Estimating fractional cointegration in the presence of polynomial trends

Willa W. Chen[a], Clifford M. Hurvich[b],*

[a] *Department of Statistics, Texas A&M University, College Station, TX, 77843, USA*
[b] *New York University, Statistics, 44 W. 4'th Street, New York, NY 10012, USA*

Abstract

We propose and derive the asymptotic distribution of a tapered narrow-band least squares estimator (NBLSE) of the cointegration parameter β in the framework of fractional cointegration. This tapered estimator is invariant to deterministic polynomial trends. In particular, we allow for arbitrary linear time trends that often occur in practice. Our simulations show that, in the case of no deterministic trends, the estimator is superior to ordinary least squares (OLS) and the nontapered NBLSE proposed by P.M. Robinson when the levels have a unit root and the cointegrating relationship between the series is weak. In terms of rate of convergence, our estimator converges faster under certain circumstances, and never slower, than either OLS or the nontapered NBLSE. In a data analysis of interest rates, we find stronger evidence of cointegration if the tapered NBLSE is used for the cointegration parameter than if OLS is used.
© 2003 Elsevier B.V. All rights reserved.

1. Introduction

We say that a process is integrated of order d, denoted by $I(d)$, if its kth difference has spectral density

$$f(\lambda) \sim C|\lambda|^{-2(d-k)}, \quad \lambda \to 0,$$

where $C > 0$, and k is a nonnegative integer such that $d - k < 0.5$. We often call d the memory parameter. An $I(d)$ process without deterministic trends is weakly

---

* Corresponding author. Tel.: +1-212-998-0449; fax: +1-212-995-4003.
E-mail addresses: wchen@stat.tamu.edu (W.W. Chen), churvich@stern.nyu.edu (C.M. Hurvich).

stationary if $d < 0.5$ and nonstationary otherwise. Consider a two-dimensional process (X_t, Y_t) such that both variates are $I(d)$ processes. We say that $\{X_t\}$ and $\{Y_t\}$ are fractionally cointegrated if there exists a linear combination $U_t = Y_t - \beta X_t$ such that $\{U_t\}$ is $I(d_U)$, with $d_U < d$. Fractional cointegration is a generalization of standard cointegration, where d and d_U are 1 and 0, respectively. Though both fractional and standard cointegration were originally defined simultaneously in Engle and Granger (1987), standard cointegration has been studied far more extensively both in theory and applications. See, for example, Johansen (1988, 1991), Stock and Watson (1988), Phillips (1991). Nevertheless, criticism on standard cointegration has also emerged. The standard cointegration paradigm allows only integer values for the memory parameter, and tests for the existence of cointegration rely on unit root theory (see Dickey and Fuller, 1979, 1981; Phillips and Perron, 1988). Many economic and financial processes in practice possess fractional memory parameters, a property known as long memory. There has been mounting evidence on the existence of long-run relationships among long-memory processes. See, for example, Cheung and Lai (1993), Diebold et al. (1994), Baillie and Bollerslev (1994). The fractional cointegration framework then provides more information since it allows the memory parameter to take fractional values and $d - d_U$ to be any positive real number.

Much work done in fractional cointegration analysis has focused on the reduction of the memory parameter from $d \geqslant \frac{1}{2}$, to $d_U < \frac{1}{2}$. One reason for this is that cointegration is commonly thought of as a stationary relationship between nonstationary variables. But cases where $d < \frac{1}{2}$ are also of interest, particularly if one wishes to study fractional cointegration in volatility. A popular method for estimating the cointegration parameter β in standard cointegration analysis is the ordinary least squares (OLS) estimator. Robinson (1994) noted that for $0 < d < \frac{1}{2}$, the OLS estimator will in general be inconsistent in the presence of correlation between $\{X_t\}$, $\{U_t\}$, and he proposed a narrow-band least squares estimator (NBLSE) of β in the frequency domain, demonstrating its consistency in this stationary framework. We now briefly review this estimator, which was studied in the nonstationary case $d \geqslant \frac{1}{2}$ by Robinson and Marinucci (2001), under a different definition for an $I(d)$ process than we use here.

