Semiparametric Estimation of Multivariate Fractional Cointegration

Willa W. Chen [*]   Clifford M. Hurvich [†]

June 14, 2002

Abstract. We consider the semiparametric estimation of fractional cointegration in a multivariate process of cointegrating rank $r > 0$. We estimate the cointegrating relationships by the eigenvectors corresponding to the r smallest eigenvalues of an averaged periodogram matrix of tapered, differenced observations. The number of frequencies m used in the periodogram average is held fixed as the sample size grows. We first show that the averaged periodogram matrix converges in distribution to a singular matrix whose null eigenvectors span the space of cointegrating vectors. We then show that the angle between the estimated cointegrating vectors and the space of true cointegrating vectors is $O_p(n^{d_u - d})$ where d and d_u are the memory parameters of the observations and cointegrating errors, respectively. The proposed estimator is invariant to the labeling of the component series, and therefore does not require one of the variables to be specified as a dependent variable. We determine the rate of convergence of the r smallest eigenvalues of the periodogram matrix, and present a criterion which allows for consistent estimation of r. Finally, we apply our methodology to the analysis of fractional cointegration in interest rates.

Keywords: Fractional cointegration, long memory, tapering, periodogram.

1 Introduction

Fractional cointegration was originally defined in Engle and Granger (1987), and has been the subject of increasing recent attention in the econometric literature. Robinson (1994) proposed a narrow band least squares (NBLS) estimator of the cointegrating parameter, with bandwidth tending to ∞. Further results on this estimator in the nonstationary case were obtained by Robinson and Marinucci (2001). Chen and Hurvich (2002) considered a tapered NBLS estimator based on differenced data, and showed that this estimator, which is invariant to additive polynomial trends of a certain order, can converge faster under some circumstances than the non-tapered NBLS. Most of the existing theory for NBLS has been derived in the bivariate case. For series of dimension three or higher, NBLS suffers from the drawback that it requires the specification of one of the component series as a dependent variable. The estimator is not

[*] Department of Statistics, Texas A&M University, College Station, Texas 77843, USA

[†] New York University, 44 W. 4'th Street, New York NY 10012, USA

The authors thank Rohit Deo for helpful comments.

invariant to this choice, and not all choices are even permissible, since the chosen series may not appear in a cointegrating relationship with any of the other series.

In this paper, we consider the properties of eigenvectors of the averaged periodogram matrix of tapered, differenced observations using a fixed amount of averaging. Such eigenvectors are invariant to the labeling of the variables, and are also invariant to additive polynomial trends. Since we hold the amount of averaging fixed, we do not need to estimate the (common) memory parameter of the original series. Such estimation, which cannot be carried out accurately in the presence of fractional cointegration, was needed in Robinson and Yajima (2002), who studied the problem of determining the cointegrating rank based on the eigenvalues of a standardized averaged periodogram matrix, where the amount of averaging tends to ∞. (See also Marinucci and Robinson 2002).

We will first derive the limit distribution of the averaged periodogram of the tapered, differenced data, generalizing results obtained by Chen and Hurvich (2002) to the multivariate case. The averaged periodogram matrix converges in distribution to a singular matrix with null space equal to the space of cointegrating vectors. We then use recent results by Barlow and Slapničar (2000) on perturbed (non-stochastic) singular symmetric matrices to obtain bounds in probability on the angle between the data-based eigenvectors and the space of cointegrating vectors. We then derive the rate of convergence of the r smallest eigenvalues of the averaged periodogram matrix, and present a model selection criterion, similar to one presented in Robinson and Yajima (2002), which allows for the consistent estimation of the cointegrating rank, r. We then present an application of our methodology to the analysis of interest rates. Next, we give a discussion, including possible generalizations of our methodology to a situation where there are varying degrees of cointegration. We conclude with a mathematical appendix.

2 Model and Averaged Periodogram Matrix

Suppose that the original data are a $q \times 1$ time series such that the $p - 1^{th}$ differences $\{y_t\}$ are weakly stationary with a common memory parameter $d \in (-p + 1/2, 1/2)$, where $p \geq 1$ is a fixed integer. The use of $p - 1^{th}$ differences converts any additive polynomial trend of order $p - 1$ in the original series into an additive constant. The value of this constant is irrelevant for our purposes since the estimators considered here are functions of the discrete Fourier transform at nonzero Fourier frequencies. We can therefore take the mean of $\{y_t\}$ to be zero, without loss of generality.

In order to guarantee that the cointegrating relationships in the stochastic component of the levels are preserved in the differences, we apply a taper to the differences, that is, we multiply the differences by a sequence of constants prior to Fourier transformation. A convenient family of tapers for use on the differences was given in Hurvich and Chen (2000), where it was used for semiparametric estimation of the memory parameter. The same family of tapers was used on differences in Chen and Hurvich (2002) for the tapered NBLS of fractional cointegration in a bivariate process.

We assume that the differenced series $\{y_t\}$ has the generalized common-components representation

$$y_t = Ax_t + Bu_t \quad , \tag{1}$$

where A is a $q \times (q - r)$ unknown nonstochastic matrix of rank $q - r$ $(1 \leq r < q)$, B is a $q \times r$ unknown nonstochastic matrix of rank r, $\{x_t\}$ is a $(q - r) \times 1$ unobserved series, the entries of which are called *common components*, and $\{u_t\}$ is an $r \times 1$ unobserved error series. We further assume that each entry of $\{x_t\}$ has memory parameter d, and each entry of $\{u_t\}$ has memory parameter d_u, where $d, d_u \in (-p+1/2, 1/2)$ and $d_u < d$.

Here, we specify a linear model for the vector series $(x_t', u_t')'$. Let ψ_k be a sequence of $q \times q$ matrices such that

$$\psi_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\lambda}\mathbf{\Psi}(\lambda)\, d\lambda \quad ,$$

where for each $\lambda \in [-\pi, \pi]$, $\mathbf{\Psi}(\lambda)$ is a complex-valued matrix such that $\mathbf{\Psi}(-\lambda) = \overline{\mathbf{\Psi}(\lambda)}$ and ψ_0 is an identity matrix. Define the $q \times 1$ vector process $(x_t', u_t')'$ as

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k} \quad ,$$

where $\left\{\varepsilon_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,q})'\right\} \sim iid\left(\mathbf{0}, 2\pi\mathbf{\Sigma}\right)$, $\mathbf{\Sigma}$ is a symmetric positive definite matrix with diagonal entries σ_{aa}^2 and off diagonal entries $\sigma_{ab} = \sigma_{ba}$, $a \neq b$, $a, b \in \{1, \ldots, q\}$ and $E\left\|\varepsilon_t\right\|^4 < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. The spectral density matrix of $(x_t', u_t')'$ is

$$\mathbf{f}(\lambda) = \mathbf{\Psi}(\lambda)\,\mathbf{\Sigma}\mathbf{\Psi}^*(\lambda) \quad , \quad \lambda \in [-\pi, \pi] \quad ,$$

where the superscript $*$ denotes conjugate transposition. We assume that all entries of x_t are $I(d)$ processes (that is, integrated of order d) and all entries of u_t are $I(d_u)$ processes with $-p + 1/2 < d_u < d < 1/2$. We further assume that for $\lambda \in [-\pi, \pi]$, the (a, b)th entry of $\mathbf{\Psi}(\lambda)$ is given by

$$\Psi_{ab}(\lambda) = \left(1 - e^{-i\lambda}\right)^{-d_{ab}} \tau_{ab}(\lambda)\, e^{i\phi_{ab}(\lambda)} \tag{2}$$

where for $a \in \{1, \ldots, q-r\}$, $d_{aa} = d$, $d_{ab} \leq d$ if $b \neq a$ and for $a \in \{q-r+1, \ldots, q\}$, $d_{aa} = d_u$, $d_{ab} \leq d_u$ if $b \neq a$, $\tau_{ab}(\cdot)$ are positive even real-valued functions, $\phi_{ab}(\cdot)$ are odd real-valued functions, all continuously differentiable in an interval containing zero. It follows from (2) that the first derivatives of $\Psi_{ab}(\lambda)$ satisfy

$$\Psi_{ab}'(\lambda) = O\left(\left|\Psi_{aa}(\lambda)\Psi_{bb}(\lambda)\right|^{1/2}|\lambda|^{-1}\right) \quad . \tag{3}$$

It also follows from (2) that we can write the spectral density matrix as

$$\mathbf{f}(\lambda) = \mathbf{\Upsilon}(\lambda)\mathbf{f}^{\dagger}(\lambda)\,\mathbf{\Upsilon}^*(\lambda) \quad , \tag{4}$$

where $\mathbf{\Upsilon}(\lambda) = \operatorname{diag}\left\{\left(1-e^{-i\lambda}\right)^{-d}, \ldots, \left(1-e^{-i\lambda}\right)^{-d}, \left(1-e^{-i\lambda}\right)^{-d_u}, \ldots, \left(1-e^{-i\lambda}\right)^{-d_u}\right\}$, i.e, the first $q-r$ diagonal entries are $\left(1-e^{-i\lambda}\right)^{-d}$ and the remaining diagonal entries are $\left(1-e^{-i\lambda}\right)^{-d_u}$, and $\mathbf{f}^{\dagger}(\lambda)$ is nonnegative definite, Hermitian, continuous at zero frequency, and therefore real-valued at zero frequency. We further assume that $\mathbf{f}_{xx}^{\dagger}(0)$, the $(q-r) \times (q-r)$ leading diagonal block matrix of $\mathbf{f}^{\dagger}(0)$, is positive definite. Thus, $\{x_t\}$ is not fractionally cointegrated, but $\{u_t\}$ is allowed to be fractionally cointegrated (see Robinson and Marinucci, 1998). Our assumptions imply that $\{y_t\}$ is fractionally cointegrated. Indeed, for any nonzero $q \times 1$ vector α in the null space of A', denoted by $Ker(A')$, the linear combination $\{\alpha' y_t\}$ has memory parameter less than or equal to d_u, with equality holding if $\{u_t\}$ is not fractionally cointegrated. The space of cointegrating vectors $Ker(A')$ has dimension r, where r is the *cointegrating rank*, assumed known.

