$ distribution, and Johansen (1995, Table 15.1), respectively.

sample. As expected, the parametric tests with no lag-augmentation are overwhelmingly oversized and are thus not useful for Model B. The augmented BH test with 4 lags is moderately size distorted with finite sample rejection frequencies of 10 and 7 percent under the null ($b = 0$) for $T = 100$ and $T = 250$, respectively. On the other hand, the variance ratio test is just slightly undersized. This model hence illustrates the usefulness of the nonparametric nature of the variance ratio test when compared to the parametric BH test. In the smaller sample, the size-corrected power of the variance ratio test with $d_1 = 0.1$ is superior to both the BH and JT tests for all three lag-augmentations. For the larger sample size, the $\Lambda_{2,0}(0.1)$ test has size-corrected finite sample rejection frequencies that are in between those of the parametric tests with lag-augmentations 1 and 4.

For comparison, Table 6 presents simulation results for the autoregressive alternative given in Model C, which is simulated for $a = 0.5, 0.6, \ldots, 1.0$ where $a = 1.0$ is the null of no cointegration. As expected, the JT test is superior in this case, especially the correctly specified parametric JT_0 test. However, perhaps more surprisingly, the nonparametric variance ratio test with $d_1 = 0.1$ has much better power properties than the BH test against this type of alternative. Indeed, the variance ratio test with $d_1 = 0.1$ actually has finite sample power comparable to (and in some cases better than) the JT_1 and JT_4 tests. Hence, the power loss due to the nonparametric nature of the variance ratio test is about as small as the power loss from overparameterization of the JT test (1-4 lags vs no lags), even in Model C for which the JT test is designed.

As a general observation, the variance ratio test with $d_1 = 0.1$ has much better finite sample power properties than the test with $d_1 = 1$. Since the model (23)–(24) was simulated with $d = 1$, the test with $d_1 = 1$ coincides with the cointegration rank test of Breitung (2002). It is therefore clear from the simulation results reported in this section that non-trivial power gains can be obtained by allowing fractional values of d_1 less than unity. This is the case even in Model C which is a non-fractional traditional $I(1) - I(0)$ cointegration model.

It is also worth observing that, based on Tables 4 and 5, the variance ratio tests have power when $b = 0.2$ or $b = 0.4$, thus detecting cointegration even when $d - b \geq 1/2$. This suggests that, although it is necessary for the proofs, the condition $d - b < 1/2$ from Assumption 3 may not be necessary in practice, and hence that this condition may not be a serious drawback of the variance ratio approach in practice.

It is clear from the simulations that the nonparametric variance ratio test is useful with good size and power in the models considered here, even though the sample sizes $T = 100$ and $T = 250$ would typically be considered quite small for nonparametric inference. In particular, the variance ratio test seems to be robust to different specifications of the model, including cointegrated alternatives of both fractional and autoregressive types. This is in contrast to the parametric Johansen (1988) and Breitung & Hassler (2002) cointegration rank tests, whose properties are very sensitive to the choice of lag-augmentation.

6 Empirical Application

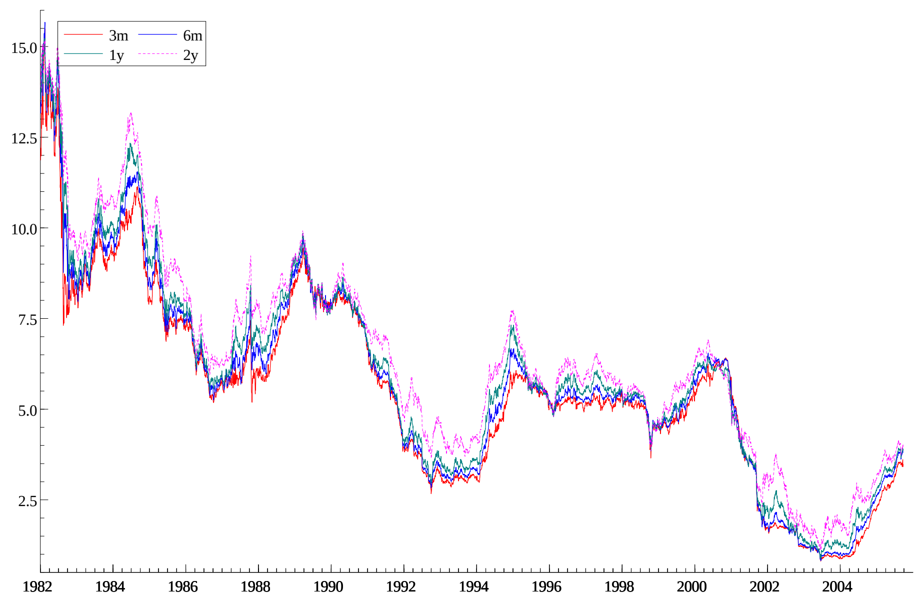
The proposed nonparametric methodology is applied to the expectations hypothesis for the term structure of interest rates and its implications in a cointegrated system which have attracted much attention in the last two decades. The restrictions laid down by the expectations hypothesis in terms of cointegration rank are tested within the fractional cointegration framework. For a general overview of term structure theory, see e.g. the survey by Pagan, Hall & Martin (1996).

Denote the interest rate at time t with maturity h by $i_t(h)$, where h is measured in years, and consider a system of n interest rates of different maturities, $h_1, h_2, \ldots, h_n$. Assuming that interest rates are (fractionally) integrated processes, the expectations hypothesis has a number of cointegration implications, see e.g. Hall, Anderson & Granger (1992). First of all, the expectations hypothesis for the term structure of interest rates implies that among these n variables there should be $n-1$ cointegrating relations, or equivalently that there should be only one common stochastic trend moving the interest rates. Second, for any linear combination of such n interest rates with different maturities, say $a_1 i_t(h_1) + a_2 i_t(h_2) + \ldots + a_n i_t(h_n)$, the coefficients should sum to zero when the expectations hypothesis holds, i.e. in equilibrium. In terms of the cointegration implications, this means the coefficients in each of the cointegration vectors, say β_k for $k = 1, \ldots, n-1$, should add up to zero. In principle, the zero-sum restriction can be tested by

$$\beta = H\phi, \quad H = \begin{bmatrix} I_{n-1} \\ -\iota \end{bmatrix}, \tag{29}$$

where $\iota = [1, \ldots, 1]$ such that H is a $n \times (n-1)$ matrix and ϕ is a $(n-1) \times (n-1)$ matrix of free parameters, see Hall et al. (1992) and Engsted & Tanggaard (1994). Geometrically, the cointegration vectors should be spanned by the columns of H, $\mathcal{C}(\beta) \subseteq \mathcal{C}(H)$. In (29), note that the restrictions

Figure 1: Time series plots of interest rates covering all business days from 1/4/82 to 9/23/05



in the first $n-1$ rows of H are identifying and the restrictions in the last row are overidentifying. Thus, the zero-sum restriction in (29) imposes only $n-1$ testable parameter restrictions on β. However, the present methodology is meant for testing the cointegration rank, so the focus here is on testing the rank restriction although the estimated cointegration space is also briefly compared informally to H.

These ideas were applied by Campbell & Shiller (1987) who tested the rank and zero-sum restrictions jointly by conducting cointegration tests on spreads between two interest rates. The analysis was generalized to a systems framework by Stock & Watson (1988) who applied their common trends test to testing the rank condition, and by Hall et al. (1992) and Engsted & Tanggaard (1994), among others, who applied the Johansen (1988, 1991) VAR methodology to test both the rank and zero-sum restrictions. The analysis most closely related to the present study is Chen & Hurvich (2003), who conduct a fractional cointegration analysis of eight different interest rates using their semiparametric cointegration rank test. They find that, although each series is insignificantly different from $I(1)$, fractional cointegration is important because the cointegration residuals are fractionally integrated. They also conclude that there are six cointegrating relations among their eight interest rates series, and argue that there could be one common trend driving the short maturity interest rates and another common trend driving the long maturity interest rates. It has also been argued in the literature that the expectations hypothesis may hold better in the shorter end of the yield curve, and thus the interest rates considered here are for relatively short maturities.