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

$$w_{\xi,j} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \xi_t e^{i\lambda_j t}$$

and the cross-periodogram of $\{\xi_t\}$ and $\{\zeta_t\}$ by

$$I_{\xi\zeta,j} = w_{\xi,j}\bar{w}_{\zeta,j},$$

where the bar indicates complex conjugation and $\lambda_j = 2\pi j/n$ is the jth Fourier frequency, $j = 1, \ldots, n$. Define the averaged periodogram

$$\hat{F}_{\xi\zeta}(m) = \frac{2\pi}{n} \sum_{j=1}^m Re\{I_{\xi\zeta,j}\}, \quad 1 \leqslant m < n/2,$$

where m is a bandwidth parameter. The narrow-band estimator $\hat{\beta}_m$ is defined as

$$\hat{\beta}_m = \hat{F}_{XY}(m)/\hat{F}_{XX}(m).$$

Using their definition of an $I(d)$ process, Robinson and Marinucci (2001) derived the limiting distribution of $\hat{\beta}_m$, assuming that $d \geqslant \frac{1}{2}$, and that $m \to \infty$, $m/n \to 0$. This distribution depends on the degree of nonstationarity of the levels and the strength of the cointegration between them. Therefore the theoretical properties of $\hat{\beta}_m$ with $m \to \infty$ were derived for five mutually exclusive subsets of $\{d \geqslant \frac{1}{2}\} \cap \{0 \leqslant d_U < d\}$.

Asymptotically, if $m/n + m^{-1} \to 0$, the narrow-band estimator $\hat{\beta}_m$ either converges faster than or is equivalent to OLS in all of the situations considered. In the case where m is held fixed as $n \to \infty$, Robinson and Marinucci (2001) did not derive any limiting distribution theory for $\hat{\beta}_m$, although they did give expressions for the limiting covariance of $w_{X,j}$ and $w_{U,k}$, with $d \geqslant 0.5$, and j, k fixed positive integers.

It is known that many economic series possess linear trends. In their original definition of cointegration, Engle and Granger (1987) excluded deterministic components, so that the cointegrating relationship in this two-dimensional process is a stochastic one. Much subsequent work on cointegration includes linear trend, that is

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix} t + \begin{pmatrix} \tilde{X}_t \\ \tilde{Y}_t \end{pmatrix},$$

where $(\tilde{X}_t, \tilde{Y}_t)$ is a bivariate process with memory parameter d and no deterministic component. In practice, (α_x, α_y) is often unknown. However, most of the literature on estimating the cointegrating parameter β requires that

$$(-\beta, 1)(\alpha_x, \alpha_y)' = 0. \tag{1}$$

This condition is imposed in the standard cointegration framework to ensure that the residual process $\{U_t\}$ is stationary. There is no compelling reason for such a condition in the fractional cointegration framework, since in fractional cointegration $\{U_t\}$ need not be stationary. Even in standard cointegration, recent literature has allowed for linear trends without the above restriction; see Lütkepohl and Saikkonen (2000). The philosophy we adopt here is that the concept of cointegration should entail a relationship solely between the stochastic components of the series, and should not be affected in any way by additive trends.

We note that the results of Robinson and Marinucci (2001) do not allow for the possibility of deterministic trends in the levels. In a different paper, Robinson and Marinucci (2000) studied the properties of averaged periodograms in the presence of deterministic trends, under certain conditions to be described more fully below.

In this paper, we propose a tapered narrow-band least squares estimator of the cointegration parameter β, based on a family of tapers introduced in Hurvich and Chen (2000). The new estimator is invariant with respect to deterministic polynomial trends in the series. In the proposed method, the data are differenced, say k times in the presence of potential kth order polynomial trends. The differenced data are then tapered, i.e., multiplied by a sequence of constants, which depend on k. The use of differencing followed by tapering for the purpose of detrending was originally advocated by Hart (1989) in the context of nonparametric regression with autocorrelated errors.