For any vector sequence of observations $\{\xi_t\}_{t=1}^n$, define the tapered discrete Fourier transform by

$$w_{\xi,j} = \frac{1}{\sqrt{2\pi \sum \left| h_t^{p-1} \right|^2}} \sum_{t=1}^{n} h_t^{p-1} \xi_t e^{i\mu_j t} \quad ,$$

where $\mu_j = 2\pi j/n$ is the j'th Fourier frequency, and $\{h_t\}$ is the complex-valued taper of Hurvich and Chen (2000),

$$h_t = 0.5\left(1 - e^{i2\pi(t-0.5)/n}\right) \quad , \quad t = 1, \ldots, n \quad .$$

Note that $p = 1$ yields the no-tapering case. Next, define the tapered cross-periodogram matrix of two vector sequences $\{\xi_t\}_{t=1}^n$ and $\{\zeta_t\}_{t=1}^n$ by

$$I_{\xi\zeta,j} = w_{\xi,j} w_{\zeta,j}^* \quad .$$

We will work with the (real part of the) averaged periodogram matrix of a sample of n observations $\{y_t\}_{t=1}^n$,

$$I_m = \sum_{j=1}^{m} Re(I_{yy,j}) \quad ,$$

where m is a fixed positive integer, $m \geq q - r$. It follows from (1) that

$$I_m = A \sum_{j=1}^{m} Re(I_{xx,j})\, A' + A \sum_{j=1}^{m} Re(I_{xu,j}) B' + B \sum_{j=1}^{m} Re(I_{ux,j}) A' + B \sum_{j=1}^{m} Re(I_{uu,j}) B' \quad . \tag{5}$$

We will show that the righthand side of (5) is dominated by the first term,

$$M = A \sum_{j=1}^{m} Re(I_{xx,j})\, A' \quad . \tag{6}$$

We will also show that, with probability approaching one, $\sum_{j=1}^{m} Re(I_{xx,j})$ is positive definite, so that M is a singular $q \times q$ matrix of rank $q - r$, and $Ker(M) = Ker(A')$. Thus, with probability approaching one, the space of cointegrating vectors is precisely the null space of M (except for the zero vector), that is, the space of eigenvectors of M with corresponding eigenvalue zero. Note that both I_m and M are symmetric nonnegative definite matrices, so all of the eigenvalues of both matrices are real and nonnegative. Furthermore, all eigenvectors of I_m corresponding to distinct eigenvalues are mutually orthogonal.

From the discussion above, it seems plausible that the eigenvectors of I_m corresponding to its r smallest eigenvalues should be close in some sense to the space of cointegrating vectors. The primary goal of this paper is to prove a precise version of this claim. It must be stressed, however, that the eigenvectors of I_m cannot be considered as estimators of the cointegrating vectors. Even though it is useful to view I_m as a perturbed version of M (with a "small" symmetric perturbation), it is not necessarily true that any of the eigenvectors of I_m converge in a meaningful sense to any of the eigenvectors of M, even after appropriate labeling and standardization. Instead, we seek to show that any set of r orthonormal eigenvectors $\{\alpha_i\}_{i=1}^r$ of I_m corresponding to the r smallest eigenvalues is, with high probability, close to the space of cointegrating vectors, in the sense that $\sin\Theta = O_p(n^{d_u - d})$, where $\sin\Theta$ is the square root of the sum of the squared lengths of the residuals from the orthogonal projections of $\{\alpha_i\}_{i=1}^r$ on the space of cointegrating vectors.

3 Limiting Distribution of the Averaged Periodogram

The limiting distribution of the averaged periodogram follows from that of the tapered discrete Fourier transform. Let $\mathbf{G}$ be the $q \times q$ matrix-valued spectral measure on $[-\pi, \pi]$ defined by $\mathbf{G}(d\lambda) = \mathbf{f}(\lambda)d\lambda$. Let $\mathbf{G}_n$ be the $q \times q$ renormalized spectral measure on $\mathbb{R}$ defined by

$$\mathbf{G}_n(dx) = \mathbf{\Lambda}_n \mathbf{G}(dx/n)\mathbf{\Lambda}_n = \mathbf{\Lambda}_n \mathbf{\Psi}(x/n)\mathbf{\Sigma}\mathbf{\Psi}^*(x/n)\mathbf{\Lambda}_n\, dx$$

where $\mathbf{\Lambda}_n = \text{diag}\left(n^{-d}, \ldots, n^{-d}, n^{-d_u}, \ldots, n^{-d_u}\right)$, i.e., the first $q - r$ diagonal entries are n^{-d} and the remaining diagonal entries are n^{-d_u}. It follows from our assumptions that there exists a Hermitian nonnegative definite $q \times q$ matrix-valued measure $\mathbf{G}_0$ on $\mathbb{R}$ such that $\mathbf{G}_n(S) \to \mathbf{G}_0(S)$ as $n \to \infty$ for all bounded Lebesgue measurable sets S. For $x > 0$, we have

$$\mathbf{G}_0\,(dx) = \mathbf{\Pi}\,(x)\,\mathbf{f}^{\dagger}(0)\mathbf{\Pi}^*\,(x)\,dx \tag{7}$$

and $\mathbf{G}_0\,(-dx) = \overline{\mathbf{G}_0\,(dx)}$ where

$$\mathbf{\Pi}(x) = \text{diag}\left(e^{-i\pi d/2}\,|x|^{-d}, \ldots, e^{-i\pi d/2}\,|x|^{-d}, e^{-i\pi d_u/2}\,|x|^{-d_u}, \ldots, e^{-i\pi d_u/2}\,|x|^{-d_u}\right) \quad .$$

We will make use of the spectral representation for the vector process $\{\epsilon_t\}$,

$$\epsilon_t = \int_{-\pi}^{\pi} e^{i\lambda t} dZ_{\epsilon}(\lambda)$$

where $dZ_\epsilon(\lambda)$ is a $(q \times 1)$ complex-valued random vector with the following properties:

$$\overline{dZ_\epsilon(-\lambda)} = dZ_\epsilon(\lambda) \quad , \quad E[dZ_\epsilon(\lambda)] = 0 \quad , \tag{8}$$

$$E[dZ_\epsilon(\lambda)dZ_\epsilon^*(\mu)] = 0 \quad (\mu \neq \lambda) \quad , \quad E[dZ_\epsilon(\lambda)dZ_\epsilon^*(\lambda)] = \mathbf{\Sigma} d\lambda. \tag{9}$$

For any bounded set Δ in $\mathbb{R}$, define

$$\psi_\Delta\,(s) = \frac{1}{2\pi}\int_{\Delta/n} e^{-isx}\,\mathbf{\Psi}\,(x)\,dx, \tag{10}$$

$$Z_n\,(\Delta) = n^{1/2}\mathbf{\Lambda}_n \sum_s \psi_\Delta\,(s)\,\varepsilon_s = n^{1/2}\mathbf{\Lambda}_n \int_{\Delta/n} \mathbf{\Psi}\,(x)\,dZ_\varepsilon\,(x)\,. \tag{11}$$

Define $Z_{\mathbf{G}_0}$ as the $(q \times 1)$ multivariate complex Gaussian random measure satisfying

$$E\left[Z_{\mathbf{G}_0}\,(S)\right] = 0, \quad E\left[Z_{\mathbf{G}_0}\,(S)\,Z_{\mathbf{G}_0}^*\,(S)\right] = \mathbf{G}_0\,(S)\,, \quad \overline{Z_{\mathbf{G}_0}\,(-S)} = Z_{\mathbf{G}_0}\,(S)\,,$$

$$E\left[Z_{\mathbf{G}_0}\,(S_1)\,Z_{\mathbf{G}_0}^*\,(S_2)\right] = 0, \text{ if } S_1 \cap S_2 = \emptyset, \tag{12}$$

where S is any Borel set in $\mathbb{R}$.

Lemma 1 *If $\Delta_1, \ldots, \Delta_M$ are intervals in $\mathbb{R}$ with nonzero endpoints and $\pm\Delta_1, \ldots, \pm\Delta_M$ are disjoint, then*

$$\left(Z_n\,(\Delta_1)\,, \ldots, Z_n\,(\Delta_M)\right) \xrightarrow{d} \left(Z_{\mathbf{G}_0}\,(\Delta_1)\,, \ldots, Z_{\mathbf{G}_0}\,(\Delta_M)\right)$$

where the $q \times q$ matrix-valued measure $\mathbf{G}_0\,(S)$ is defined above, and , $Z_{\mathbf{G}_0}$ is the multivariate complex Gaussian random measure satisfying the properties given in (12).

Given the process $(x_t', u_t')' = \sum_k \psi_k \varepsilon_{t-k}$ described above, consider the m tapered DFT vectors $w_1^T, \ldots, w_m^T$,

$$w_j^T = \left[\begin{array}{c} w_{x,j}^T \\ w_{u,j}^T \end{array} \right] = \frac{1}{\sqrt{2\pi \sum_{t=1}^n \left| h_t^{p-1} \right|^2}} \sum_{t=1}^n h_t^{p-1} \left[\begin{array}{c} x_t \\ u_t \end{array} \right] \exp\left(i\mu_j t\right), \quad j = 1, \ldots, m.$$

It is useful to note that $\sum_{t=1}^n \left| h_t^{p-1} \right|^2 = nc_p$, where

$$c_p = 2^{-2(p-1)} \left(\begin{array}{c} 2p-2 \\ p-1 \end{array} \right) \quad .$$

We define the function (for $x \in \mathbb{R}$)

$$\Delta_p(x) = \left(\begin{array}{c} 2p-2 \\ p-1 \end{array} \right)^{-1/2} \sum_{k=0}^{p-1} \left(\begin{array}{c} p-1 \\ k \end{array} \right) (-1)^k \Delta(x + 2\pi k) \quad ,$$

where

$$\Delta(x) = \frac{1}{\sqrt{2\pi}} \frac{e^{ix} - 1}{ix} \quad .$$

Theorem 1 *As $n \to \infty$, with m a fixed positive integer,*

$$\{\Lambda_n w_j^T\}_{j=1}^m \xrightarrow{d} \left\{ \int_{\mathbb{R}} \Delta_p(x + 2\pi j)\, dZ_{\mathbf{G}_0}(x) \right\}_{j=1}^m \quad ,$$

where $\mathbf{\Lambda}_n = \operatorname{diag}\left(n^{-d}, \ldots, n^{-d}, n^{-d_u}, \ldots, n^{-d_u}\right)$ and $Z_{\mathbf{G}_0}$ is the multivariate complex Gaussian random measure satisfying the properties given in (12).

We partition $\mathbf{G}_0$ into four sub-matrices

$$\mathbf{G}_0 = \left[\begin{array}{cc} \mathbf{G}_{0,xx} & \mathbf{G}_{0,xu} \\ \mathbf{G}_{0,ux} & \mathbf{G}_{0,uu} \end{array} \right],$$

where $\mathbf{G}_{0,xx}$ is a $(q-r) \times (q-r)$ matrix and $\mathbf{G}_{0,uu}$ is an $r \times r$ matrix. Defining

$$\upsilon_j(x) = \frac{1}{2} \left[\overline{\Delta_p(-x + 2\pi j)} + \Delta_p(x + 2\pi j) \right],$$
$$\nu_j(x) = \frac{i}{2} \left[\overline{\Delta_p(-x + 2\pi j)} - \Delta_p(x + 2\pi j) \right],$$

we have the following Corollary.

Corollary 1 *Under the conditions of Theorem 1,*

$$n^{-2d} I_m \xrightarrow{d} \mathbf{A} \sum_{j=1}^m \left(U_j U_j' + V_j V_j' \right) \mathbf{A}' \quad , \tag{13}$$

where $\mathrm{vec}\left\{U_j, V_k\right\}_{j,k=1}^m$ *is a* $2m(q-r)$*-variate normal random variable with zero mean, and covariance determined by*

$$E\left(U_j U_k'\right) = \int_{\mathbb{R}} \upsilon_j\left(x\right) \overline{\upsilon_k\left(x\right)} \mathbf{G}_{0,xx}\left(dx\right),$$
$$E\left(V_j V_k'\right) = \int_{\mathbb{R}} \nu_j\left(x\right) \overline{\nu_k\left(x\right)} \mathbf{G}_{0,xx}\left(dx\right),$$
$$E\left(U_j V_k'\right) = \int_{\mathbb{R}} \upsilon_j\left(x\right) \overline{\nu_k\left(x\right)} \mathbf{G}_{0,xx}\left(dx\right).$$

Lemma 2 *Let* Ξ *denote the covariance matrix of* $\mathrm{vec}\{U_1, \ldots, U_m, V_1, \ldots, V_m\}$ *in Corollary 1. Then* Ξ *is positive definite.*

Remark: *It follows from the proof of Lemma 2 that if* $\{x_t\}$ *were cointegrated, then* Ξ *would not be positive definite. Thus, positive definiteness of* Ξ *provides a characterization of lack of cointegration in* $\{x_t\}$.