The interest rate data considered here is a system of four U.S. Treasury Bill interest rates with constant maturities of 3 months, 6 months, 1 year, and 2 years, respectively. The data

Table 7: Univariate analysis

Panel A: GSP estimates of d				
Bandwidth	$i_t\,(1/4)$	$i_t\,(1/2)$	$i_t\,(1)$	$i_t\,(2)$
$m = \lfloor T^{0.4} \rceil = 32$	0.96	1.02	1.02	1.01
$m = \lfloor T^{0.6} \rceil = 183$	1.08*	1.07	1.09*	1.10**
$m = \lfloor T^{0.8} \rceil = 1043$	1.01	1.01	1.03*	1.03*
Panel B: Univariate variance ratio test				
d_1	$i_t\,(1/4)$	$i_t\,(1/2)$	$i_t\,(1)$	$i_t\,(2)$
0.1	1.93	1.93	1.93	1.94
1.0	255.51	240.15	228.97	214.29

Note: Panel A reports GSP estimates based on the first difference of the interest rate data with unity added back and Panel B reports the univariate variance ratio test $\rho\,(d_1)$ for the detrended interest rate data. Standard errors for the GSP estimates can be found from the asymptotic distribution $\sqrt{m}(\hat{d}-d) \xrightarrow{D} N\,(0, 1/4)$, see Robinson (1995). The critical values for the variance ratio test of the hypothesis $d = 1$ can be found in Table 3 under $n - r = 1$. One and two asterisks denote rejection of $d = 1$ at 5% and 1% significance levels, respectively.

set is extracted from the Federal Reserve Board of Governors H.15 release and consists of daily observations for all business days from 4 January, 1982, until 23 September, 2005, for a total of $T = 5,932$ observations covering a span of over 23 years. Hence, the data set is quite long, both in terms of the number of observations and in terms of the span of time covered. A larger data set including also interest rates with maturities 3, 5, 7, and 10 years was also analyzed, and the results were qualitatively very similar to the results shown here. Specifically, the nonparametric variance ratio approach indicated a cointegration rank of $r = 6$ (i.e. two common stochastic I(1) trends), whereas the parametric approaches indicated a much lower rank. Since it has been argued in the literature that the expectations hypothesis may hold better in the shorter end of the yield curve only the results for the four interest rates with maturities up to 2 years are reported here.

In Figure 1 the four interest rates are plotted. The graphical evidence shows quite clearly that the interest rates tend to move together over time, but also that deviations can be quite persistent, e.g. towards the end of the sample period. The existence of a non-zero maturity risk premium and some trending behavior is evident from Figure 1, and consequently the following analysis allows for both non-zero mean and trend terms.

Table 7 presents univariate results for the interest rate series. Panel A reports Gaussian semi-parametric (GSP) estimates of d based on the first difference with unity added back, and Panel B reports the univariate variance ratio test $\rho\,(d_1)$ for the detrended interest rate data. One and two asterisks denote rejection of $d = 1$ at 5% and 1% significance levels, respectively. Note that the tests of $d = 1$ in Panel A are against the two-sided alternative, whereas the alternative in Panel B is given in Theorems 1–2. Also note that the GSP estimates are invariant to non-zero mean and trend terms, and that their standard errors can be found from the asymptotic distribution $\sqrt{m}(\hat{d}-d) \xrightarrow{D} N\,(0, 1/4)$, see Robinson (1995). The critical values for the variance ratio test of the hypothesis $d = 1$ can be found in Table 3 under $n-r = 1$. It is seen that, with the exception of some evidence from the middle bandwidth for the GSP estimates, the interest rates appear insignificantly different from $I\,(1)$. This supports previous evidence in the literature, e.g. Chen & Hurvich (2003).

In Panel A of Table 8 the eigenvalues of (15) are reported and in Panel B the corresponding

Table 8: Variance ratio cointegration rank tests

Panel A: Eigenvalues				
d_1	λ_1	λ_2	λ_3	λ_4
0.1	338.15	383.39	412.69	521.51
1.0	0.0057	0.0126	0.0218	0.0730
Panel B: $\Lambda_{n,r}(d_1)$				
d_1	$n-r=1$	$n-r=2$	$n-r=3$	$n-r=4$
0.1	1.92	4.10^{**}	6.45^{**}	9.41^{**}
1.0	201.68	645.46	1412.61^{*}	3979.62^{**}

Note: Panel A reports the eigenvalues (×1,000) of the problem (15) and Panel B reports the variance ratio cointegration rank test $\Lambda_{n,r}(d_1)$ for the detrended interest rate data. The critical values assuming $d = 1$ can be found in Table 3 and the critical values using the average GSP estimate with $m = 32$ for $\hat{d}$ can be found in Table 9. One and two asterisks denote rejection at 5% and 1% significance levels, respectively, according to the critical values in Table 9.

Table 9: Simulated critical values $CV_{\xi,n-r}(\hat{d}, d_1)$ with mean and trend

		$n-r$			
d_1	ξ	1	2	3	4
0.1	0.10	1.93	3.81	5.75	7.74
	0.05	1.98	3.87	5.83	7.82
	0.01	2.08	4.00	5.97	7.97
1.0	0.10	228.81	586.32	1159.92	1960.74
	0.05	293.45	691.22	1330.46	2198.69
	0.01	447.33	950.59	1691.45	2695.75

Note: The simulated critical values are based on 10,000 replications and sample size 1,000. The average GSP estimate with $m = 32$ for the detrended interest rate data was used for $\hat{d}$. The test rejects when the test statistic in Table 8 is larger than the critical values in this table.

variance ratio cointegration rank tests are reported for the detrended interest rate series. Both panels of Table 8 report the results for $d_1 = 0.1$ as well as $d_1 = 1$. Assuming $d = 1$ the critical values of the variance ratio tests can be found in Table 3. On the other hand, using the average GSP estimate with $m = 32$ for $\hat{d}$, the critical values have been simulated and are reported in Table 9. The asterisks in Table 8 denote rejection using these simulated critical values. Note that since $\hat{d}$ is close to one, the critical values in Table 9 are close to those for $d = 1$, and the conclusions reached do not change qualitatively if the critical values for $d = 1$ from Table 3 were used instead. Also note that the estimates for the larger bandwidth parameters in Table 7 are generally a little higher, and the larger bandwidths could have induced a bias in the estimation. Given the large number of observations in the data set it seems prudent to use the results for the bandwidth $m = 32$ which presumably should be less influenced by bias.

The results of the variance ratio tests reported in Tables 8 and 9 depend on the value of d_1. For the reasons given above, i.e. to maximize the power of the cointegration rank tests, most weight should be given to the $d_1 = 0.1$ results which clearly indicate the presence of only one common stochastic trend among the four interest rates, or in other words, the presence of three cointegrating relations. The results for $d_1 = 1$ are less clear but seem to indicate $r = 2$. These results are presumably a consequence of the improved power of the variance ratio test when $d_1 = 0.1$ compared

Table 10: Parametric cointegration rank tests

Panel A: Johansen (1988, 1991) trace test				
Lag length	$n-r=1$	$n-r=2$	$n-r=3$	$n-r=4$
0	5.67^{*}	29.59^{**}	139.06^{**}	527.11^{**}
21	10.11^{*}	31.05^{**}	69.32^{**}	149.42^{**}
63	15.27^{**}	37.80^{**}	70.13^{**}	105.69^{**}
126	15.19^{**}	33.22^{**}	51.46^{**}	79.56^{**}
Panel B: Breitung & Hassler (2002) fractional trace test				
Lag length	$n-r=1$	$n-r=2$	$n-r=3$	$n-r=4$
0	4.68^{*}	36.94^{**}	428.08^{**}	879.26^{**}
21	3.51	11.00	40.66^{**}	115.21^{**}
63	2.35	8.98	27.84^{**}	58.10^{**}
126	0.02	2.56	17.17	47.71^{**}

Note: In Panel A the Johansen (1988, 1991) trace test for cointegration rank is reported for the interest rate data with a unrestricted constant and trend terms. The critical values can be found in Johansen (1995, Table 15.5). In Panel B the Breitung & Hassler (2002) fractional trace test with estimated mean and trend is reported, and the critical values can be found from the $\chi^2((n-r)^2)$ distribution. One and two asterisks denote rejection at 5% and 1% significance levels, respectively.

to $d_1 = 1$, c.f. the simulation results above. Note that support of the expectations hypothesis is by rejection of the sequence of null hypotheses, $r_0 = 0, r_0 = 1, \ldots, r_0 = n-2$, and thus the more powerful test will offer more support of the expectations hypothesis, if it is indeed true.

For comparison, Table 10 presents the results of the Johansen (1991) cointegration rank trace tests with unrestricted constant and trend terms (Panel A) and the Breitung & Hassler (2002) fractional trace tests with estimated mean and trend in levels (Panel B). The lag-augmentations chosen for the two tests are 0, 21, 63, and 126 business days, corresponding to about 0, 1, 3, and 6 calendar months, respectively. The critical values for the Johansen tests can be found in Johansen (1995, Table 15.5) and the critical values for the Breitung & Hassler (2002) tests can be found from their asymptotic $\chi^2((n-r)^2)$ distribution. It appears from Panel A of Table 10 that the Johansen trace tests clearly reject all the hypotheses. The same rejection problem appears for the Breitung-Hassler tests in Panel B when there is no lag-augmentation. With a lag-augmentation, the conclusions of the Breitung-Hassler tests vary and depend on the length of the augmentation, with $r = 1$ as the result in the case with the most generous lag-augmentation. These seemingly conflicting results could be due to misspecification of the lag-augmentation of the parametric tests, and it is not attempted here to conduct a more thorough investigation. Instead, note once more that the nonparametric variance ratio test does not depend on specification of short-run dynamics.