Though differencing is a widely used technique for detrending, it is typically considered highly undesirable in standard cointegration analysis. Note, for example, that

if (X_t, Y_t) exhibits standard cointegration with $\{X_t\}$, $\{U_t\}$ mutually uncorrelated, then the differenced process $\triangle U_t = \triangle Y_t - \beta \triangle X_t$ will be noninvertible, and the OLS estimator of β based on the nontapered differences is only $n^{1/2}$-consistent, while the OLS estimator based on the levels is n-consistent. Therefore, it is commonly believed that β should not be estimated from differences. For example, Hamilton (1994, p. 573) says that the levels of $\{X_t\}$ contain information that is useful for forecasting $\{Y_t\}$ beyond that contained in changes of $\{X_t\}$ alone. The message of this paper is that tapering of the differenced data can not only strongly mitigate any detrimental leakage effects due to noninvertibility but will also preserve information originally contained in the levels of the processes. In particular, the tapered NBLSE based on the differences achieves n-consistency for β in the case of standard cointegration.

Robinson and Marinucci (2000) demonstrated that in the presence of deterministic trends a consistent and asymptotically normal estimator of β can be obtained from averaged periodograms with $m \to \infty$ at any rate, without differencing or tapering. By contrast, we consider m to be fixed in the asymptotics, we do not require that deterministic trends be present nor do we require the knowledge as to which deterministic components dominate which stochastic ones. Furthermore, we do not require that the stochastic component of the levels be nonstationary, and we do not require constraints such as (1) on the deterministic components.

The remainder of this paper is arranged as follows. In the next section, the tapered NBLS estimator is presented. In Sections 3 and 4, the theoretical results are given. Section 5 gives a discussion, where it is pointed out that in terms of rate of convergence, the tapered NBLSE converges faster under certain circumstances, and never slower, than either OLS or the nontapered NBLSE. In Section 6, an application of the new estimator to the study of interest rates is provided. We also present simulations showing that, in the case of no deterministic trends, the tapered NBLSE is superior to OLS and the nontapered NBLSE when the levels have a unit root and the cointegrating relationship between the series is weak. All the technical proofs are detailed in the Appendices.

2. The tapered narrow-band least squares estimator

Suppose that the observed series (X_t, Y_t) consists of two $I(d)$ components with $d \in (-0.5, p - 0.5)$, where $p \geqslant 1$ is a fixed integer, and the series may have additive deterministic polynomial trends of order less than or equal to $(p - 1)$. The stochastic component of the process is nonstationary when $d \geqslant \frac{1}{2}$. A widely used technique for detrending and inducing stationarity is differencing. The $(p - 1)$th difference will convert the memory parameter to $d - p + 1$, and will completely remove a polynomial trend of the form described above, converting the trend into a constant. However, overdifferencing may arise as an unintended consequence of differencing, causing problems such as bias in parameter estimation. Tapering (see, e.g., Tukey, 1967) is a technique for reducing periodogram bias. A number of studies have suggested that the routine use of differencing followed by tapering may be helpful in many situations. See, for instance, Hurvich and Ray (1995), Deo and Hurvich (1998), Hurvich and Chen (2000),

Hurvich et al. (2002). With a class of tapers due to Kolmogorov (see Zhurbenko, 1979), Velasco (1999a, b), Lobato and Velasco (2000), obtained general consistency and asymptotic normality results for periodogram and log-periodogram semiparametric estimates of d in the potential presence of additive polynomial trends of arbitrary degree by tapering the observations without differencing. A family of tapers introduced by Hurvich and Chen (2000), used on differenced data, was proven to be more efficient for periodogram-based semiparametric estimation of d. The narrow-band tapered NBLS estimator introduced in this paper is an application of this family of tapers on the $(p-1)$th differenced data.