Following from Lemma 2 and the Theorem of Okamoto (1973), we have the following Corollary.

Corollary 2 *The averaged periodgram of* $\{x_t\}$,

$$n^{-2d} \sum_{j=1}^m \mathrm{Re}\left(I_{xx,j}\right),$$

under its limiting distribution, is positive definite with probability one.

Proof. Let $\mathbf{U} = (U_1, \ldots, U_m)$ and $\mathbf{V} = (V_1, \ldots, V_m)$, then the limiting distribution of $n^{-2d} \sum_{j=1}^m \mathrm{Re}\left(I_{xx,j}\right)$ is

$$\sum_{j=1}^m \left(U_j U_j' + V_j V_j'\right) = \left(\mathbf{UU}' + \mathbf{VV}'\right).$$

Since $\mathrm{vec}\{U_1, \ldots, U_m, V_1, \ldots, V_m\}$ is a $2m(q-r)$-variate normal random variable with positive definite covariance matrix, the joint distribution of $\mathrm{vec}\{U_1, \ldots, U_m, V_1, \ldots, V_m\}$ is absolutely continuous. Since we are assuming that $m \geq q-r$, it follows from the theorem of Okamoto (1973) that both $\mathbf{UU}'$ and $\mathbf{VV}'$ are positive definite with probability one $\square$

It follows from (5) together with Corollary 1 and the proof of Corollary 2 that the limiting distribution of $n^{-2d}I_m$ is the same as the limiting distribution of $n^{-2d}M$, where M is given by (6).

4 Angle between estimated and true spaces of cointegrating vectors

If we define $\tilde{H} = n^{-2d}I_m$ and $H = n^{-2d}M$, then from (5) we can write

$$\tilde{H} = H + \Delta H \quad , \tag{14}$$

where the entries of the $q \times q$ matrix ΔH are $O_p(n^{d_u - d})$. The spaces of eigenvectors of I_m and $\tilde{H}$ are identical, but it is convenient here to work with $\tilde{H}$. We see from (14) that $\tilde{H}$ is a perturbed version of the singular random symmetric matrix H, and the perturbation is a random symmetric matrix with entries that are $o_p(1)$. This setup in the non-stochastic case was studied by Barlow and Slapničar (2000), and we will generalize their results to the stochastic situation, (14). We will use the notation of Barlow and Slapničar (2000), who considered Hermitian matrices with Hermitian perturbations. Since our matrices H and ΔH are real and symmetric, all eigenvectors are real.

Consider the family of perturbed matrices

$$H(\zeta) = H + \zeta \Delta H \quad , \quad \zeta \in [0,1] \quad .$$

Note that $H = H(0)$, and $\tilde{H} = H(1)$. For each value of ζ, let $0 \leq \lambda_1(\zeta) \leq \lambda_2(\zeta) \ldots \leq \lambda_q(\zeta)$ be the eigenvalues of $H(\zeta)$, and let $\chi_i(\zeta)$, $i \in \{1, \ldots, q\}$, be a corresponding set of orthonormal eigenvectors, so that

$$H(\zeta)\chi_i(\zeta) = \lambda_i(\zeta)\chi_i(\zeta) \quad , \quad i = 1, \ldots, q \quad .$$

We assume that $Ker(H)$ has dimension r, so that $\lambda_1(0) = \ldots = \lambda_r(0) = 0$.

Define

$$\mathcal{J} = \{i : H(\zeta)\chi_i(\zeta) \neq 0 \text{ for all } \zeta \in [0,1]\} \quad . \tag{15}$$

The set $\mathcal{J}$ and its complement $\mathcal{J}^C$ play an important role in the results of Barlow and Slapničar (2000) in the nonstochastic case. Note that $\mathcal{J}$ is the set of indices i such that the i^{th} smallest eigenvalue of $H(\zeta)$ remains positive for all perturbations $\zeta \in [0,1]$. Clearly, $\{1, \ldots r\} \subset \mathcal{J}^C$. Since the matrices $H(\zeta)$ are random, $\mathcal{J}$ and $\mathcal{J}^C$ are random subsets of $\{1, \ldots q\}$. It seems plausible that since $dim(Ker(H)) = r$, and since the difference between H and $H(\zeta)$ is $o_p(1)$, there is a very high probability that $\mathcal{J}^C = \{1, \ldots r\}$. The following lemma is proved in the Appendix.

Lemma 3 *$Prob\{\mathcal{J}^C \neq \{1, \ldots, r\}\} \to 0$ as $n \to \infty$.*

Define

$$X_1 = [\chi_{r+1}(0), \ldots, \chi_q(0)] \quad ,$$

a matrix of orthonormal eigenvectors of H associated with its $q - r$ largest eigenvalues, $\lambda_{r+1}(0), \ldots, \lambda_q(0)$. Now, define

$$\sin \Theta = ||X_1^* \tilde{X}_2||_F \quad ,$$

where for any matrix E, $||E||_F = \sqrt{\sum |E_{ij}|^2}$ is the Frobenius norm, and

$$\tilde{X}_2 = [\chi_1(1), \ldots, \chi_r(1)]$$

is a matrix of orthonormal estimated cointegrating vectors, that is, eigenvectors of $\tilde{H}$ associated with its r smallest eigenvalues, $\lambda_1(1), \ldots, \lambda_r(1)$. We can think of $\sin \Theta$ as the sine of the angle Θ between the space spanned by the estimated cointegrating vectors and the space spanned by the true cointegrating vectors. A related interpretation of $\sin \Theta$ in terms of orthogonal projections was given in Section 2. Roughly speaking, the smaller $\sin \Theta$ is, the closer the spaces of estimated and true cointegrating vectors are to each other. Barlow and Slapničar (2000) show that if $\mathcal{J} = \{r+1, \ldots, q\}$ then

$$\sin \Theta = ||X_1^* \tilde{X}_2||_F \leq \frac{||H^\dagger(0) \Delta H X_2||_F}{\text{relgap}_0(\Lambda_1, \tilde{\Lambda}_2)} \quad , \tag{16}$$

where the quantities on the righthand side of this inequality are defined below, without requiring the assumption that $\mathcal{J} = \{r+1, \ldots, q\}$. We define

$$X_2 = [\chi_1(0), \ldots, \chi_r(0)] \quad ,$$

an orthogonal matrix of true cointegrating vectors. We can write $H(0) = X(0)\Lambda(0)X^*(0)$, where

$$X(0) = [\chi_1(0), \ldots, \chi_q(0)] \quad ,$$

and $\Lambda(0) = diag(0, \ldots, 0, \lambda_{r+1}(0), \ldots, \lambda_q(0))$. The Moore-Penrose inverse of $H(0)$ is given by

$$H^\dagger(0) = X(0)\Lambda^\dagger(0)X^*(0) \quad ,$$

where $\Lambda^\dagger(0) = diag(0, \ldots, 0, \lambda_{r+1}^{-1}(0), \ldots, \lambda_q^{-1}(0))$. The matrices Λ_1 and $\tilde{\Lambda}_2$ are defined by $\Lambda_1 = diag(\lambda_{r+1}(0), \ldots, \lambda_q(0))$, and $\tilde{\Lambda}_2 = diag(\lambda_1(1), \ldots, \lambda_r(1))$. The generalized relative gap in the eigenvalues of Λ_1, $\tilde{\Lambda}_2$ is given by

$$\text{relgap}_0(\Lambda_1, \tilde{\Lambda}_2) = \min_{j \in \{1, \ldots, r\}\,,\, i \in \{r+1, \ldots, q\}} \left| \frac{\lambda_i(0) - \lambda_j(1)}{\lambda_i(0)} \right| \quad .$$

Since the eigenvalues of a matrix are continuous functions of the entries of the matrix, and since $\Delta H = o_p(1)$, it follows that $\lambda_1(1), \ldots, \lambda_r(1)$ all converge in probability to zero, and the limiting distribution of $\lambda_{r+1}(1), \ldots, \lambda_q(1)$ is the same as the limiting distribution of $\lambda_{r+1}(0), \ldots, \lambda_q(0)$. Since H, under its limiting distribution, has rank $q-r$ with probability one, it follows that the limiting marginal distributions of $\lambda_{r+1}(0), \ldots, \lambda_q(0)$ have no mass at zero. Thus,

$$\text{relgap}_0(\Lambda_1, \tilde{\Lambda}_2) \xrightarrow{P} 1 \quad . \tag{17}$$

Since $\Delta H = O_p(n^{d_u - d})$, we have

$$||H^\dagger(0)\Delta H X_2||_F = O_p(n^{d_u - d}) \quad . \tag{18}$$

Note that both (17) and (18) hold without the need to assume that $\mathcal{J} = \{r+1, \ldots, q\}$.

We now present our main theorem.

Theorem 2 $\sin\Theta = O_p(n^{d_u - d})$.

Proof of Theorem 2: We need to show that $n^{d-d_u}\sin\Theta$ is bounded in probability, that is,

$$\forall \epsilon > 0 \; \exists M_\epsilon \; : \; Prob\{n^{d-d_u}\sin\Theta > M_\epsilon\} < \epsilon \; \forall \, \text{sufficiently large} \, n \quad .$$

For any constant C,

$$Prob\{n^{d-d_u}\sin\Theta > C\} = Prob\{n^{d-d_u}\sin\Theta > C \,|\, \mathcal{J} = \{r+1, \ldots, q\}\} Prob\{\mathcal{J} = \{r+1, \ldots, q\}\}$$
$$+ Prob\{n^{d-d_u}\sin\Theta > C \,|\, \mathcal{J} \neq \{r+1, \ldots, q\}\} Prob\{\mathcal{J} \neq \{r+1, \ldots, q\}\} \quad . \tag{19}$$

The final term on the righthand side of (19) tends to zero as $n \to \infty$ by Lemma 3. Define

$$R = n^{d-d_u}||H^\dagger(0)\Delta H X_2||_F / \text{relgap}_0(\Lambda_1, \tilde{\Lambda}_2) \quad .$$

By (16), we have

$$Prob\{n^{d-d_u}\sin\Theta > C \mid \mathcal{J} = \{r+1,\dots,q\}\} \le \frac{Prob\{(R>C)\cap(\mathcal{J}=\{r+1,\dots,q\})\}}{Prob\{\mathcal{J}=\{r+1,\dots,q\}\}}$$

$$\le \frac{Prob\{R>C\}}{Prob\{\mathcal{J}=\{r+1,\dots,q\}\}} \quad .$$

From (17) and (18), we conclude that $R = O_p(1)$, so we can say

$$\forall\epsilon>0\ \exists M^*_\epsilon\ :\ Prob\{R > M^*_\epsilon\} < \epsilon/2\ \forall\,\text{sufficiently large}\,n \quad .$$

Then for the same ϵ and M^*_ϵ as above and all sufficiently large n, we have

$$Prob\{n^{d-d_u}\sin\Theta > M^*_\epsilon\} < \frac{\epsilon/2}{Prob\{\mathcal{J}=\{r+1,\dots,q\}\}} + o(1) < \epsilon \quad \Box$$

5 Consistent Estimation of the Cointegrating Rank

Robinson and Yajima (2002) proposed to use a model selection criterion described in Fujikoshi (1985), Fujikoshi & Veitch (1979) and Gunderson & Muirhead (1997) to estimate the cointegrating rank, $r > 0$. Here, we adapt this criterion to our setting, obtaining a consistent estimate of r. The method requires some information about d and d_u. This shortcoming applies also to the analogous method presented in Robinson and Yajima (2002), who require that d be estimated, using a bandwidth which cannot be adequately set without the knowledge of a lower bound for $d - d_u$.