Table 11 presents the eigenvectors of (15) for the detrended interest rate data. The eigenvectors are sorted in the same order as the eigenvalues in Panel A of Table 8, i.e. η_j corresponds to λ_j for $j = 1, \ldots, 4$, and the final row reports the sum of all the elements of the corresponding eigenvector. According to the expectations hypothesis, the elements of all the columns of any basis for the cointegration space should sum to zero, and it is noted from Table 11 that this may indeed be the case for all the eigenvectors. Since the eigenvectors are not normalized in the most intuitive way and should not be considered estimates of particular cointegration vectors, the interpretation of the eigenvectors is not so straightforward. Instead, if the eigenvectors corresponding to the three largest eigenvalues are gathered in the matrix $\eta(3) = [\eta_2, \eta_3, \eta_4]$, the cointegration space can be rotated

Table 11: Eigenvectors of (15)

	$d_1 = 0.1$				$d_1 = 1.0$			
	η_1	η_2	η_3	η_4	η_1	η_2	η_3	η_4
$i_t\,(1/4)$	4.25	−1.75	−14.91	68.93	−0.02	−0.10	−0.11	0.77
$i_t\,(1/2)$	−2.65	−17.14	−31.50	−129.87	0.06	0.23	−0.25	−1.42
$i_t\,(1)$	3.43	60.53	68.87	66.54	−0.09	−0.33	0.58	0.57
$i_t\,(2)$	0.01	−45.16	−23.76	−2.21	0.07	0.19	−0.24	0.13
sum	5.03	−3.52	−1.30	3.39	0.02	−0.01	−0.02	0.05

Note: The eigenvectors (×1,000) of the problem (15) are reported for the detrended interest rate data. The eigenvectors are sorted in the same order as the eigenvalues in Panel A of Table 8, i.e. η_j corresponds to λ_j for $j = 1, \ldots, 4$. The final row reports the sum of all the elements of the corresponding eigenvector.

Table 12: Estimated cointegration space for $r = 3$

	$d_1 = 0.1$			$d_1 = 1.0$		
	η_2	η_3	η_4	η_2	η_3	η_4
$i_t\,(1/4)$	1	0	0	1	0	0
$i_t\,(1/2)$	0	1	0	0	1	0
$i_t\,(1)$	0	0	1	0	0	1
$i_t\,(2)$	−1.09	−1.12	−1.10	−0.89	−0.97	−1.00
sum	−0.09	−0.12	−0.10	0.11	0.03	0.00

Note: The estimated and rotated cointegration space assuming $r = 3$, i.e. the renormalized $\eta\,(3)$, is reported for the detrended interest rate data. Only identifying restrictions are imposed on the cointegration space.

to facilitate the interpretation. Thus, Table 12 reports the elements of $\eta\,(3)\,([I_3, 0_{3\times1}]\,\eta\,(3))^{-1}$, which is normalized such that each vector of the estimated and rotated cointegration space can be interpreted as an estimated spread between two interest rates, c.f. (29). Note that no restrictions are imposed in Table 12 except identification.

The results in Table 12 show informally that the estimated cointegration space is very close to the hypothesized space in (29). Although the nonparametric variance ratio testing approach does not yield a straightforward distribution theory for the estimated cointegration space, and hence does not lead to a natural formal test of (29), it is consistent as shown in Theorem 3. In particular, for both choices of d_1, the columns of the estimated and rotated cointegration space are close to being just the differences between two interest rates of different maturities.

In general, this empirical application illustrates the usefulness of the nonparametric cointegration rank test. While the parametric tests of Johansen (1988, 1991) and Breitung & Hassler (2002) depend heavily on the choice of lag-augmentation, the nonparametric test does not depend on any such tuning parameters, but only on d_1 which it is argued should be set at $d_1 = 0.1$. Intuitively, there may be several reasons for the parametric testing procedures to fail to indicate the correct cointegration rank. In finite samples, the Johansen tests may fail to detect some cointegrating relations if the cointegrating strength, i.e. b, for those relations is low. In particular, if $d - b > 0$, the cointegration errors have long memory, and may not be detected as a cointegrating relation by the $I\,(1) - I\,(0)$-motivated Johansen tests. On the other hand, the Breitung-Hassler tests should be able to detect the presence of such fractional cointegrating relations. Most importantly, however, the parametric nature of the Johansen and Breitung-Hassler tests may cause the tests to lead to incorrect conclusions simply due to misspecified autocorrelation structure and lag-augmentation. Hence, this illustrates the usefulness of the nonparametric variance ratio methodology, and in particular highlights the advantages of its nonparametric nature and its ability to detect fractional cointegration among nonstationary fractionally integrated variables.

7 Concluding Remarks

In this paper a nonparametric variance ratio testing approach has been proposed for determining the cointegration rank in fractionally integrated systems. An integral part of the approach is to allow the integration orders to be unknown, i.e., if there is no strong a priori reason to pick particular integration orders, it is important to make only minimal assumptions on their values and possibly allow them to be estimated as part of the procedure. The proposed approach is nonparametric in the sense that it does not require the specification of a particular data generating process or model and is asymptotically invariant to the short-run dynamics of the time series. Furthermore, the approach does not require estimating the cointegration vector(s) and is thus easier to implement than regression-based approaches, especially when examining relationships between several variables with possibly multiple cointegrating vectors. Finally, a consistent estimator of (a basis for) the cointegration space can be obtained from the procedure.

The usefulness of the proposed approach was documented in a simulation study with samples as small as $T = 100$ or $T = 250$ which would often be considered quite small for nonparametric inference. The test appears to be robust to different specifications of the simulated model, including cointegrated alternatives of both fractional and autoregressive types. Finally, the proposed methodology was applied to an analysis of the term structure of interest rates, where, contrary to both fractional and integer-based parametric approaches, clear evidence in favor of the restrictions implied by the expectations hypothesis was found using the nonparametric approach.

Appendix: Proofs

This section begins with three lemmas that describe the behavior of processes and product moments in the stationary and nonstationary directions, complementing similar results in, e.g., Robinson & Hualde (2003, Appendix D) and Johansen & Nielsen (2009, Appendix C). Throughout C denotes a finite positive constant which may take different values in different places.

Lemma 4 *Let $\pi_k(u) = \Gamma(k+u)/(\Gamma(u)\Gamma(k+1))$ denote the k'th coefficient in the binomial expansion of Δ^{-u} and define*

$$\zeta_T(u, v) = \max_{1 \le l,m \le T} \sum_{t=\max(l,m)}^{T} |\pi_{t-l}(u)\pi_{t-m}(v)|. \tag{30}$$

Then

$$\zeta_T(u, v) \le C(\log T)^{\mathbb{I}_{\{u+v=1\}}} T^{\max(u+v-1,u-1,v-1,0)}, \tag{31}$$

where $\mathbb{I}_{\{A\}}$ denotes the indicator function of the event A and the constant does not depend on u, v, or T.

Proof. By symmetry it is enough to consider $v \leq u$. Note that $\pi_k(u)$ satisfies $|\pi_k(u)| \leq Ck^{u-1}, k \geq 1$, and $|\pi_0(u)| \leq C$. In the case where $v = 0$ it holds that $\pi_{t-m}(0) = C\mathbb{I}_{\{t=m\}}$ so that

$$\sum_{t=\max(l,m)}^{T} |\pi_{t-l}(u)\pi_{t-m}(0)| = C|\pi_{m-l}(u)| \leq CT^{\max(u-1,0)},$$

and similarly for $u = 0$. Thus, assume that $u \neq 0$ and $v \neq 0$.

For the term with $t = \max(l,m)$ and $l = m$, it follows that $|\pi_{t-l}(u)\pi_{t-m}(v)| = |\pi_0(u)\pi_0(v)| \leq C$. When $t = \max(l,m)$ and $l > m$, $|\pi_{t-l}(u)\pi_{t-m}(v)| = |\pi_0(u)\pi_{l-m}(v)| \leq CT^{\max(v-1,0)}$; $l < m$ is similar. So now consider only the terms with $t \geq \max(l,m)+1$.

Replacing $\pi_{t-l}(u)$ and $\pi_{t-m}(v)$ by their upper bounds for $t \geq \max(l,m)+1$ and $u \neq 0, v \neq 0$, yields the term $\sum_{t=\max(l,m)+1}^{T}(t-l)^{u-1}(t-m)^{v-1}$. If $u-1 \leq 0$ then also $v-1 \leq 0$ and hence $(t-l)^{u-1}(t-m)^{v-1} \leq (t-\max(l,m))^{(u+v-1)-1}$ such that

$$\sum_{t=\max(l,m)+1}^{T} (t-l)^{u-1}(t-m)^{v-1} \leq C(\log T)^{\mathbb{I}\{u+v=1\}}T^{\max(u+v-1,0)}.$$

If $u-1 > 0$ then $(t-l)^{u-1} \leq T^{u-1}$ and

$$\sum_{t=\max(l,m)+1}^{T} (t-l)^{u-1}(t-m)^{v-1} \leq T^{u-1} \sum_{t=\max(l,m)+1}^{T} (t-m)^{v-1} \leq CT^{u-1+\max(v,0)} \leq CT^{\max(u+v-1,u-1)}.$$

■

Lemma 5 *Let $X_{j,t} = \sum_{k=0}^{\infty} \tau_{j,k}\varepsilon_{j,t-k}, j = 1,2$, be n_j-dimensional stationary linear processes with $\sum_{k=0}^{\infty} \|\tau_{j,k}\| < \infty$ for $n_j \times n_j$ matrices $\tau_{j,k}$ and n_j-dimensional i.i.d. random variables $\varepsilon_{j,t}$ with zero mean and finite fourth moments. Define the product moment $Q_{ijT}(u,v) = T^{-1}\sum_{t=1}^{T} \Delta_+^{-u}X_{i,t}\Delta_+^{-v}X'_{j,t}$.*