Throughout the paper, we suppose that observations on $\{X_t\}$ and $\{Y_t\}$ are available for $t = -p+2, \ldots, n$. Equivalently, the $(p-1)$th differences $\{x_t\}$ and $\{y_t\}$, $t = 1, \ldots, n$, are generated from weakly stationary processes $\{x_t\} = \{\triangle^{p-1} X_t\}$ and $\{y_t\} = \{\triangle^{p-1} Y_t\}$, with common memory parameter $d_x = d_y = d - p + 1 \in (-p + 0.5, 0.5)$, where $\varDelta$ is the differencing operator. We will give more detailed assumptions on the differenced process in the next section.

Note that if $\{X_t\}$ and $\{Y_t\}$ are fractionally cointegrated with cointegration parameter β, then $\{x_t\}$ and $\{y_t\}$ are also fractionally cointegrated with the same cointegration parameter. Furthermore the relation between the linear combination $U_t = Y_t - \beta X_t$ and the linear combination $u_t = y_t - \beta x_t$ is that $\{u_t\} = \{\triangle^{p-1} U_t\}$.

The complex-valued taper in Hurvich and Chen (2000) is given by

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

Define the tapered discrete Fourier transform of a series $\{\xi_t\}$ by

$$w_{\xi,j}^{\mathrm{T}} = \frac{1}{\sqrt{2\pi \sum |h_t^{p-1}|^2}} \sum_{t=1}^{n} h_t^{p-1} \xi_t e^{i\lambda_j t}$$

and the tapered cross-periodogram of $\{\xi_t\}$ and $\{\zeta_t\}$ by

$$I_{\xi\zeta,j}^{\mathrm{T}} = w_{\xi,j}^{\mathrm{T}} \bar{w}_{\zeta,j}^{\mathrm{T}}.$$

Since $w_{\xi,j}^{\mathrm{T}}$ can be written as a linear combination of $w_{\xi,j}, \ldots, w_{\xi,j+p-1}$ (see Hurvich and Chen, 2000), it follows that the tapered Fourier transform values at nonzero Fourier frequencies are invariant to the mean of the series. Thus, the tapered DFT, tapered periodogram and tapered cross-periodogram based on the $(p-1)$th differences $\{x_t\}$, $\{y_t\}$ are invariant to $(p-1)$th degree polynomial trends in the levels $\{X_t\}$, $\{Y_t\}$.

Next, define the averaged tapered periodogram

$$\hat{F}_{\xi\zeta}^{\mathrm{T}}(m) = \frac{2\pi}{n} \sum_{j=1}^{m} Re\{I_{\xi\zeta,j}^{\mathrm{T}}\}, \quad 1 \leqslant m < n/2.$$

The proposed tapered estimator of the cointegration parameter β is

$$\hat{\beta}_m^{\mathrm{T}} = \hat{F}_{xy}^{\mathrm{T}}(m)/\hat{F}_{xx}^{\mathrm{T}}(m), \tag{2}$$

where $m \geqslant 1$ is fixed. We can interpret $\hat{\beta}_m^{\mathrm{T}}$ as estimating the unknown β in the regression model

$$y_t = \beta x_t + u_t, \quad t = 1, 2, \ldots . \tag{3}$$

The mean-invariance of the tapered DFT allows us to ignore the intercept which would otherwise need to be included in (3). Eqs. (2) and (3) imply that

$$\hat{\beta}_m^{\mathrm{T}} - \beta = \hat{F}_{xu}^{\mathrm{T}}(m)/\hat{F}_{xx}^{\mathrm{T}}(m). \tag{4}$$

In the next two sections, we obtain the limiting joint distribution of the tapered discrete Fourier transforms $\{w_{x,j}^{\mathrm{T}}\}_{j=1}^{m}$ and $\{w_{u,j}^{\mathrm{T}}\}_{j=1}^{m}$ for fixed m, and obtain as a consequence the limiting distribution for $\hat{\beta}_m^{\mathrm{T}}$. Our main theorem is based on bivariate generalizations of some of the results of Terrin and Taqqu (1991). The proofs of the theoretical results are given in the Appendices.