We start by obtaining a rate of convergence for the r smallest eigenvalues $\lambda_1(1),\dots,\lambda_r(1)$ of $\tilde{H}$.

Lemma 4 $(\lambda_1(1),\dots,\lambda_r(1)) = O_p(n^{d_u-d})$.

Proof of Lemma 4: Since $\lambda_r(1)\chi_r(1) = \tilde{H}\chi_r(1)$, using $||\cdot||_2$ for the Euclidean norm, we have

$$\lambda_r^2(1) = ||\tilde{H}\chi_r(1)||_2^2 \le 2||H\chi_r(1)||_2^2 + 2||\Delta H\chi_r(1)||_2^2 = 2||H\chi_r(1)||_2^2 + O_p(n^{2(d_u-d)}) \quad . \tag{20}$$

Note that

$$H\chi_r(1) = X(0)\Lambda(0)X^*(0)\chi_r(1) = X(0)\begin{pmatrix} 0 \\ \vdots \\ \lambda_{r+1}(0)\chi'_{r+1}(0)\chi_r(1) \\ \vdots \\ \lambda_{r+1}(0)\chi'_q(0)\chi_r(1)\end{pmatrix} \quad .$$

The entries of $X(0)$ are $O(1)$. Since H converges in distribution, we have $(\lambda_{r+1}(0),\dots,\lambda_q(0)) = O_p(1)$. By Theorem 2,

$$\chi'_j(0)\chi_r(1) = O_p(n^{d_u-d}) \quad , \quad j = r+1,\dots,q \quad .$$

It follows that $H\chi_r(1) = O_p(n^{d_u-d})$. The lemma now follows from (20) $\Box$

Our estimator of r is based on eigenvalues of I_m, which may be obtained directly from the observed data. We denote these eigenvalues by $\hat{\delta}_1 \leq \ldots \leq \hat{\delta}_q$, so that $\hat{\delta}_i = n^{2d}\lambda_i(1)$ for $i = 1, \ldots, q$. Define

$$\hat{\sigma}_{j,k} = \sum_{i=j}^{k} \hat{\delta}_i \quad .$$

Let $\hat{r}$ be the minimizer of $L(u)$ for $1 \leq u < q$, where

$$L(u) = V(n)(q-u) - \hat{\sigma}_{u+1,q} \quad , \tag{21}$$

and $V(n)$ is a deterministic sequence. Under certain conditions on $V(n)$, the estimator $\hat{r}$ is consistent for r.

Theorem 3 *If $V(n)$ is a deterministic sequence such that*

$$\frac{n^{d_u+d}}{V(n)} + \frac{V(n)}{n^{2d}} \to 0 \quad ,$$

then $Prob\{\hat{r} = r\} \to 1$.

Proof of Theorem 3: We follow the same lines as the proof of Theorem 4 of Robinson and Yajima (2002). We have

$$Prob\{\hat{r} > r\} \leq \sum_{u=r+1}^{q} Prob\{L(u) < L(r)\} \leq q\, Prob\{\hat{\delta}_{r+1} < V(n)\} \to 0 \quad ,$$

since $\hat{\delta}_{r+1} = O_p(n^{2d})$ and $V(n)/n^{2d} \to 0$. Next,

$$Prob\{\hat{r} < r\} \leq \sum_{u=1}^{r-1} Prob\{L(u) < L(r)\} \leq r\, Prob\{V(n) < \hat{\delta}_r\} \to 0 \quad ,$$

since $\hat{\delta}_r = O_p(n^{d_u+d})$ and $V(n)/n^{d_u+d} \to \infty$ $\square$

6 Application

We applied the methods of this paper to a multivariate series of interest rates on United States Treasury securities, with maturities of 3 months, 6 months, 1 year, 3 years, 5 years, 7 years, 10 years and 30 years. The observations were daily, spanning the period from January 1, 1982 to Dec 31, 2001. The sample size is $n = 4999$. The data were obtained from the database of the Federal Reserve Board, at http://www.federalreserve.gov/releases/. We took logarithms of the interest rates, differenced the logarithms once, and applied a first-order taper. We then computed eigenvalues and orthonormal eigenvectors of the resulting averaged periodogram matrix, using a bandwidth of $m = 10$. Estimates of the memory parameter of the tapered differences were not significantly different from zero for any of the series. This provides some justification for our assumption that all of the series share the same memory parameter. Figure 1 plots the log interest rates for two short-term maturities (3 months, 6 months) and two long-term maturities (7 years, 10 years). Two groups are apparent, one for the short-term rates, the other for the long-term rates. Table 1 gives the eigenvalues λ_j of the averaged periodogram matrix, sorted in order from lowest to highest, as well as the corresponding eigenvectors, χ_j. The ratio of the largest to the smallest eigenvalue λ_8/λ_1 is 4.7×10^4. The largeness of this ratio is an indication of near-singularity of the averaged periodogram matrix, and therefore of potential fractional cointegration of the time series. It is perhaps of interest to note that, for χ_1, the only components with an absolute coefficient exceeding 0.1 are those corresponding to maturities of 3 years, 7 years, and 10 years. Furthermore, χ_3 has large coefficients only for maturities of 3 months, 6 months, 1 year and 5 years. For $j \leq 5$, the coefficients of χ_j for 3 months and 6 months are either both small or both large, as are the coefficients for 7 years and 10 years.

We multiplied the transpose of each eigenvector by the multivariate series of logarithms of non-tapered interest rates, yielding eight univariate residual series. For each residual series, we computed the Gaussian semiparametric estimator of the memory parameter of the tapered differences, using a first-order taper and bandwidths of 30. For the bandwidth of 30, we plot in Figure 2 the log periodogram of the tapered differenced series versus log frequency for each of the residual series. It can be seen that the differenced residual series corresponding to the largest eigenvalues seem to have a memory parameter of nearly zero, as the scatterplot shows a slope of approximately zero, while those corresponding to the smaller eigenvalues seem to have increasingly negative memory parameters. This suggests the potential presence of fractional cointegration.

The Gaussian semiparametric estimators of the memory parameters d_u for each series are given in Figure 2. In each case, the approximate standard error for the estimator is 0.146. The standard error was computed using the finite-sample expression given in Hurvich and Chen (2000). Only the series corresponding to the six smallest eigenvalues have estimated memory parameters which are significantly less than zero. Furthermore, there is evidence that the underlying memory parameters corresponding to these six residual series are not all the same. This suggests that some of the cointegrating relationships are stronger than others. We will discuss this point further in the next section.

It should be stressed that there is as yet no full theoretical basis to justify the above Gaussian semiparametric estimators. However, it seems clear that the presence of fractional cointegration would place rather strict upper bounds on the bandwidth that could be used. This is the reason why we have used the relatively small bandwidth of 30. We also tried a bandwidth of 70. This gave estimated memory parameters that are somewhat closer to zero, but which may be contaminated by bias.

We computed $\hat{r}$ as the minimizer of $L(u)$ given by (21). We tried several choices for the tuning parameter $V(n)$. Most of these choices yielded either $\hat{r} = 6$ or $\hat{r} = 7$. Overall, this suggests a cointegrating rank of $r = 6$, since the estimated memory parameter is significantly negative for χ_6 but not for χ_7.

7 Discussion

In the case of classical nonparametric cointegration, in which (d, d_u) is known to be $(1, 0)$, Bierens (1997) obtained estimators of the cointegrating vectors based on solutions to a generalized eigenvalue problem, and showed that for these estimators $\sin\Theta = O_p(n^{-1})$, where $\sin\Theta$ is defined in the same way as in the current paper. It might be interesting to compare Bierens' (1997) estimators to ours in the case $(d, d_u) = (1, 0)$, especially since one of the matrices used in Bierens' generalized eigenvalue problem is closely related to the averaged periodogram matrix, with a fixed degree, m, of smoothing.

Although we have considered the consistent estimation of the cointegrating rank r assuming that cointegration is known to exist ($r > 0$), we have not presented a test for the presence of fractional cointegration. A Hausman-type test for fractional cointegration was proposed by Marinucci and Robinson (2002), where it was shown to work well in simulations. Presumably, the methods of the current paper could be safely applied once the null hypothesis of no cointegration ($r = 0$) is rejected by such a test.

We would like to stress here that the results obtained in this paper allow the cointegrating errors to have different memory parameters. Since we do not assume $\mathbf{f}_{uu}^{\dagger}(0)$ to be positive definite, it follows that $\{u_t\}$, unlike $\{x_t\}$, is allowed to be cointegrated. If $\{u_t\}$ is cointegrated then analogously to (1) we can write

$$u_t = \tilde{A}_1 u_t^{(1)} + \tilde{A}_2 u_t^{(2)}$$

so that

$$y_t = A_0 x_t + A_1 u_t^{(1)} + A_2 u_t^{(2)}$$

where $A_1 = B\tilde{A}_1$ and $A_2 = B\tilde{A}_2$. In general we can elaborate the common trend representation in (1) as follows. Let $\left\{u_t^{(k)}\right\}$, $k = 1, \dots, s$, be $r_k \times 1$ series with memory parameter d_{u_k}, with $\sum_{k=1}^{s} r_k = r$ and $d > d_{u_1} > d_{u_2} > \cdots > d_{u_s}$. Let $\mathbf{A}_k$ be $q \times r_k$ full-rank matrices for $k = 1, \dots, s$. We have the common trend representation

$$y_t = \mathbf{A}_0 x_t + \mathbf{A}_1 u_t^{(1)} + \cdots + \mathbf{A}_s u_t^{(s)}, \tag{22}$$

where A_0 is a full-rank $q \times (q - r)$ matrix as was the matrix A in (1). This representation implies that some of the cointegrating relationships among y_t are stronger than others, as we saw in our analysis of the interest rate data. All of the theoretical results in this paper hold for this elaborated model, if we take $d_u = d_{u_1}$ throughout. Indeed, the elaborated model is simply a special case of the model we have assumed in Section 2.

Under model (22), there will exist cointegrating vectors such that the cointegrating residual has memory parameter d_{u_k} for all k. We are currently exploring the ability of the estimators we have studied in this paper to approximately recover such cointegrating vectors.

8 Appendix

For convenience of notation, throughout this appendix we write $v_x = q - r$, $v_u = r$, and $v = v_x + v_u = q$.