(a) For $1/2 < v \leq u$ it holds that

$$T^{-u-v+1}Q_{ijT}(u,v) = O_P(1), \quad i,j = 1,2.$$

(b) For $v < 1/2 < u$ it holds that

$$T^{-u+1/2}Q_{ijT}(u,v) \overset{P}{\to} 0, \quad i,j = 1,2.$$

(c) For $v = 1/2 \leq u$ it holds that

$$T^{-u+1/2}(\log T)^{-\mathbb{I}_{\{u=1/2\}}}Q_{ijT}(u,v) = O_P(1), \quad i,j = 1,2.$$

Proof. The proof is given for the scalar case $n_1 = n_2 = 1$. The proof for $n_j \geq 2$ is the same, but the notation is more complicated. The second moment of $Q_{ijT}(u,v)$ is

$$\begin{aligned}
& E(T^{-1}\sum_{t=1}^{T}\sum_{l=1}^{t}\sum_{m=1}^{t} \pi_{t-l}(u)\pi_{t-m}(v)X_{i,l}X_{j,m})^2 \\
= \; & T^{-2}\sum_{t_1,t_2=1}^{T}\sum_{l_1,m_1=1}^{t_1}\sum_{l_2,m_2=1}^{t_2} \pi_{t_1-l_1}(u)\pi_{t_1-m_1}(v)\pi_{t_2-l_2}(u)\pi_{t_2-m_2}(v)E(X_{i,l_1}X_{i,l_2}X_{j,m_1}X_{j,m_2}) \\
\leq \; & \zeta_T(u,v)^2 \sum_{l_1,l_2,m_1,m_2=1}^{T} T^{-2}\left\|E(X_{i,l_1}X_{i,l_2}X_{j,m_1}X_{j,m_2})\right\|,
\end{aligned}$$

and in view of (31) in Lemma 4 the proof is completed by showing that the second factor is bounded by a constant that does not depend on T.

The last factor is

$$T^{-2}\sum_{l_1,l_2,m_1,m_2=1}^{T}\sum_{p_1=0}^{l_1-1}\sum_{p_2=0}^{l_2-1}\sum_{q_1=0}^{m_1-1}\sum_{q_2=0}^{m_2-1}\left\|\tau_{i,p_1}\tau_{i,p_2}\tau_{j,q_1}\tau_{j,q_2}E(\varepsilon_{i,l_1-p_1}\varepsilon_{i,l_2-p_2}\varepsilon_{j,m_1-q_1}\varepsilon_{j,m_2-q_2})\right\|.$$

Note that the number of terms in the eight summations is proportional to T^8, but most are zero because $E(\varepsilon_{j,t})=0$ for $j=1,2$. Specifically, there is a non-zero contribution only if the second subscripts on the four ε's are equal in pairs (or all equal, which is a special case), i.e. if $l_1 = l_2+p_1-p_2, m_1 = m_2+q_1-q_2$ or $l_1 = m_1+p_1-q_1, l_2 = m_2+p_2-q_2$ or $l_1 = m_2+p_1-q_2, l_2 = m_1+p_2-q_1$. Each of these three cases impose two constraints, thus eliminating two summations and leaving in each case six summations of the generic form

$$\begin{aligned} T^{-2}\sum_{l,m=1}^{T}\sum_{p_1=0}^{T}\sum_{p_2=0}^{l-1}\sum_{q_1=0}^{T}\sum_{q_2=0}^{m-1}\left\|\tau_{i,p_1}\tau_{i,p_2}\tau_{j,q_1}\tau_{j,q_2}\right\| &\leq C\sum_{p_1=0}^{T}||\tau_{i,p_1}||\sum_{p_2=0}^{T}||\tau_{i,p_2}||\sum_{q_1=0}^{T}||\tau_{j,q_1}||\sum_{q_2=0}^{T}||\tau_{j,q_2}|| \\ &\leq C\left(\sum_{k=0}^{\infty}||\tau_{i,k}||\right)^2\left(\sum_{k=0}^{\infty}||\tau_{j,k}||\right)^2. \end{aligned}$$

This is finite because $\sum_{k=0}^{\infty}\|\tau_{j,k}\|<\infty$ for $j=1,2$, showing that $T^{-2}\sum_{l_1,l_2,m_1,m_2=1}^{T}\|E(X_{i,l_1}X_{i,l_2}X_{j,m_1}X_{j,m_2})\|$ is bounded by a constant which only depends on $\tau_{j,k}$ and $\tau_{i,k}$ and not on T. ■

For the next lemma, define the number $\psi=\max(d-b+d_1,1/2)$, the process $\tilde{Y}_t=\Delta_+^{-d_1}\hat{Z}_t$, c.f. (14), and the functionals

$$\begin{aligned} Z_T(s) &= T^{1/2-d}R_{n-r}'Z_{\lfloor sT\rfloor}, \\ \hat{Z}_T(s) &= T^{1/2-d}R_{n-r}'\hat{Z}_{\lfloor sT\rfloor}, \\ \tilde{Z}_T(s) &= T^{1/2-d-d_1}R_{n-r}'\tilde{Z}_{\lfloor sT\rfloor} = T^{1/2-d-d_1}\Delta_+^{-d_1}R_{n-r}'Z_{\lfloor sT\rfloor}, \\ \tilde{Y}_T(s) &= T^{1/2-d-d_1}R_{n-r}'\tilde{Y}_{\lfloor sT\rfloor} = T^{1/2-d-d_1}\Delta_+^{-d_1}R_{n-r}'\hat{Z}_{\lfloor sT\rfloor}. \end{aligned}$$

Also define the matrices $F_T=\sum_{t=1}^{T}R'\hat{Z}_t\hat{Z}_t'R$ and $G_T=\sum_{t=1}^{T}R'\tilde{Y}_t\tilde{Y}_t'R$, which are decomposed as

$$F_T=\begin{bmatrix}F_{11T} & F_{12T}\\ F_{12T}' & F_{22T}\end{bmatrix} \text{ and } G_T=\begin{bmatrix}G_{11T} & G_{12T}\\ G_{12T}' & G_{22T}\end{bmatrix},$$

where

$$\begin{aligned} F_{11T} &= \sum_{t=1}^{T}R_{n-r}'\hat{Z}_t\hat{Z}_t'R_{n-r}, \quad F_{12T}=\sum_{t=1}^{T}R_{n-r}'\hat{Z}_t\hat{Z}_t'R_r, \quad F_{22T}=\sum_{t=1}^{T}R_r'\hat{Z}_t\hat{Z}_t'R_r, \\ G_{11T} &= \sum_{t=1}^{T}R_{n-r}'\tilde{Y}_t\tilde{Y}_t'R_{n-r}, \quad G_{12T}=\sum_{t=1}^{T}R_{n-r}'\tilde{Y}_t\tilde{Y}_t'R_r, \quad G_{22T}=\sum_{t=1}^{T}R_r'\tilde{Y}_t\tilde{Y}_t'R_r. \end{aligned}$$

Finally, define the limits

$$\begin{aligned} F_{11} &= \Phi\int_0^1 B_{j,d}^{n-r}(s)\,B_{j,d}^{n-r}(s)'\,\mathsf{d}s\Phi', \quad F_{22}=Var(\Delta^{-(d-b)}u_{2t}), \\ G_{11} &= \Phi\int_0^1 \tilde{B}_{j,d,d_1}^{n-r}(s)\,\tilde{B}_{j,d,d_1}^{n-r}(s)'\,\mathsf{d}s\Phi'. \end{aligned}$$

Lemma 6 *Under the assumptions of Theorem 1, as* $T \to \infty$*,*

$(a) : T^{-2d}F_{11T} = T^{-2d}\sum_{t=1}^{T} R_{n-r}'\hat{Z}_t\hat{Z}_t'R_{n-r} \xrightarrow{D} F_{11}$,

$(b) : T^{-1}F_{22T} = T^{-1}\sum_{t=1}^{T} R_r'\hat{Z}_t\hat{Z}_t'R_r \xrightarrow{P} F_{22}$,

$(c) : T^{-d-1/2}F_{12T} = T^{-d-1/2}\sum_{t=1}^{T} R_{n-r}'\hat{Z}_t\hat{Z}_t'R_r \xrightarrow{P} 0$,

$(d) : T^{-2d-2d_1}G_{11T} = T^{-2d-2d_1}\sum_{t=1}^{T} R_{n-r}'\tilde{Y}_t\tilde{Y}_t'R_{n-r} \xrightarrow{D} G_{11}$,

$(e) : T^{-2\psi}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}G_{22T} = T^{-2\psi}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T} R_r'\tilde{Y}_t\tilde{Y}_t'R_r$

$$\begin{cases} \xrightarrow{P} Var(\Delta^{-(d-b+d_1)}u_{2t}) & \text{if } d-b+d_1 < 1/2, \\ = O_P(1) & \text{if } d-b+d_1 \geq 1/2, \end{cases}$$

$(f) : T^{-d-d_1-\psi}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}G_{12T} = T^{-d-d_1-\psi}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T} R_{n-r}'\tilde{Y}_t\tilde{Y}_t'R_r$

$$\begin{cases} \xrightarrow{P} 0 & \text{if } d-b+d_1 \leq 1/2, \\ = O_P(1) & \text{if } d-b+d_1 > 1/2, \end{cases}$$

where $u_{2,t}$ *is the* r*-vector with the last* r *elements of* u_t*.*

Proof. *Proof of (a) and (d)*: It is first shown that $\hat{Z}_T(s) = T^{1/2-d}R_{n-r}'\hat{Z}_{\lfloor sT\rfloor} \Longrightarrow \Phi B_{j,d}^{n-r}(s)$ and $\tilde{Y}_T(s) = T^{1/2-d-d_1}R_{n-r}'\tilde{Y}_{\lfloor sT\rfloor} \Longrightarrow \Phi\tilde{B}_{j,d,d_1}^{n-r}(s)$, from which results (a) and (d) will follow by the continuous mapping theorem.