3. Model and complex gaussian random measure

Let ψ_k be a sequence of real-valued 2×2 matrices such that

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

where for each $\lambda \in [-\pi, \pi]$, $\mathbf{\Psi}(\lambda)$ is a complex-valued matrix such that $\mathbf{\Psi}(-\lambda) = \mathbf{\Psi}^*(\lambda)$, the superscript $*$ denotes conjugate transposition, and ψ_0 is an identity matrix. Let $(x_t, u_t)'$ be a vector process,

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=-\infty}^{\infty} \psi_k \varepsilon_{t-k} = \sum_{k=-\infty}^{\infty} \begin{bmatrix} \psi_{k,11} & \psi_{k,12} \\ \psi_{k,21} & \psi_{k,22} \end{bmatrix} \begin{bmatrix} \varepsilon_{t-k,1} \\ \varepsilon_{t-k,2} \end{bmatrix},$$

where $\{\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2})'\} \sim iid(\mathbf{0}, 2\pi\mathbf{\Sigma})$,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}$$

is a symmetric positive definite matrix ($\sigma_{12} = \sigma_{21}$), and $\mathrm{E}[\varepsilon_{t,1}^4] < \infty$, $\mathrm{E}[\varepsilon_{t,2}^4] < \infty$. Let

$$y_t = \beta x_t + u_t. \tag{5}$$

The spectral density matrix of $(x_t, u_t)'$ is

$$\mathbf{f}(\lambda) = \mathbf{\Psi}(\lambda)\mathbf{\Sigma}\mathbf{\Psi}^*(\lambda)$$

so that

$$f_{11}(\lambda) = |\Psi_{11}(\lambda)|^2 \sigma_{11}^2 + \Psi_{12}(\lambda)\overline{\Psi_{11}(\lambda)}\sigma_{21} + \Psi_{11}(\lambda)\overline{\Psi_{12}(\lambda)}\sigma_{12} + |\Psi_{12}(\lambda)|^2 \sigma_{22}^2,$$

$$f_{12}(\lambda) = \Psi_{11}(\lambda)\overline{\Psi_{21}(\lambda)}\sigma_{11}^2 + \Psi_{11}(\lambda)\overline{\Psi_{22}(\lambda)}\sigma_{12} + \Psi_{12}(\lambda)\overline{\Psi_{21}(\lambda)}\sigma_{21}$$
$$+ \Psi_{12}(\lambda)\overline{\Psi_{22}(\lambda)}\sigma_{22}^2,$$

$$f_{21}(\lambda) = \overline{f_{12}(\lambda)}$$

and

$$f_{22}(\lambda) = |\Psi_{21}(\lambda)|^2 \sigma_{11}^2 + \Psi_{21}(\lambda)\overline{\Psi_{22}(\lambda)}\sigma_{12} + \Psi_{22}(\lambda)\overline{\Psi_{21}(\lambda)}\sigma_{21} + |\Psi_{22}(\lambda)|^2 \sigma_{22}^2.$$

We assume that $\{x_t\}$ and $\{u_t\}$ have memory parameters d_x and d_u respectively, with $-p + \frac{1}{2} < d_u < d_x < \frac{1}{2}$. We further assume that for $\lambda \in [-\pi, \pi]$,

$$\boldsymbol{\Psi}(\lambda) = \begin{pmatrix} (1 - \mathrm{e}^{-\mathrm{i}\lambda})^{-d_x}\tau_{11}(\lambda)\,\mathrm{e}^{\mathrm{i}\phi_{11}(\lambda)} & (1 - \mathrm{e}^{-\mathrm{i}\lambda})^{-d_{12}}\tau_{12}(\lambda)\,\mathrm{e}^{\mathrm{i}\phi_{12}(\lambda)} \\ (1 - \mathrm{e}^{-\mathrm{i}\lambda})^{-d_{21}}\tau_{21}(\lambda)\,\mathrm{e}^{\mathrm{i}\phi_{21}(\lambda)} & (1 - \mathrm{e}^{-\mathrm{i}\lambda})^{-d_u}\tau_{22}(\lambda)\,\mathrm{e}^{\mathrm{i}\phi_{22}(\lambda)} \end{pmatrix} \tag{6}$$