Proof of Lemma 1: From (11) and the properties (8), (9) of the spectral representation, we conclude that $Z_n(\Delta_j) = \overline{Z_n(-\Delta_j)}$ for $j = 1, \dots, M$ and

$$E[\operatorname{Re} Z_n(\Delta_j) \operatorname{Re} Z_n'(\Delta_k)] = E[\operatorname{Re} Z_n(\Delta_j) \operatorname{Im} Z_n'(\Delta_k)] = E[\operatorname{Im} Z_n(\Delta_j) \operatorname{Re} Z_n'(\Delta_k)]$$

$$= E[\operatorname{Im} Z_n(\Delta_j) \operatorname{Im} Z_n'(\Delta_k)] = 0 \quad , \quad (j \neq k) \quad .$$

It therefore suffices to prove that $Z_n(\Delta) \xrightarrow{d} Z_{\mathbf{G}_0}(\Delta)$ for any interval Δ with nonzero endpoints and $\Delta \cap -\Delta = \emptyset$. By the Cramer-Wold device, this is equivalent to showing that for all $\alpha, \beta \in \mathbb{R}^2$, $\alpha' \operatorname{Re} Z_n(\Delta) + \beta' \operatorname{Im} Z_n(\Delta) \xrightarrow{d} \alpha' \operatorname{Re} Z_{\mathbf{G}_0}(\Delta) + \beta' \operatorname{Im} Z_{\mathbf{G}_0}(\Delta)$. Note that $E[Z_n(\Delta) Z_n^*(\Delta)] = \mathbf{G}_n(\Delta) \to \mathbf{G}_0(\Delta)$ as

$n \to \infty$. Note also that from the properties of the spectral representation, $E[Z_n(\Delta)Z_n^*(-\Delta)] = 0$. It follows that

$$
\begin{aligned}
Var[\alpha' \operatorname{Re} Z_n(\Delta) + \beta' \operatorname{Im} Z_n(\Delta)] &= \frac{1}{4}(\alpha' - i\beta')\,\mathbf{G}_n(\Delta)\,(\alpha + i\beta) + \frac{1}{4}(\alpha' + i\beta')\,\overline{\mathbf{G}_n(\Delta)}\,(\alpha - i\beta) \\
&\to \frac{1}{4}(\alpha' - i\beta')\,\mathbf{G}_0(\Delta)\,(\alpha + i\beta) + \frac{1}{4}(\alpha' + i\beta')\,\overline{\mathbf{G}_0(\Delta)}\,(\alpha - i\beta) \\
&= Var[\alpha' \operatorname{Re} Z_{\mathbf{G}_0}(\Delta) + \beta' \operatorname{Im} Z_{\mathbf{G}_0}(\Delta)] := \sigma_0^2 \geq 0 \quad .
\end{aligned}
$$

Note that it is possible that $\sigma_0^2 = 0$ since the limiting normal distribution as asserted in Lemma 1 may have a singular covariance matrix. We will require bounds on the entries of $\psi_\Delta(s,n)$, where here we explicitly denote the dependence on n as well as s. Without loss of generality, we assume that $\Delta = (A, B]$ where $0 < A < B$. We denote the (a,b) entry, $\psi_\Delta(s,n)_{ab}$. Using integration by parts, we have

$$
\psi_\Delta(s,n)_{ab} = \frac{1}{2\pi}\int_{A/n}^{B/n} e^{-isx}\Psi_{ab}(x)\,dx = \frac{1}{2\pi}\frac{1}{-is}e^{-isx}\Psi_{ab}(x)|_{A/n}^{B/n} - \frac{1}{2\pi}\frac{1}{-is}\int_{A/n}^{B/n} e^{-isx}\Psi'_{ab}(x)\,dx \quad .
$$

In the sequel, we use C to denote a generic constant. From Equations (2) and (3) we have, for all sufficiently small $x > 0$, $|\Psi_{ab}(x)| < Cx^{-d_x}$ and $|\Psi'_{ab}(x)| < Cx^{-d-1}$ for $a \in \{1, \dots, v_x\}$, $|\Psi_{ab}(x)| < Cx^{-d_u}$ and $|\Psi'_{ab}(x)| < Cx^{-d_u-1}$ for $a \in \{(v_x+1), \dots, v\}$ and We obtain for all $s \neq 0$

$$
|\psi_\Delta(s,n)_{ab}| \leq \frac{C}{|s|}[(A/n)^{-d} + (B/n)^{-d}] + \frac{C}{|s|}\frac{B-A}{n}[(A/n)^{-d-1} + (B/n)^{-d-1}] \leq Cn^d\frac{1}{|s|} \tag{23}
$$

for $a \in \{1, \dots, v_x\}$ and

$$
|\psi_\Delta(s,n)_{ab}| \leq Cn^{d_u}\frac{1}{|s|}. \tag{24}
$$

for $a \in \{(v_x+1), \dots, v\}$. We can write

$$
\begin{aligned}
\alpha' \operatorname{Re} Z_n(\Delta) + \beta' \operatorname{Im} Z_n(\Delta) &= \alpha' n^{1/2}\mathbf{\Lambda}_n \sum_s \operatorname{Re}\,\psi_\Delta(s)\epsilon_s + \beta' n^{1/2}\mathbf{\Lambda}_n \sum_s \operatorname{Im}\psi_\Delta(s)\epsilon_s \\
&= n^{1/2-d}\sum_{a=1}^{v_x}\sum_{b=1}^{v}\sum_s [\alpha_a \operatorname{Re}\psi_\Delta(s)_{ab}\epsilon_{s,b} + \beta_a \operatorname{Im}\psi_\Delta(s)_{ab}\epsilon_{s,b}] \\
&+ n^{1/2-d_u}\sum_{a=v_x+1}^{v}\sum_{b=1}^{v}\sum_s [\alpha_a \operatorname{Re}\,\psi_\Delta(s)_{ab}\epsilon_{s,b} + \beta_a \operatorname{Im}\,\psi_\Delta(s)_{ab}\epsilon_{s,b}] \\
&= \sum_s W_{ns},
\end{aligned}
$$

where for each n, the $\{W_{ns}\}_{s=-\infty}^{\infty}$ are independent random variables given by

$$
\begin{aligned}
&W_{ns} \\
&= n^{1/2}\sum_{b=1}^{v}\left\{n^{-d}\sum_{a=1}^{v_x}[\alpha_a \operatorname{Re}\,\psi_\Delta(s)_{ab} + \beta_a \operatorname{Im}\psi_\Delta(s)_{ab}] + n^{-d_u}\sum_{a=v_x+1}^{v}[\alpha_a \operatorname{Re}\,\psi_\Delta(s)_{ab} + \beta_a \operatorname{Im}\,\psi_\Delta(s)_{ab}]\right\}\epsilon_{s,b}
\end{aligned} \tag{25}
$$

Since $|\operatorname{Re}\,\psi_\Delta(s)_{ab}| \leq |\psi_\Delta(s)_{ab}|$ and $|\operatorname{Im}\,\psi_\Delta(s)_{ab}| \leq |\psi_\Delta(s)_{ab}|$ for $a, b \in \{1, \dots, v\}$, we conclude that for $s \neq 0$

$$
E[W_{ns}^2] \leq Cn\sum_{b=1}^{v}\left\{n^{-2d}\sum_{a=1}^{v_x}|\psi_\Delta(s)_{ab}|^2 + n^{-2d_u}\sum_{a=v_x+1}^{v}|\psi_\Delta(s)_{ab}|^2\right\} \leq Cn/s^2 \tag{26}
$$

where the final inequality follows from the bounds (23), (24) for the entries of $\psi_\Delta(s,n)$.

Let $V_0(n)$ be a non-decreasing sequence, to be determined later. We have

$$\sum_s W_{ns} = \sum_{|s|\le V_0(n)} W_{ns} + \sum_{|s|>V_0(n)} W_{ns} \quad .$$

Using (26), we have

$$E[|\sum_{|s|>V_0(n)} W_{ns}|^2] = \sum_{|s|>V_0(n)} E[W_{ns}^2] \le Cn \sum_{|s|>V_0(n)} \frac{1}{s^2} \le C\frac{n}{V_0(n)} \quad .$$

If we choose $V_0(n)$ so that $n/V_0(n) \to 0$ as $n \to \infty$, the Lemma will follow if we can show that

$$\sum_{|s|\le V_0(n)} W_{ns} \xrightarrow{d} N(0,\sigma_0^2) \quad if \quad \sigma_0^2 > 0 \quad ,$$

and that

$$\sum_{|s|\le V_0(n)} W_{ns} \xrightarrow{p} 0 \quad if \quad \sigma_0^2 = 0 \quad .$$

In the case $\sigma_0^2 = 0$, the equation above follows since $\sum_{|s|\le V_0(n)} E[W_{ns}^2] = \sigma_0^2 + o(1)$. Now suppose that $\sigma_0^2 > 0$. By the Lyapounov condition (see, e.g., Billingsley 1986, p. 371) it suffices to show that

$$\frac{\sum_{|s|\le V_0(n)} E[W_{ns}^4]}{\left(\sum_{|s|\le V_0(n)} E[W_{ns}^2]\right)^2} \to 0 \quad .$$

Since $\sum_{|s|\le V_0(n)} E[W_{ns}^2] = \sigma_0^2 + o(1)$, it suffices to show that $\sum_{|s|\le V_0(n)} E[W_{ns}^4] \to 0$ for a suitably chosen non-decreasing sequence with $n/V_0(n) \to 0$.

Since $E\,\|\varepsilon_t\|^4 < \infty$, we have from (25)

$$E[W_{ns}^4] \le Cn^2 \sum_{b=1}^{v} \left\{ \sum_{a=1}^{v_x} \left|n^{-d}\psi_\Delta(s)_{ab}\right|^4 + \sum_{a=v_x+1}^{v} \left|n^{-d_u}\psi_\Delta(s)_{ab}\right|^4 \right\}.$$

For $a \in \{1,\ldots,v_x\}$, we have from the Cauchy-Schwarz inequality

$$\max_s |n_\Delta^{-d}\psi_\Delta(s,n)_{ab}| \le \frac{1}{2\pi} n^{-d} \left(\int_{A/n}^{B/n} |\Psi_{ab}(x)|^2\,dx\right)^{1/2} \left(\int_{A/n}^{B/n} (1)\,dx\right)^{1/2}$$

$$\le Cn^{-d}\left[(B/n)^{-2d+1} + (A/n)^{-2d+1}\right]^{1/2} \left(\frac{B-A}{n}\right)^{1/2} = C\,n^{-1}\,.$$

Using similar arguments, we obtain overall that $\max_s E[W_{ns}^4] = O(1/n^2)$, and therefore that

$$\sum_{|s|\le V_0(n)} E[W_{ns}^4] \le \frac{CV_0(n)}{n^2} \quad .$$

The proof of the lemma is therefore completed by choosing $V_0(n)$ to be any non-decreasing sequence such that $n/V_0(n) \to 0$ and $V_0(n)/n^2 \to 0$, for example, $V_0(n) = [n^{1.5}]$. $\square$