Recall that $\hat{Z}_t = Z_t - (\hat{\alpha}-\alpha)'\delta_t$, where Z_t is defined in Assumption 3. From (11) the convergence

$$Z_T(s) = T^{1/2-d}R_{n-r}'Z_{\lfloor sT\rfloor} \Longrightarrow \Phi W_d^{n-r}(s)$$

holds, which proves the result when $\delta_t = 0$. For $\delta_t \neq 0$ define $N_1(T) = 1$ and $N_2(T) = diag(1, T^{-1})$ and write

$$T^{1/2-d}R_{n-r}'(\hat{\alpha}-\alpha)'\delta_{\lfloor sT\rfloor} = T^{1/2-d}R_{n-r}'(\hat{\alpha}-\alpha)'N_j(T)^{-1}N_j(T)\delta_{\lfloor sT\rfloor},$$

where

$$\begin{aligned} T^{1/2-d}R_{n-r}'(\hat{\alpha}-\alpha)'N_j(T)^{-1} &= \left(T^{-1}\sum_{s=1}^{T} T^{1/2-d}R_{n-r}'Z_s\delta_s'N_j(T)\right)\left(T^{-1}\sum_{s=1}^{T} N_j(T)\delta_s\delta_s'N_j(T)\right)^{-1} \\ &= \left(T^{-1}\sum_{s=1}^{T} T^{1/2-d}R_{n-r}'Z_sD_j(s/T)\right)\left(T^{-1}\sum_{s=1}^{T} D_j(s/T)D_j(s/T)'\right)^{-1} \\ &\xrightarrow{D} \Phi\left(\int_0^1 W_d^{n-r}(s)\,D_j(s)'\,\mathsf{d}s\right)\left(\int_0^1 D_j(s)\,D_j(s)'\,\mathsf{d}s\right)^{-1} \end{aligned} \tag{32}$$

as $T \to \infty$ by application of (11) and the continuous mapping theorem, and

$$N_j(T)\delta_{\lfloor sT\rfloor} = D_j(\lfloor sT\rfloor/T) \to D_j(s) \text{ as } T \to \infty. \tag{33}$$

It thus follows that

$$\hat{Z}_T(s) = T^{1/2-d}R_{n-r}'\hat{Z}_{\lfloor sT\rfloor} \Longrightarrow \Phi B_{j,d}^{n-r}(s), \quad j = 0,1,2. \tag{34}$$

Next, for

$$\begin{aligned} \tilde{Y}_T(s) &= T^{1/2-d-d_1}\Delta_+^{-d_1}R_{n-r}'\hat{Z}_{\lfloor sT\rfloor} \\ &= T^{1/2-d-d_1}\Delta_+^{-d_1}R_{n-r}'Z_{\lfloor sT\rfloor} - T^{1/2-d-d_1}\Delta_+^{-d_1}R_{n-r}'(\hat{\alpha}-\alpha)'\delta_{\lfloor sT\rfloor}, \end{aligned}$$

the convergence

$$\tilde{Z}_T(s) = T^{1/2-d-d_1}\Delta_+^{-d_1} R'_{n-r} Z_{\lfloor sT \rfloor} \Longrightarrow \Phi W^{n-r}_{d+d_1}(s)$$

holds from (11) because the operator Δ_+ satisfies $\Delta_+^{d_1}\Delta_+^{d_2} Z_t = \Delta_+^{d_1+d_2} Z_t$ for any d_1 and d_2. For the remaining term,

$$\begin{aligned}
T^{1/2-d-d_1}\Delta_+^{-d_1} R'_{n-r}(\hat{\alpha}-\alpha)'\delta_{\lfloor sT \rfloor} &= T^{1/2-d-d_1}\sum_{k=0}^{\lfloor sT \rfloor -1}\pi_k(d_1) R'_{n-r}(\hat{\alpha}-\alpha)'\delta_{\lfloor sT \rfloor - k} \\
&= T^{1/2-d-d_1}\sum_{k=1}^{\lfloor sT \rfloor}\pi_{\lfloor sT \rfloor - k}(d_1) R'_{n-r}(\hat{\alpha}-\alpha)'\delta_k \\
&= \left(T^{1/2-d} R'_{n-r}(\hat{\alpha}-\alpha)' N_j(T)^{-1}\right)\left(T^{-d_1}\sum_{k=1}^{\lfloor sT \rfloor}\pi_{\lfloor sT \rfloor - k}(d_1) N_j(T)\delta_k\right),
\end{aligned}$$

the first factor converges by (32) and the last factor is deterministic and satisfies the convergence

$$\begin{aligned}
T^{-d_1}\sum_{k=1}^{\lfloor sT \rfloor}\pi_{\lfloor sT \rfloor - k}(d_1) N_j(T)\delta_k &= T^{-d_1}\sum_{k=1}^{\lfloor sT \rfloor}\pi_{\lfloor sT \rfloor - k}(d_1) D_j(k/T) \\
&= T^{-d_1}\sum_{k=1}^{\lfloor sT \rfloor}\frac{(\lfloor sT \rfloor - k)^{d_1-1}}{\Gamma(d_1)} D_j(k/T) + o(1) \\
&= T^{-1}\sum_{k=1}^{\lfloor sT \rfloor}\frac{\left(\frac{\lfloor sT \rfloor}{T} - \frac{k}{T}\right)^{d_1-1}}{\Gamma(d_1)} D_j(k/T) + o(1) \\
&\to \int_0^s \frac{(s-r)^{d_1-1}}{\Gamma(d_1)} D_j(r)\,\mathsf{d}r \text{ as } T\to\infty. \qquad (35)
\end{aligned}$$

The second line follows from the substitution $\pi_k(d) = k^{d-1}(\Gamma(d)^{-1} + O(k^{-1}))$, which implies that the remainder term in the second line is of order $T^{-d_1}\sum_{k=1}^{\lfloor sT \rfloor}(\lfloor sT \rfloor - k)^{d_1-2} = O(T^{\max(-d_1,-1)}(\log T)) = o(1)$ (uniformly in s). Hence, it follows that

$$\tilde{Y}_T(s) \Longrightarrow \Phi \tilde{B}^{n-r}_{j,d,d_1}(s), \quad j = 0,1,2,$$

which proves results (a) and (d).

Proof of (b): First suppose $\delta_t = 0$. Since $d-b < 1/2$, define the zero-mean stationary time series (with no truncation) $\breve{Z}_t = \Delta^{-(d-b)}u_{2,t} = \sum_{k=0}^{\infty}\pi_k(d-b)u_{2,t-k}$. For $\breve{Z}_t$ the law of large numbers for stationary ergodic time series implies that

$$T^{-1}\sum_{t=1}^{T}\breve{Z}_t\breve{Z}'_t \xrightarrow{P} E(\breve{Z}_t\breve{Z}'_t) = Var(\Delta^{-(d-b)}u_{2,t}).$$