where $d_{12} \leqslant d_x$ and $d_{21} \leqslant d_u$, and for $(j,k) \in \{1,2\}$, $\tau_{jk}(\cdot)$ are positive even real-valued functions, $\phi_{jk}(\cdot)$ are odd real-valued functions, all continuously differentiable in an interval containing zero. This formulation is semiparametric, but includes the noncointegrated bivariate ARFIMA models as special cases; see Lobato (1997). It follows from (6) that we can write the spectral density matrix as

$$f(\lambda) = \Upsilon(\lambda)f^{\dagger}(\lambda)\Upsilon^*(\lambda) \tag{7}$$

where $\Upsilon(\lambda) = diag[(1 - \mathrm{e}^{-\mathrm{i}\lambda})^{-d_x}, (1 - \mathrm{e}^{-\mathrm{i}\lambda})^{-d_u}]$ and $f^{\dagger}(\lambda)$ is positive definite, Hermitian, continuous at zero frequency, and therefore real-valued at zero frequency. It also follows from (6) that as $\lambda \to 0^+$ the matrix of first derivatives of $\boldsymbol{\Psi}(\lambda)$ is given by

$$\boldsymbol{\Psi}'(\lambda) \sim (-1) \begin{pmatrix} \mathrm{e}^{-\mathrm{i}d_x\pi/2}\,d_x\tau_{11}(0)\,|\lambda|^{-d_x-1} & \mathrm{e}^{-\mathrm{i}d_{12}\pi/2}\,d_{12}\tau_{12}(0)\,|\lambda|^{-d_{12}-1} \\ \mathrm{e}^{-\mathrm{i}d_{21}\pi/2}\,d_{21}\tau_{21}(0)\,|\lambda|^{-d_{21}-1} & \mathrm{e}^{-\mathrm{i}d_u\pi/2}\,d_u\tau_{22}(0)\,|\lambda|^{-d_u-1} \end{pmatrix}, \tag{8}$$

and the behavior of $\boldsymbol{\Psi}'(\lambda)$ as $\lambda \to 0^-$ is obtained by noting that $\boldsymbol{\Psi}'(-\lambda) = \overline{\boldsymbol{\Psi}'(\lambda)}$ for $\lambda > 0$. It was implicitly assumed in Eq. (8) that d_x, d_{12}, d_{21} and d_u are all nonzero. If any of these quantities is zero, the corresponding (j,k) entry of the righthand side of Eq. (8) would be replaced by the constant $\tau_{jk}(0)\mathrm{i}\phi'_{jk}(0) + \tau'_{jk}(0)$.

Let G be the 2×2 matrix spectral measure on $[-\pi, \pi]$ defined by $G(\mathrm{d}\lambda) = f(\lambda)\,\mathrm{d}\lambda$. Let G_n be the 2×2 renormalized spectral measure on $\mathbb{R}$ defined by

$$G_n(\mathrm{d}x) = \Lambda_n G(\mathrm{d}x/n)\Lambda_n = \Lambda_n \boldsymbol{\Psi}(x/n)\boldsymbol{\Sigma}\boldsymbol{\Psi}^*(x/n)\Lambda_n\,\mathrm{d}x \tag{9}$$

where $\Lambda_n = diag(n^{-d_x}, n^{-d_u})$.

It follows from our assumptions that there exists a Hermitian positive definite 2×2 matrix-valued measure G_0 on $\mathbb{R}$ such that $G_n(S) \to G_0(S)$ as $n \to \infty$ for all bounded Lebesgue measurable sets S. For $x > 0$, we have

$$G_0(\mathrm{d}x) = \Pi(x)f^{\dagger}(0)\Pi^*(x)\,\mathrm{d}x \tag{10}$$

and $G_0(-\mathrm{d}x) = \overline{G_0(\mathrm{d}x)}$, where $\Pi(x) = diag(\mathrm{e}^{-\mathrm{i}d_x\pi/2}|x|^{-d_x}, \mathrm{e}^{-\mathrm{i}d_u\pi/2}|x|^{-d_u})$.