Proof of Theorem 1

By the Cramer-Wold device, applied to complex-valued random variables, it suffices to show that any linear combination of the vm complex-valued random variables contained in $\{\Lambda_n w_j^T\}_{j=1}^m$, with fixed complex-valued coefficients, converges in distribution to the corresponding linear combination of the limit distribution given in the statement of Theorem 1. The initial linear combination can be expressed as

$$Y_n = \sum_{j=1}^{m} a_j^* \mathbf{\Lambda}_n w_j^T$$

where a_j are $v \times 1$ vectors of complex numbers. Using the definitions given here and in the preceding section, together with the change of variable $s = t - k$, we can write

$$\begin{aligned}
\sum_{j=1}^{m} a_j^* \left(\mathbf{\Lambda}_n w_j^T\right) &= \frac{1}{\sqrt{2\pi n c_p}} \sum_{j=1}^{m} a_j^* \mathbf{\Lambda}_n \sum_{t=1}^{n} h_t^{p-1} \sum_{k=-\infty}^{\infty} \psi_k \epsilon_{t-k} \exp(i\mu_j t) \\
&= \frac{1}{\sqrt{2\pi n c_p}} \sum_{j=1}^{m} a_j^* \mathbf{\Lambda}_n \sum_{t=1}^{n} h_t^{p-1} \sum_{k=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \mathbf{\Psi}(\lambda)\, d\lambda\, \epsilon_{t-k} \exp(i\mu_j t) \\
&= \frac{1}{\sqrt{2\pi n c_p}} \sum_{j=1}^{m} a_j^* \mathbf{\Lambda}_n \sum_{t=1}^{n} h_t^{p-1} \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(t-s)\lambda} \mathbf{\Psi}(\lambda)\, d\lambda\, \epsilon_s \exp(i\mu_j t) \\
&= \frac{1}{\sqrt{2\pi n c_p}} \sum_{s=-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-is\lambda} \sum_{j=1}^{m} a_j^* \sum_{t=1}^{n} h_t^{p-1} \exp(it(\lambda + \mu_j) \mathbf{\Lambda}_n \mathbf{\Psi}(\lambda)\, d\lambda \right\} \epsilon_s \ .
\end{aligned}$$

Using a similar argument to that given above and defining

$$h_n^*(\lambda) = \frac{1}{\sqrt{2\pi n c_p}} \sum_{j=1}^{m} a_j^* \sum_{t=1}^{n} h_t^{p-1} \exp(it(\lambda + \mu_j)) \quad ,$$

we conclude that

$$Y_n = \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-is\lambda} h_n^*(\lambda) \mathbf{\Lambda}_n \mathbf{\Psi}(\lambda)\, d\lambda\, \epsilon_s \quad . \tag{27}$$

Let A be a real number with $0 < A < n\pi$. We write $Y_n = Y_n^A + R_n$, where

$$Y_n^A = \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-is\lambda} h_n^*(\lambda) \mathbf{\Lambda}_n \mathbf{\Psi}(\lambda)\, d\lambda\, \epsilon_s \quad , \tag{28}$$

$$R_n = \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{[-\pi,\pi]\setminus[-A/n,A/n]} e^{-is\lambda} h_n^*(\lambda) \mathbf{\Lambda}_n \mathbf{\Psi}(\lambda)\, d\lambda\, \epsilon_s \quad . \tag{29}$$

By an argument similar to that given in the proof of Proposition 2 of Terrin and Hurvich (1994), it can be shown that $K_n^*(x) := n^{-1/2} h_n^*(x/n) \to K_0^*(x)$ and uniformly on $[-A, A]$, where

$$K_0^*(x) = \sum_{j=1}^{m} a_j^* \Delta_p(x + 2\pi j) \ .$$

Similarly, it can be shown that for $-p+1/2 < d_u < d < 1/2$,

$$\lim_{n\to\infty}\left[\int_{[-\pi,\pi]} \|h_n(x)\|^2 \operatorname{trace}\{\mathbf{\Lambda}_n d\mathbf{G}(x)\mathbf{\Lambda}_n\}\right] = \int_{\mathbb{R}} \|K_0(x)\|^2 \operatorname{trace}\{d\mathbf{G}_0(x)\} \ < \infty \tag{30}$$

and

$$\lim_{A\to\infty}\left[\int_{\mathbb{R}\setminus[-A,A]} \|K_n(x)\|^2 \operatorname{trace}\{\mathbf{\Lambda}_n d\mathbf{G}(x)\mathbf{\Lambda}_n\}\right] = 0 \quad , \tag{31}$$

uniformly for $n = 1, 2, \ldots$.

It follows from the properties given above that we can approximate K_n on $[-A, A]$ by step functions of form

$$g_A(x) = \sum_{\ell=-L}^{L} g_{\Delta_\ell} 1_{\Delta_\ell}(x) \ ,$$

where $\Delta_{-L}, \ldots, \Delta_L$ partitions $[-A, A]$ into equal subintervals, and $g_{\Delta_0} = 0$. Specifically, we have

Lemma 5 *There exist step functions $\{g_A\}_{A>0}$ as above, such that for any $\epsilon > 0$*

$$\int_{-A}^{A} \|K_n - g_A\|^2 \operatorname{trace}\{d\mathbf{G}_n(x)\} < \varepsilon \tag{32}$$

when $A > A(\epsilon)$ and $n \geq n(\epsilon)$, and such that

$$\int_{-A}^{A} \|K_0 - g_A\|^2 \operatorname{trace}\{d\mathbf{G}_0(x)\} < \varepsilon \tag{33}$$

when $A > A(\epsilon)$.

Define

$$\begin{aligned} I_n^A &= \sum_s \left\{\frac{1}{2\pi}\int_{-A/n}^{A/n} e^{-isx} n^{1/2} g_A^*(nx)\mathbf{\Lambda}_n \mathbf{\Psi}(x)dx\right\} \epsilon_s \\ &= n^{1/2}\sum_s\sum_\ell g_{\Delta_\ell}^* \frac{1}{2\pi}\int_{\Delta_\ell/n} e^{-isx}\mathbf{\Lambda}_n\mathbf{\Psi}(x)dx\,\epsilon_s \\ &= \sum_\ell g_{\Delta_\ell}^* Z_n(\Delta_\ell)\,. \end{aligned} \tag{34}$$

It follows from Lemma 1 that as $n \to \infty$ for fixed A,

$$I_n^A \xrightarrow{d} I_0^A := \sum_\ell g_{\Delta_\ell}^* Z_{G_0}(\Delta_\ell). \tag{35}$$

We will complete the proof of the theorem by showing that $Y_n \xrightarrow{d} Y$, where Y is a complex normal random variable given by

$$Y := \int_{-\infty}^{\infty} K_0^*(x)\, dZ_{\mathbf{G}_0}(x).$$

For a given A, we have shown that $I_n^A \xrightarrow{d} I_0^A$. If we can show that $I_0^A \xrightarrow{d} Y$ as $A \to \infty$ and that for all $\epsilon > 0$

$$\lim_A \limsup_n P[|Y_n - I_n^A| \geq \epsilon] = 0 \tag{36}$$

it will follow that $Y_n \xrightarrow{d} Y$ by Theorem 25.5 of Billingsley (1986). We prove Equation (36) in Lemma 6. It remains to show that $I_0^A \xrightarrow{d} Y$ as $A \to \infty$.

We have

$$I_0^A - Y = \int_{-\infty}^{\infty} (g_A^*(x) - K_0^*(x))\, dZ_{\mathbf{G}_0}(x).$$

By Equation (12) and Cauchy's inequality,

$$E|I_0^A - Y|^2 \leq 3 \int_{-\infty}^{\infty} \|g_A(x) - K_0(x)\|^2 \operatorname{trace}\{d\mathbf{G}_0(x)\}.$$

It follows from (33) and (30) that $E|I_0^A - Y|^2 \to 0$ and hence that $I_0^A \xrightarrow{d} Y$ as $A \to \infty$. $\square$

Lemma 6 *For every $\epsilon > 0$,*

$$\lim_A \limsup_n P\left(\left|Y_n - I_n^A\right| \geq \epsilon\right) = 0.$$

Proof of Lemma 6: Since $Y_n - I_n^A = Y_n^A - I_n^A + R_n$, it suffices to show that

$$\lim_A \limsup_n E\left|Y_n^A - I_n^A\right|^2 \longrightarrow 0 \tag{37}$$

and

$$\lim_A \limsup_n E\left|R_n\right|^2 \longrightarrow 0. \tag{38}$$

We start by proving (37). We have

$$Y_n^A - I_n^A = \sum_s \left\{ \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-isx}[h_n^*(x) - n^{1/2} g_A^*(nx)] \mathbf{\Lambda}_n \mathbf{\Psi}(x) dx \right\} \varepsilon_s \quad .$$

Thus, using C to denote a generic constant, we have

$$\begin{aligned}
&Var[Y_n^A - I_n^A] \\
&= \sum_s \iint_{-A/n}^{A/n} e^{-isx}\left[h_n^*(x) - n^{1/2} g_A^*(nx)\right] \mathbf{\Lambda}_n \mathbf{\Psi}(x)\, \mathbf{\Sigma} \mathbf{\Psi}^*(y) \mathbf{\Lambda}_n' e^{isy} \left[h_n(y) - n^{1/2} g_A(ny)\right] dxdy \\
&\leq C \sum_s \iint_{-A/n}^{A/n} e^{-isx}\left[h_n^*(x) - n^{1/2} g_A^*(nx)\right] e^{isy}[h_n(y) - n^{1/2} g_A(ny)] \operatorname{trace}\{\mathbf{\Lambda}_n \mathbf{\Psi}(x)\, \mathbf{\Sigma} \mathbf{\Psi}^*(y) \mathbf{\Lambda}_n'\}\, dxdy \\
&= C \int_{-A/n}^{A/n} \left\| h_n(x) - n^{1/2} g_A(nx)\right\|^2 \operatorname{trace}\{\mathbf{\Lambda}_n \mathbf{\Psi}(x)\, \mathbf{\Sigma} \mathbf{\Psi}^*(x) \mathbf{\Lambda}_n'\}\, dx \quad ,
\end{aligned}$$

by Parseval's equality. With a change of variables, we obtain

$$Var[Y_n^A - I_n^A] \leq C \int_{-A}^{A} \|K_n(x) - g_A(x)\|^2 \operatorname{trace}\{d\mathbf{G}_n(x)\}$$

Equation (37) now follows from Lemma 3, Equation (32). We next prove (38). Using an argument very similar to that given in proving (37), we conclude from (29) that

$$Var[R_n] \leq C \int_{[-n\pi,n\pi]\backslash[-A,A]} \|K_n(x)\|^2 \operatorname{trace}\{d\mathbf{G}_n(x)\}.$$

Thus, (38) follows from (31) $\square$

Proof of Lemma 2: Denote the $2m \times 1$ vector $\varphi(x) = \left(\{\upsilon_j(x)\}_{j=1}^m, \{\nu_k(x)\}_{k=1}^m\right)'$. Then

$$\Xi = \int_{\mathbb{R}} (\varphi(x)\varphi^*(x)) \otimes \mathbf{G}_{0,xx}(dx).$$

We will show that

$$a'\Xi a = \int_{\mathbb{R}} a' \left[(\varphi(x)\varphi^*(x)) \otimes \mathbf{G}_{0,xx}(dx)\right] a > 0,$$

where a is a real-valued $2v_x m \times 1$ constant vector. Since both $(\varphi(x)\varphi^*(x))$ and $\mathbf{G}_{0,xx}(dx)$ are nonnegative definite in $\mathbb{R}$, it suffices to show that

$$a' \left[(\varphi(x)\varphi^*(x)) \otimes \mathbf{G}_{0,xx}(dx)\right] a \neq 0 \tag{39}$$

for $x \in S$, where S is a Borel set with positive measure. Partitioning $a' = (a_1', \ldots, a_{2m}')$, where a_j are $v_x \times 1$ vectors, we write the lefthand side of (39)

$$\left(\sum_{j=1}^m [a_j \upsilon_j(x) + a_{m+j}\nu_j(x)]\right)^* \mathbf{G}_{0,xx}(dx) \left(\sum_{k=1}^m [a_k \upsilon_k(x) + a_{m+k}\nu_k(x)]\right).$$