The processes $\breve{Z}_t$ and $R'_r Z_t$ are related as $\breve{Z}_t = R'_r Z_t + V_t$, where $V_t = \sum_{k=t}^{\infty}\pi_k(d-b)u_{2,t-k}$. From Assumption 3, $u_t = \sum_{l=0}^{\infty}\Psi_l\varepsilon_{t-l}$ with $\sum_{l=0}^{\infty} l^{1/2}||\Psi_l|| < \infty$ such that the Beveridge-Nelson decomposition implies that $u_t = (\sum_{l=0}^{\infty} \Psi_l)\varepsilon_t + \sum_{l=0}^{\infty} \tilde{\Psi}_l \Delta\varepsilon_{t-l}$ with $\sum_{l=0}^{\infty} ||\tilde{\Psi}_l||^2 < \infty$. Then, letting $\Psi_{2,l}$ and $\tilde{\Psi}_{2,l}$ denote the last r rows of Ψ_l and $\tilde{\Psi}_l$, respectively,

$$V_t = (\sum_{l=0}^{\infty} \Psi_{2,l}) \sum_{k=t}^{\infty} \pi_k(d-b)\varepsilon_{t-k} + \sum_{l=0}^{\infty} \tilde{\Psi}_{2,l} \sum_{k=t}^{\infty} \pi_k(d-b-1)\varepsilon_{t-k-l}$$

using that $\sum_{k=t}^{\infty} \pi_k(d-b)\Delta\varepsilon_{t-k} = \sum_{k=t}^{\infty} \pi_k(d-b-1)\varepsilon_{t-k}$. For V_t it holds that $E(V_t) = 0$ and

$$\begin{aligned}
E(T^{-1}\sum_{t=1}^{T} V_t V_t') &= (\sum_{l=0}^{\infty} \Psi_{2,l})(\sum_{l=0}^{\infty} \Psi_{2,l})' T^{-1} \sum_{t=1}^{T} \sum_{k=t}^{\infty} \pi_k(d-b)^2 \\
&\quad + T^{-1} \sum_{t=1}^{T} \sum_{l=0}^{\infty} \sum_{k_1=t}^{\infty} \sum_{k_2=t}^{\infty} \tilde{\Psi}_{2,l} \tilde{\Psi}_{2,l+k_1-k_2}' \pi_{k_1}(d-b-1)\pi_{k_2}(d-b-1) \\
&\quad + (\sum_{l_1=0}^{\infty} \Psi_{2,l_1}) \sum_{k_1=t}^{\infty} \sum_{k_2=t}^{\infty} \pi_{k_1}(d-b)\pi_{k_2}(d-b-1)\tilde{\Psi}_{2,k_1-k_2}' \\
&\quad + \sum_{k_1=t}^{\infty} \sum_{k_2=t}^{\infty} \pi_{k_1}(d-b)\pi_{k_2}(d-b-1)\tilde{\Psi}_{2,k_1-k_2}(\sum_{l_1=0}^{\infty} \Psi_{2,l_1})'.
\end{aligned}$$

The first two terms have norms bounded by

$$\begin{aligned}
CT^{-1}\sum_{t=1}^{T}\sum_{k=t}^{\infty} k^{2(d-b)-2} &\leq CT^{-1}\sum_{t=1}^{T} t^{2(d-b)-1} \to 0, \\
CT^{-1}\sum_{t=1}^{T}\sum_{k_1=t}^{\infty}\sum_{k_2=t}^{\infty} k_1^{d-b-2} k_2^{d-b-2} &\leq CT^{-1}\sum_{t=1}^{T} t^{2(d-b)-2} \to 0,
\end{aligned}$$

and the norms of the last two terms hence also converge to zero by the Cauchy-Schwarz inequality. It follows that the probability limit of $T^{-1}\sum_{t=1}^{T} R_r' Z_t Z_t' R_r$ is the same as that of $T^{-1}\sum_{t=1}^{T} \breve{Z}_t \breve{Z}_t'$.

When $\delta_t \neq 0$, $T^{-1}\sum_{t=1}^{T} N_j(T)\delta_t Z_t' R_r$ has mean zero and variance with norm bounded by

$$\begin{aligned}
||2ET^{-2}\sum_{t=1}^{T}\sum_{s=t}^{T} N_j(T)\delta_t Z_t' R_r R_r' Z_s \delta_s' N_j(T)|| &\leq ||CT^{-2}\sum_{t=1}^{T}\sum_{s=t}^{T} N_j(T)\delta_t \sum_{k=0}^{t-1} \tau_k' \tau_{s-t+k} \delta_s' N_j(T)|| \\
&\leq CT^{-2}\sum_{t=1}^{T}\sum_{s=t}^{T}\sum_{k=1}^{t-1} k^{d-b-1}(s-t+k)^{d-b-1} \\
&\leq CT^{-2}\sum_{t=1}^{T}\sum_{s=t}^{T}(s-t+1)^{d-b-1}\sum_{k=1}^{t-1} k^{d-b-1} \leq CT^{2(d-b)-1}
\end{aligned}$$

such that

$$T^{-1}\sum_{t=1}^{T} N_j(T)\delta_t Z_t' R_r = O_P(T^{d-b-1/2}), \tag{36}$$

$$N_j(T)^{-1}(\hat{\alpha}-\alpha)R_r = \left(T^{-1}\sum_{t=1}^{T} N_j(T)\delta_t\delta_t' N_j(T)\right)^{-1} T^{-1}\sum_{t=1}^{T} N_j(T)\delta_t Z_t' R_r = O_P(T^{d-b-1/2}), \tag{37}$$

$$N_j(T)\delta_t = D_j(t/T) = O_P(1). \tag{38}$$

From (36)-(38) it follows that

$$T^{-1}\sum_{t=1}^{T}R_r'Z_t\delta_t'(\hat{\alpha}-\alpha)R_r = O_P(T^{2(d-b)-1}), \tag{39}$$

$$T^{-1}\sum_{t=1}^{T}R_r'\left(\hat{\alpha}-\alpha\right)'\delta_t\delta_t'(\hat{\alpha}-\alpha)R_r = O_P(T^{2(d-b)-1}), \tag{40}$$

such that the result is the same as when $\delta_t = 0$.

Proof of (c): From Lemma 5(b) it follows that $T^{-d-1/2}\sum_{t=1}^{T}R_{n-r}'Z_tZ_t'R_r \xrightarrow{P} 0$ which proves the result for $\delta_t = 0$. When $\delta_t \neq 0$ the result is unchanged if

$$T^{-d-1/2}\sum_{t=1}^{T}R_{n-r}'Z_t\delta_t'\left(\hat{\alpha}-\alpha\right)R_r \xrightarrow{P} 0, \tag{41}$$

$$T^{-d-1/2}\sum_{t=1}^{T}R_{n-r}'\left(\hat{\alpha}-\alpha\right)'\delta_tZ_t'R_r \xrightarrow{P} 0, \tag{42}$$

$$T^{-d-1/2}\sum_{t=1}^{T}R_{n-r}'\left(\hat{\alpha}-\alpha\right)'\delta_t\delta_t'\left(\hat{\alpha}-\alpha\right)R_r \xrightarrow{P} 0. \tag{43}$$

By the Cauchy-Schwarz inequality the left-hand side of (41) is

$$O_P\left(\left(T^{-2d}\sum_{t=1}^{T}E||R_{n-r}'Z_t||^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}E||\delta_t'\left(\hat{\alpha}-\alpha\right)R_r||^2\right)^{1/2}\right),$$

where the first factor is $O_P(1)$ by Lemma 5(a) and the second is $o_P(1)$ by (40). Next, using (32), (42) is

$$T^{1/2-d}R_{n-r}'\left(\hat{\alpha}-\alpha\right)'N_j(T)^{-1}T^{-1}\sum_{t=1}^{T}N_j(T)\delta_tZ_t'R_r = O_P\left(T^{-1}\sum_{t=1}^{T}N_j(T)\delta_tZ_t'R_r\right),$$

which is $O_P(T^{d-b-1/2})$ as in (36). The term (43) follows in the same way from (32) and (40).

Proof of (e): When $d-b+d_1 < 1/2$ this result follows exactly as in (b). When $d-b+d_1 \geq 1/2$ and $\delta_t = 0$ write

$$T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}R_r'\tilde{Z}_t\tilde{Z}_t'R_r = O_P\left(T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}E||R_r'\tilde{Z}_t||^2\right) \tag{44}$$

and the result follows from Lemma 5(a) for $d-b+d_1 > 1/2$ and Lemma 5(c) for $d-b+d_1 = 1/2$. When $d-b+d_1 \geq 1/2$ and $\delta_t \neq 0$, the additional term $J_t = \sum_{k=1}^{t}\pi_{t-k}\left(d_1\right)R_r'\left(\hat{\alpha}-\alpha\right)'\delta_k$ arises and we have

$$
\begin{aligned}
& T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}R_r'\tilde{Y}_t\tilde{Y}_t'R_r \\
= \; & T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}R_r'\tilde{Z}_t\tilde{Z}_t'R_r + T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}J_tJ_t' \qquad (45) \\
& -T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}J_t\tilde{Z}_t'R_r - T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}R_r'\tilde{Z}_tJ_t'. \qquad (46)
\end{aligned}
$$

The result therefore follows if (46) and the second term of (45) are all $O_P(1)$. Note from (35) that $\sum_{k=1}^{t}\pi_{t-k}(d_1)N_j(T)\delta_k = O(T^{d_1})$, which together with (37) implies that

$$J_t = (R_r'(\hat{\alpha}-\alpha)'N_j(T)^{-1})(\sum_{k=1}^{t}\pi_{t-k}(d_1)N_j(T)\delta_k) = O_P(T^{d-b+d_1-1/2}).$$

Thus, the second term of (45) is

$$O_P(T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}E||J_t||^2) = O_P(1),$$

and by application of the Cauchy-Schwarz inequality and the bounds in Lemma 5(a,c), (46) is

$$O_P\left(T^{-2(d-b+d_1)}(\log T)^{-\mathbb{I}_{\{\psi=1/2\}}}\left(\sum_{t=1}^{T}E||J_t||^2\right)^{1/2}\left(\sum_{t=1}^{T}E||\tilde{Z}_t'R_r||^2\right)^{1/2}\right) = O_P(1),$$

which proves the result.

Proof of (f): If $d-b+d_1 < 1/2$ the proof follows as in part (c). When $d-b+d_1 \geq 1/2$, applying the Cauchy-Schwarz inequality shows that the left-hand side is

$$O_P\left(\left(T^{-2d-2d_1}\sum_{t=1}^{T}E||R_{n-r}'\tilde{Y}_t||^2\right)^{1/2}\left(T^{-2\psi}(\log T)^{-2\mathbb{I}_{\{\psi=1/2\}}}\sum_{t=1}^{T}E||R_r'\tilde{Y}_t||^2\right)^{1/2}\right),$$

where the first factor is $O_P(1)$ by part (d) and the second factor is $O_P((\log T)^{-\mathbb{I}_{\{\psi=1/2\}}})$ by part (e). ■

Note that the proofs of (e) and (f) in Lemma 6 in the case with $d-b+d_1 > 1/2$ would follow easily from the functional central limit theorem and continuous mapping theorem if the assumptions included the existence of $q > \max(2, 2/(2d-2b+2d_1-1))$ moments and full rank of $\Psi(1)$, in which case the limits of G_{12T} and G_{22T} would be functionals of fractional Brownian motion and the convergence would be weak. However, with only $q > \max(4, 2/(2d-1))$ moments the proof is completed instead by application of the bounds in Lemma 5, and G_{12T} and G_{22T} can only be described as $O_P(1)$ when $d-b+d_1 \geq 1/2$.

Proof of Theorem 1. Following the method of proof of Johansen (1988, 1991), the stochastic orders of magnitude of all the eigenvalues of (15), corresponding to eigenvectors in both the cointegrating and non-cointegrating directions, are derived first. The statistic $\Lambda_{n,r}(d_1)$ in Theorem 1 is a simple function of the ordered eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ of (15), which are also eigenvalues of the eigenproblem

$$|\lambda G_T - F_T| = 0, \tag{47}$$

where $F_T = \sum_{t=1}^{T} R'\hat{Z}_t\hat{Z}_t'R$ and $G_T = \sum_{t=1}^{T} R'\tilde{Y}_t\tilde{Y}_t'R$. The eigenproblem (47) has the advantage over (15) that the cointegrating and non-cointegrating directions are delivered directly through the pre- and post-multiplication by the orthonormal matrix R, which does not change the eigenvalues of the problem.

Letting

$$M_T = \begin{bmatrix} T^{-d-d_1-1/2+\psi} I_{n-r} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & T^{-1/2} I_r \end{bmatrix} \text{ and } \bar{F} = \begin{bmatrix} 0_{(n-r)\times(n-r)} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & F_{22} \end{bmatrix}$$

it follows by direct application of Lemma 6 that

$$\begin{aligned} M_T F_T M_T &= \begin{bmatrix} T^{-2d-2d_1-1+2\psi} F_{11T} & T^{-d-d_1-1+\psi} F_{12T} \\ T^{-d-d_1-1+\psi} F_{12T}' & T^{-1} F_{22T} \end{bmatrix} \xrightarrow{P} \bar{F}, \\ T^{1-2\psi} M_T G_T M_T &= \begin{bmatrix} T^{-2d-2d_1} G_{11T} & T^{-d-d_1-\psi} G_{12T} \\ T^{-d-d_1-\psi} G_{12T}' & T^{-2\psi} G_{22T} \end{bmatrix}, \end{aligned} \tag{48}$$

where $T^{-2d-2d_1} G_{11T} \xrightarrow{D} G_{11}$, which is symmetric and positive definite a.s., $T^{-d-d_1-\psi} G_{12T} = O_P(1), T^{-2\psi} G_{22T} = O_P(1)$, and F_{22} is deterministic, symmetric, and positive definite.

Next define $\rho = \lambda T^{2\psi-1}$. Then the ordered eigenvalues of (47) are the same as those of

$$\left|\rho T^{1-2\psi} M_T G_T M_T - M_T F_T M_T\right| = 0 \tag{49}$$

Denote by $\Upsilon(A, B)$ the function that associates with a pair of matrices (A, B) the ordered eigenvalues of the problem $|\rho A - B| = 0$. Since this function is continuous in its two matrix arguments and $(T^{1-2\psi} M_T G_T M_T, M_T F_T M_T)$ is tight, it follows from (48) that

$$||\Upsilon(T^{1-2\psi} M_T G_T M_T, M_T F_T M_T) - \Upsilon(T^{1-2\psi} M_T G_T M_T, \bar{F})|| \xrightarrow{P} 0.$$

The solutions of $\Upsilon(T^{1-2\psi} M_T G_T M_T, \bar{F})$ satisfy

$$\begin{aligned} 0 &= \left|\rho \begin{bmatrix} T^{-2d-2d_1} G_{11T} & T^{-d-d_1-\psi} G_{12T} \\ T^{-d-d_1-\psi} G_{12T}' & T^{-2\psi} G_{22T} \end{bmatrix} - \begin{bmatrix} 0_{(n-r)\times(n-r)} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & F_{22} \end{bmatrix}\right| \\ &= |\rho T^{-2d-2d_1} G_{11T}||\rho(T^{-2\psi} G_{22T} - (T^{-d-d_1-\psi} G_{12T})'(T^{-2d-2d_1} G_{11T})^{-1}(T^{-d-d_1-\psi} G_{12T})) - F_{22}|. \end{aligned} \tag{50}$$

This decomposition is valid because $T^{-2d-2d_1} G_{11T}$ is symmetric and positive definite a.s.

From the first factor of (50) it holds that $\rho_j \xrightarrow{P} 0$ for $j = 1, ..., n-r$ because $T^{-2d-2d_1} G_{11T} \xrightarrow{D} G_{11}$ which is symmetric and positive definite a.s. For the second factor, introduce $\nu = \rho^{-1}$ which solves

$$\begin{aligned} 0 &= |(T^{-2\psi} G_{22T} - (T^{-d-d_1-\psi} G_{12T})'(T^{-2d-2d_1} G_{11T})^{-1}(T^{-d-d_1-\psi} G_{12T})) - \nu F_{22}| \\ &= |F_{22}^{-1/2}(T^{-2\psi} G_{22T} - (T^{-d-d_1-\psi} G_{12T})'(T^{-2d-2d_1} G_{11T})^{-1}(T^{-d-d_1-\psi} G_{12T}))F_{22}^{-1/2} - \nu I_r|, \quad (51) \end{aligned}$$

where the second equality holds because F_{22} is symmetric and positive definite. By the continuous mapping theorem $(T^{-2d-2d_1} G_{11T})^{-1} \xrightarrow{D} G_{11}^{-1}$ since G_{11} is positive definite almost surely. This implies, in particular, that $(T^{-2d-2d_1} G_{11T})^{-1} = O_P(1)$. Since also $T^{-d-d_1-\psi} G_{12T} = O_P(1)$ and $T^{-2\psi} G_{22T} = O_P(1)$, the solutions of (51) satisfy $\nu_j = O_P(1)$ for $j = 1, ..., r$. It follows that $\rho_j^{-1} = O_P(1)$ for $j = n-r+1, ..., n$ and hence that the r largest roots of (15) satisfy $\lambda_j^{-1} = O_P(T^{2\psi-1})$ for $j = n-r+1, \dots, n$.

Now define

$$K_T = \begin{bmatrix} T^{-d} I_{n-r} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & T^{-1/2} I_r \end{bmatrix}$$

such that, by Lemma 6,

$$K_T F_T K_T = \begin{bmatrix} T^{-2d} F_{11T} & T^{-d-1/2} F_{12T} \\ T^{-d-1/2} F_{12T}' & T^{-1} F_{22T} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} F_{11} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & F_{22} \end{bmatrix}, \quad (52)$$

$$T^{-2d_1} K_T G_T K_T = \begin{bmatrix} T^{-2d-2d_1} G_{11T} & T^{-d-1/2-2d_1} G_{12T} \\ T^{-d-1/2-2d_1} G_{12T}' & T^{-1-2d_1} G_{22T} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} G_{11} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & 0_{r\times r} \end{bmatrix}, \quad (53)$$

noting that convergence in probability implies convergence in distribution on $\mathbb{R}^{m\times q}$. Again, the ordered eigenvalues of (47) are the same as those of

$$|\lambda K_T G_T K_T - K_T F_T K_T| = 0,$$

and letting $\mu = (T^{2d_1}\lambda)^{-1}$ they are also the same as those of

$$\left| T^{-2d_1} K_T G_T K_T - \mu K_T F_T K_T \right| = 0. \quad (54)$$

Since the eigenvalues are continuous functions of the argument matrices, (52) and (53) imply that the ordered eigenvalues of (54) converge in distribution to those of

$$\left| \begin{bmatrix} G_{11} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & 0_{r\times r} \end{bmatrix} - \mu \begin{bmatrix} F_{11} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & F_{22} \end{bmatrix} \right| = 0.$$

This equation has r zero roots (because F_{22} is deterministic and positive definite) and $n-r$ a.s. positive roots (because both G_{11} and F_{11} are symmetric and positive definite a.s.) given by the solutions of $|G_{11} - \mu F_{11}| = 0$.