Note that $\mathbf{G}_{0,xx}(dx)$ is positive definite everywhere in $\mathbb{R}\backslash\{0\}$ by (7). We show (39) by showing that $\{\upsilon_j(x)\}_{j=1}^m, \{\nu_k(x)\}_{k=1}^m$ are linearly independent. First we write

$$\upsilon_j(x) = \frac{(e^{ix}-1)}{2i\sqrt{2\pi}} \begin{pmatrix} 2p-2 \\ p-1 \end{pmatrix}^{-1/2} \sum_{k=0}^{p-1} \begin{pmatrix} p-1 \\ k \end{pmatrix} (-1)^k \left\{\phi_{j+k}(x) + \phi_{-(j+k)}(x)\right\}$$

and

$$\nu_j(x) = \frac{(e^{ix}-1)}{2\sqrt{2\pi}} \begin{pmatrix} 2p-2 \\ p-1 \end{pmatrix}^{-1/2} \sum_{k=0}^{p-1} \begin{pmatrix} p-1 \\ k \end{pmatrix} (-1)^k \left\{-\phi_{j+k}(x) + \phi_{-(j+k)}(x)\right\},$$

where

$$\phi_\ell(x) = \frac{1}{x + 2\pi\ell}, \quad \ell = \pm 1, \ldots, \pm(m+p-1).$$

It can be shown that $\{\phi_\ell(x)\}_\ell$ are linearly independent, since for $x \notin \{2\pi\ell : \ell = \pm 1, \ldots, \pm(m+p-1)\}$, $\sum_\ell b_\ell \phi_\ell(x) \neq 0$ if $\{b_\ell\}$ are not all zero. Let

$$\widetilde{\upsilon}_j(x) = (-i) \sum_{k=0}^{p-1} \begin{pmatrix} p-1 \\ k \end{pmatrix} (-1)^k \left\{\phi_{j+k}(x) + \phi_{-(j+k)}(x)\right\}$$

and

$$\widetilde{\nu}_j(x) = \sum_{k=0}^{p-1} \begin{pmatrix} p-1 \\ k \end{pmatrix} (-1)^k \left\{\phi_{j+k}(x) - \phi_{-(j+k)}(x)\right\}.$$

It suffices to show that $\{\widetilde{v}_j(x)\}_{j=1}^m, \{\widetilde{\nu}_k(x)\}_{k=1}^m$ are linearly independent. We will only show this for the case $m \geq p$. The case of $m < p$ can be handled similarly. Let $\{\alpha_j\}_{j=1}^m$ and $\{\beta_j\}_{j=1}^m$ be two sequences of complex constants where neither sequence is identically zero. Then

$$
\begin{aligned}
&\sum_{j=1}^m [\alpha_j \widetilde{v}_j(x) + \beta_j \widetilde{\nu}_j(x)] \\
&= \sum_{k=0}^{p-1} \binom{p-1}{k} (-1)^k \sum_{j=1}^m \{(-i\alpha_j + \beta_j)\, \phi_{j+k}(x) - (i\alpha_j + \beta_j)\, \phi_{-(j+k)}(x)\} \\
&= \sum_{j=1}^m \sum_{\ell=j}^{j+p-1} \binom{p-1}{\ell-j} (-1)^{\ell-j} \{(-i\alpha_j + \beta_j)\, \phi_\ell(x) - (i\alpha_j + \beta_j)\, \phi_{-\ell}(x)\} \\
&= \sum_{|\ell|=1}^{m+p-1} b_\ell \phi_\ell(x) \qquad (40)
\end{aligned}
$$

where

$$
\begin{aligned}
b_\ell &= \sum_{j=\max(1,\ell-p+1)}^{\min(m,\ell)} \binom{p-1}{\ell-j} (-1)^{\ell-j} (-i\alpha_j + \beta_j), \qquad 1 \leq \ell \leq (m+p-1), \\
&= \sum_{j=\max(1,\ell-p+1)}^{\min(m,\ell)} \binom{p-1}{\ell-j} (-1)^{\ell-j} (-i\alpha_j - \beta_j), \qquad -(m+p-1) \leq \ell \leq -1.
\end{aligned}
$$

We need to show that the $\{b_\ell\}$ are not all zero, and hence that (40) is not almost everywhere the zero function. We will prove this by contradiction. Note that

$$b_1 = -i\alpha_1 + \beta_1 \text{ and } \quad b_{-1} = -i\alpha_1 - \beta_1.$$

If $b_1 = b_{-1} = 0$ then $\alpha_1 = \beta_1 = 0$. Together with the assumption that $b_2 = b_{-2} = 0$, we have $\alpha_2 = \beta_2 = 0$. Continuing this induction, we have $\alpha_j = \beta_j = 0$ for all j. This contradicts our assumptions on $\{\alpha_j\}_{j=1}^m$ and $\{\beta_j\}_{j=1}^m$. $\square$

Proof of Lemma 3: It suffices to show for $s = 1, \ldots, q-r$ that $Prob\{r+s \in \mathcal{J}^C\} = O(n^{d_u - d})$. Suppose that $r+s \in \mathcal{J}^C$. Then for some $\zeta \in [0,1]$, we have $H(\zeta)\chi_{r+s}(\zeta) = 0$. It follows that $\lambda_1(\zeta) = \ldots = \lambda_{r+s}(\zeta) = 0$, so that $dim(Ker(H(\zeta)) \geq r+s$. Thus $\chi_1(\zeta), \ldots, \chi_{r+s}(\zeta)$ are orthonormal null vectors of $H(\zeta)$. We then have for $i \in \{1, \ldots, r+s\}$,

$$0 = H(\zeta)\chi_i(\zeta) = [H + \zeta\Delta H]\chi_i(\zeta) = H\chi_i(\zeta) + O_p(n^{d_u - d}) \quad ,$$

so that

$$||H\chi_i(\zeta)||^2 = O_p(n^{2(d_u - d)}) \quad , \quad i = 1, \ldots, r+s \quad , \qquad (41)$$

where $||\cdot||$ denotes the Euclidean norm.

Since $\{\chi_j(0)\}_{j=1}^q$ are an orthonormal basis for $\mathbb{C}^q$, we can write

$$\chi_i(\zeta) = \sum_{j=1}^q \alpha_{ij}\chi_j(0) \quad i = 1, \ldots, r+s \quad ,$$

with

$$\sum_{j=1}^{q} \alpha_{Lj}\alpha_{Mj} = \left\{ \begin{array}{cc} 1 & L = M \\ 0 & L \neq M \end{array} \right\} \quad .$$

From (41) we have for $i = 1, \ldots, r+1$,

$$\lambda_{r+1}^2(0) \sum_{j=r+1}^{q} \alpha_{ij}^2 \leq \sum_{j=r+1}^{q} \alpha_{ij}^2 \lambda_j^2(0) = || \sum_{j=r+1}^{q} \alpha_{ij}\lambda_j(0)\chi_j(0)||^2 = ||H\chi_i(\zeta)||^2 = O_p(n^{2(d_u-d)}) \quad . \tag{42}$$

Since $\{\chi_j(\zeta)\}_{j=1}^{q}$ are orthonormal, we have

$$\chi_L'(\zeta)\chi_M(\zeta) = 0 = \sum_{j=1}^{r} \alpha_{Lj}\alpha_{Mj} + \sum_{j=r+1}^{q} \alpha_{Lj}\alpha_{Mj} \quad , \quad L, M = 1, \ldots, r+1 \quad , \quad L \neq M \quad . \tag{43}$$

Define

$$\tilde{\alpha}_i = \begin{pmatrix} \alpha_{i1} \\ \vdots \\ \alpha_{ir} \end{pmatrix} \quad , \quad i = 1, \ldots, r+1 \quad ,$$

and define $B = [\tilde{\alpha}_1, \ldots, \tilde{\alpha}_r]$. We now show that $B'B = T_r + o_p(1)$, where T_r is an $r \times r$ identity matrix. Note first from (42) and (43) that if $L \neq M$,

$$\tilde{\alpha}_L'\tilde{\alpha}_M = -\sum_{j=r+1}^{q} \alpha_{Lj}\alpha_{Mj} \leq \sqrt{\sum_{j=r+1}^{q} \alpha_{Lj}^2} \sqrt{\sum_{j=r+1}^{q} \alpha_{Mj}^2} = O_p\left(\frac{1}{\lambda_{r+1}^2(0)} n^{2(d_u-d)}\right) \quad .$$

Since $\lambda_{r+1}^2(0)$ converges in distribution to a random variable which has no atom at zero, we have $\tilde{\alpha}_L'\tilde{\alpha}_M \xrightarrow{P} 0$ for $L \neq M$ and similarly $\tilde{\alpha}_L'\tilde{\alpha}_L \xrightarrow{P} 1$. It follows that $B'B = T_r + o_p(1)$.

Since $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{r+1}$ are not linearly independent, there exist coefficients $\beta_1, \ldots, \beta_r$ such that

$$\tilde{\alpha}_{r+1} = \beta_1\tilde{\alpha}_1 + \ldots + \beta_r\tilde{\alpha}_r \quad . \tag{44}$$

Since $(\beta_1, \ldots, \beta_r)' = B^{-1}\tilde{\alpha}_{r+1}$ we obtain

$$\sum_{M=1}^{r} \beta_M^2 = \tilde{\alpha}_{r+1}'(BB')^{-1}\tilde{\alpha}_{r+1} = 1 + o_p(1) \quad .$$

We now obtain an upper bound for $\sum_{j=1}^{r} \alpha_{r+1,j}^2$. From (43), (44) and the Cauchy-Schwarz inequality, we have

$$\sum_{j=1}^{r} \alpha_{r+1,j}^2 = -\sum_{M=1}^{r} \beta_M \sum_{j=r+1}^{q} \alpha_{r+1,j}\alpha_{Mj} \leq \sum_{M=1}^{r} |\beta_M| \sqrt{\sum_{j=r+1}^{q} \alpha_{r+1,j}^2} \sqrt{\sum_{j=r+1}^{q} \alpha_{Mj}^2}$$

$$\leq \sqrt{\sum_{j=r+1}^{q} \alpha_{r+1,j}^2} \sqrt{\sum_{M=1}^{r} \beta_M^2} \sqrt{\sum_{L=1}^{r} \sum_{j=r+1}^{q} \alpha_{Lj}^2} = O_p\left(\frac{n^{2(d_u-d)}}{\lambda_{r+1}^2(0)}\right) \quad . \tag{45}$$

From (42), we have

$$\lambda^2_{r+1}(0) - \lambda^2_{r+1}(0)\sum_{j=1}^{r}\alpha^2_{r+1,j} = O_p\left(n^{2(d_u-d)}\right) \quad .$$

Combining this with (45), we obtain

$$\lambda^2_{r+1}(0) = O_p\left(n^{2(d_u-d)}\right) \quad .$$

Thus, $Prob\{\mathcal{J}^C \neq \{1,\ldots,r\}\} \leq Prob\{\lambda^2_{r+1}(0) \leq R\}$, where R is a random variable which is $O_p(n^{2(d_u-d)})$. By Corollary 2, the limiting distribution of $H = An^{-2d}\sum Re(I_{xx,j})A'$ has rank $q-r$ with probability 1. We conclude that in fact $\lambda^2_{r+1}(0)$ converges in distribution to a random variable which has no mass at zero. Thus,

$$Prob\{\mathcal{J}^C \neq \{1,\ldots,r\}\} \leq Prob\{R/\lambda^2_{r+1}(0) \geq 1\} \to 0 \quad ,$$

since $R/\lambda^2_{r+1}(0)$ is $o_p(1)$ $\square$

References

[1] Barlow, J.L. and Slapničar, I. (2000) "Optimal perturbation bounds for the Hermitian eigenvalue problem," *Linear Algebra and its Applications* **309**, 19-43.