It follows that $T^{2d_1}\lambda_1, \dots, T^{2d_1}\lambda_{n-r}$ converge in distribution to the solutions of $\left|G_{11} T^{2d_1}\lambda - F_{11}\right| = 0$ and hence

$$T^{2d_1} \sum_{j=1}^{n-r} \lambda_j \xrightarrow{D} \operatorname{tr}\left\{F_{11} G_{11}^{-1}\right\}.$$

■

Proof of Theorem 2. If $r > r_0$ the proof of Theorem 1 shows that $\lambda_{n-r+1}^{-1}, \ldots, \lambda_{n-r_0}^{-1}$ are $O_P(T^{2\psi-1})$. Hence, under the alternative that $r > r_0$, the test statistic satisfies $\Lambda_{n,r_0}(d_1) = T^{2d_1}\sum_{j=1}^{n-r_0}\lambda_j \geq T^{2d_1}\lambda_{n-r_0}$ and $(T^{2d_1}\lambda_{n-r_0})^{-1} = O_P(T^{2\psi-1-2d_1})$, noting that $\lambda_j \geq 0$ for all j. Since $2d_1 + 1 - 2\psi > 0$ when $d - b < 1/2$, it follows that $\Lambda_{n,r_0}(d_1)$ diverges to $+\infty$ as $T \to \infty$. This establishes consistency of the test against $r > r_0$. The asymptotic size of the test is ξ by Theorem 1 and the definition of $CV_{\xi,n-r_0}(d, d_1)$. ■

Proof of Theorem 3. The ordered eigenvectors η_j of (15) and the eigenvectors $\hat{v}_j$ of (49) are related by $\hat{v}_j = M_T^{-1}R^{-1}\eta_j = M_T^{-1}R'\eta_j$ for $j = 1, \ldots, n$. Moreover, denoting the last r eigenvectors of (49) by $\hat{v}(r)$, it follows that

$$\hat{v}(r) = M_T^{-1}R'\eta(r) = M_T^{-1}\begin{bmatrix} R'_{n-r}\eta(r) \\ R'_r\eta(r) \end{bmatrix}. \tag{55}$$

Denoting the last r eigenvectors of (50) by $v(r)$, it follows from the proof of Theorem 1 that the space $\mathcal{C}(\hat{v}(r))$ spanned by the columns of $\hat{v}(r)$ converges to $\mathcal{C}(v(r))$. The eigenvectors $v_j, j = 1, \ldots, n$, of (50) are normalized such that

$$v_j'\begin{bmatrix} 0_{(n-r)\times(n-r)} & 0_{(n-r)\times r} \\ 0_{r\times(n-r)} & F_{22} \end{bmatrix} v_j = \rho_j, \quad j = 1, \ldots, n, \tag{56}$$

$$v_j'\begin{bmatrix} T^{-2d-2d_1}G_{11T} & T^{-d-d_1-\psi}G_{12T} \\ T^{-d-d_1-\psi}G'_{12T} & T^{-2\psi}G_{22T} \end{bmatrix} v_j = 1, \quad j = 1, \ldots, n, \tag{57}$$

where, from the proof of Theorem 1, $\rho_1, \ldots, \rho_{n-r} = o_P(1)$ and $\rho_{n-r+1}^{-1}, \ldots, \rho_n^{-1} = O_P(1)$.

Since $T^{-2d-2d_1}G_{11T} \xrightarrow{D} G_{11}$, which is positive definite a.s., (57) implies that the first $n - r$ elements of v_j are (stochastically) bounded and have norm bounded away from zero for all j. The first $n - r$ rows of (55) are $T^{d+d_1+1/2-\psi}R'_{n-r}\eta(r)$ showing that $R'_{n-r}\eta(r) = O_P(T^{-d-d_1-1/2+\psi})$.

Since F_{22} is positive definite, (56) implies that the last r elements of $v_j, j = n - r + 1, \ldots, n$, i.e. the last r rows of $v(r)$, must have norm (stochastically) bounded away from zero. The last r rows of (55) are $T^{1/2}R'_r\eta(r)$ which shows that $\|\eta(r)\|^{-1} = O_P(T^{1/2})$. It follows that

$$\sin\theta = \frac{\mathrm{tr}(R'_{n-r}\eta(r))}{\|R_{n-r}\|\,\|\eta(r)\|} = O_P\left(T^{-d-d_1+\psi}\right).$$

■

Acknowledgements

I am grateful to the editor, Peter Robinson, two anonymous referees, an anonymous associate editor, Søren Johansen, Marius Ooms, and Tim Vogelsang for comments and suggestions that significantly improved the paper. I also thank seminar participants at Cornell University, Purdue University, Queen's University, and University of Wisconsin–Madison as well as participants at TSEFAR 2006, the Econometrics in Rio 2006 Conference, the European Meeting of the Econometric Society 2006, and the CREATES Symposium on Long Memory 2007 for comments and discussion, and the Social Sciences and Humanities Research Council of Canada (SSHRC grant no. 410-2009-0183), the Danish Social Sciences Research Council (FSE grant no. 275-05-0220), the Center for Analytic Economics (CAE) at Cornell University, and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) at the University of Aarhus for research support.

References

Bierens, H. J., 1997. Nonparametric cointegration analysis. Journal of Econometrics 77, 379–404.

Breitung, J., 2002. Nonparametric tests for unit roots and cointegration. Journal of Econometrics 108, 342–363.

Breitung, J., Hassler, U., 2002. Inference on the cointegration rank in fractionally integrated processes. Journal of Econometrics 110, 167–185.

Campbell, J. Y., Shiller, R. J., 1987. Cointegration and tests of present value models. Journal of Political Economy 95, 1062–1088.

Chen, W. W., Hurvich, C. M., 2003. Semiparametric estimation of multivariate fractional cointegration. Journal of the American Statistical Association 98, 629–642.

Chen, W. W., Hurvich, C. M., 2006. Semiparametric estimation of fractional cointegrating subspaces. Annals of Statistics 34, 2939–2979.

Doornik, J. A., 2006. Ox: An Object-Oriented Matrix Programming Language. Timberlake Consultants Press, London.

Engle, R., Granger, C. W. J., 1987. Cointegration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Engsted, T., Tanggaard, C., 1994. Cointegration and the US term structure. Journal of Banking and Finance 18, 167–181.

Granger, C. W. J., 1981. Some properties of time series data and their use in econometric model specification. Journal of Econometrics 16, 121–130.

Hall, A. D., Anderson, H. M., Granger, C. W. J., 1992. A cointegration analysis of Treasury bill yields. Review of Economics and Statistics 74, 116–126.

Harris, D., 1997. Principal components analysis of cointegrated time series. Econometric Theory 13, 529–557.

Harville, D. A., 1997. Matrix Algebra from a Statistician's Perspective. Springer, New York.

Henry, M., Zaffaroni, P., 2003. The long range dependence paradigm for macroeconomics and finance. In: Doukhan, P., Oppenheim, G., Taqqu, M. S. (Eds.), Theory and Applications of Long-Range Dependence. Birkhäuser, Boston, pp. 417–438.

Hualde, J., Robinson, P. M., 2007. Root-n-consistent estimation of weak fractional cointegration. Journal of Econometrics 140, 450–484.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

Johansen, S., 1995. 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., Nielsen, M. Ø., 2009. Likelihood inference for a nonstationary fractional autoregressive model. Forthcoming in Journal of Econometrics.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?. Journal of Econometrics 54, 159–178.

Marinucci, D., Robinson, P. M., 2000. Weak convergence of multivariate fractional processes. Stochastic Processes and their Applications 86, 103–120.

Nielsen, M. Ø., 2008. A powerful test of the autoregressive unit root hypothesis based on a tuning parameter free statistic. Forthcoming in Econometric Theory.

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

Pagan, A. R., Hall, A. D., Martin, V., 1996. Modeling the term structure. In: Maddala, G. S., Rao, C. R. (Eds), Handbook of Statistics 14: Statistical Methods in Finance. North-Holland, Amsterdam, pp. 91–118.

Phillips, P. C. B., Ouliaris, S., 1988. Testing for cointegration using principal components methods. Journal of Economic Dynamics and Control 12, 205–230.

Robinson, P. M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.

Robinson, P. M., Hualde, J., 2003. Cointegration in fractional systems with unkown integration orders. Econometrica 71, 1727–1766.

Robinson, P. M., Yajima, Y., 2002. Determination of cointegrating rank in fractional systems. Journal of Econometrics 106, 217–241.

Shintani, M., 2001. A simple cointegrating rank test without vector autoregression. Journal of Econometrics 105, 337–362.

Snell, A., 1999. Testing for r versus $r-1$ cointegrating vectors. Journal of Econometrics 88, 151–191.

Stock, J. H., Watson, M. W., 1988. Testing for common trends. Journal of the American Statistical Association 404, 1097–1107.

Taylor, A. M. R., 2005. Variance ratio tests of the seasonal unit root hypothesis. Journal of Econometrics 124, 33–54.

Vogelsang, T. J., 1998a. Testing for a shift in mean without having to estimate serial-correlation parameters. Journal of Business and Economic Statistics 16, 73–80.

Vogelsang, T. J., 1998b. Trend function hypothesis testing in the presence of serial correlation. Econometrica 66, 123–148.