[2] Bierens, H.J. (1997) "Nonparametric cointegration analysis," *Journal of Econometrics* **77**, 379-404.

[3] Billingsley, P. (1986). *Probability and Measure*, 2'nd Ed., New York: Wiley.

[4] Chen, W.W. and Hurvich, C.M. (2002). "Estimating fractional cointegration in the presence of polynomial trends," Preprint.

[5] Engle, R.F., and Granger, C.W.J. (1987). "Co-integration and error correction: representation, estimation and testing," *Econometrica* **55**, 251-276.

[6] Fujikoshi, Y. (1985). "Two methods for estimation of dimensionality in canonical correlation analysis and the multivariate linear model." In: Matusita, K. (Ed.), Statistical Theory and Data Analysis. Elsevier, Amsterdam, pp. 233-240.

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

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

[9] Hurvich, C.M. and Chen, W.W. (2000) "An efficient taper for potentially overdifferenced long-memory time series," *J. Time Ser. Anal.* **21**, 155-180.

[10] Marinucci, D. and Robinson, P.M. (2002). "Semiparametric fractional cointegration analysis," *J. Econometrics* **105,** 225-247.

[11] Okamoto, M. (1973) "Distinctness of the eigenvalues of a quadratic form in a multivariate sample," *Ann. Stat.* **1**, 763-765.

[12] Robinson, P.M., (1994) "Semiparametric analysis of long-memory time series." *Ann. Stat.* **22**, 515-539.

[13] Robinson, P.M. and Marinucci, D. (1998) "Semiparametric frequency domain analysis of fractional cointegration," London School of Economics STICERD Discussion Paper, Number EM/98/348.

[14] Robinson, P.M. and Marinucci, D. (2001) "Narrow-band analysis of nonstationary processes," *Ann. Stat.* **29**, 947-86.

[15] Robinson, P.M. and Yajima, Y. (2002).

[16] Terrin, N. and Hurvich, C.M. (1994). An asymptotic Wiener-Ito representation for the low frequency ordinates of the periodogram of a long memory time series. *Stochastic Processes and their Applications* **54,** 297-307.

[17] Terrin, N. and Taqqu, M.S. (1991). Convergence in distribution of sums of bivariate Appell polynomials with long-range dependence. *Probability Theory and Related Fields* **90,** 57-81.

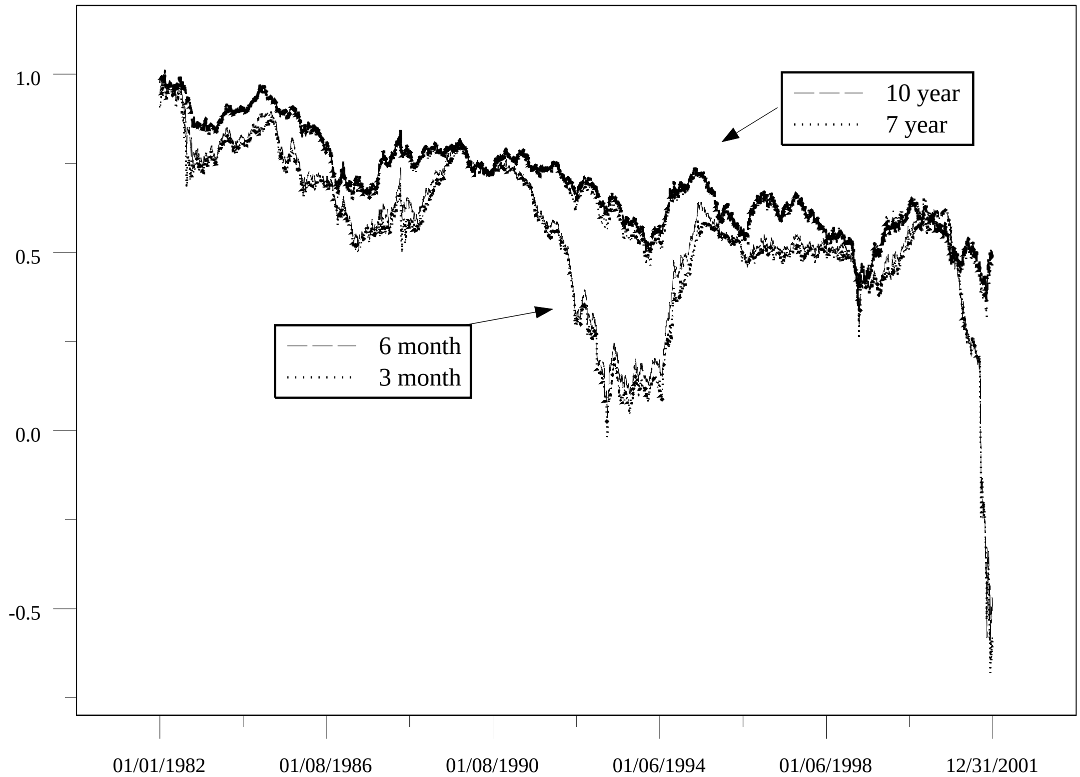

Figure 1: Log interest rates, maturities 3 months 6 months 7 years and 10 years. Daily data, 1/1/1982 to 12/31/2001

	χ_1	χ_2	χ_3	χ_4	χ_5	χ_6	χ_7	χ_8
3m	-0.0153	0.0088	0.3931	-0.0192	0.2733	0.5759	-0.4513	0.4846
6m	-0.0096	-0.0956	-0.7932	0.0587	-0.1807	-0.0122	-0.2971	0.4869
1y	0.0282	0.1764	0.4311	0.1664	-0.4245	-0.5806	-0.1202	0.4720
3y	-0.1525	-0.4270	0.0503	-0.6407	0.3301	-0.2985	0.2557	0.3434
5y	0.0863	0.7858	-0.1570	-0.0636	0.3735	-0.0217	0.3660	0.2693
7y	0.7024	-0.3594	0.0284	0.3860	0.1254	0.0713	0.3920	0.2306
10y	-0.6881	-0.1618	0.0206	0.5143	0.0447	0.1251	0.4205	0.2026
30y	0.0371	0.0674	0.0453	-0.3747	-0.6681	0.4699	0.4047	0.1436
λ_j	1.80×10^{-8}	3.50×10^{-8}	5.11×10^{-8}	1.25×10^{-7}	7.14×10^{-7}	3.94×10^{-6}	1.01×10^{-4}	8.47×10^{-4}

Table 1. Eigenvectors χ_j and corresponding eigenvalues λ_j of averaged periodogram for tapered, differenced log interest rate data, $m = 10$.

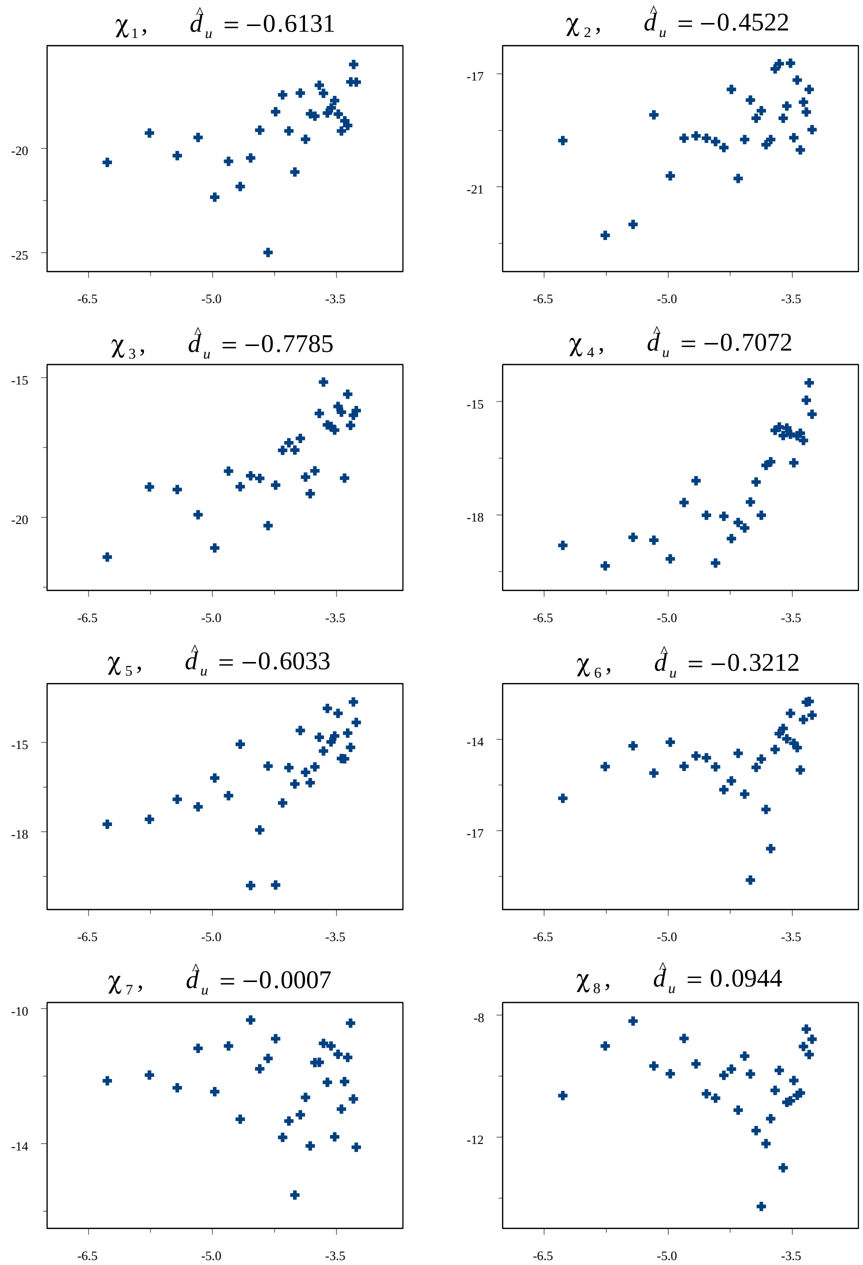

Figure 2: Log periodogram of tapered, differenced cointegrating residuals vs. log frequency. Each residual process is obtained by multiplying an eigenvector χ_j by the log interest rate data. Tapered Gaussian semiparametric estimated values of d_u with bandwidth 30 are also provided.