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

$$\varepsilon_t = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda t}\,\mathrm{d}Z_\varepsilon(\lambda)$$

where $\mathrm{d}Z_\varepsilon(\lambda)$ is a (2×1) complex-valued random vector with the following properties:

$$\overline{\mathrm{d}Z_\varepsilon(-\lambda)} = \mathrm{d}Z_\varepsilon(\lambda), \quad \mathrm{E}[\mathrm{d}Z_\varepsilon(\lambda)] = 0, \tag{11}$$

$$\mathrm{E}[\mathrm{d}Z_\varepsilon(\lambda)\,\mathrm{d}Z_\varepsilon^*(\mu)] = 0 \quad (\mu \neq \lambda), \quad \mathrm{E}[\mathrm{d}Z_\varepsilon(\lambda)\,\mathrm{d}Z_\varepsilon^*(\lambda)] = \boldsymbol{\Sigma}\,\mathrm{d}\lambda. \tag{12}$$

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

$$a_\Delta(s) = \frac{1}{2\pi} \int_{\Delta/n} e^{-isx} \Psi(x) \, dx, \tag{13}$$

$$Z_n(\Delta) = n^{1/2} \Lambda_n \sum_{s=-\infty}^{\infty} a_\Delta(s) \varepsilon_s = n^{1/2} \Lambda_n \int_{\Delta/n} \Psi(x) \, dZ_\varepsilon(x), \tag{14}$$

where the final equality in (14) follows from Theorem 4.10.1 of Brockwell and Davis (1991).

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*

$$(Z_n(\Delta_1), \ldots, Z_n(\Delta_M)) \xrightarrow{d} (Z_{G_0}(\Delta_1), \ldots, Z_{G_0}(\Delta_M)) \tag{15}$$

where for any Borel set S of $\mathbb{R}$, the 2×2 matrix measure $G_0(S)$ is defined above, and Z_{G_0} is the bivariate complex Gaussian random measure satisfying

$$E[Z_{G_0}(S)] = 0, \quad E[Z_{G_0}(S)Z_{G_0}^*(S)] = G_0(S), \quad \overline{Z_{G_0}(-S)} = Z_{G_0}(S),$$

$$E[Z_{G_0}(S_1)Z_{G_0}^*(S_2)] = 0 \quad \text{if } S_1 \cap S_2 = \emptyset. \tag{16}$$

4. Tapered DFTs and main theorem

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

$$w_j^T = \begin{bmatrix} w_{x,j}^T \\ w_{u,j}^T \end{bmatrix} = \frac{1}{\sqrt{2\pi \sum_{t=1}^n |h_t^{p-1}|^2}} \sum_{t=1}^n h_t^{p-1} \begin{bmatrix} x_t \\ u_t \end{bmatrix} \exp(i\lambda_j t),$$

$$j = 1, \ldots, m. \tag{17}$$

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

$$c_p = 2^{-2(p-1)} \binom{2p-2}{p-1}.$$

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

$$\Delta_p(x) = \binom{2p-2}{p-1}^{-1/2} \sum_{k=0}^{p-1} \binom{p-1}{k} (-1)^k \Delta(x + 2\pi k),$$

where

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

It can be shown that $\Delta_p(x)$ can also be written as

$$\Delta_p(x) = \frac{e^{ix} - 1}{i\sqrt{2\pi}} \binom{2p-2}{p-1}^{-1/2} \frac{(2\pi)^{p-1}(p-1)!}{\prod_{k=0}^{p-1}(x + 2\pi k)}. \tag{18}$$

Theorem 1. *Suppose that* $-p + \frac{1}{2} < d_u < d_x < \frac{1}{2}$, *and* m *is a fixed positive integer. Then*

$$\{\Lambda_n w_j^{\mathrm{T}}\}_{j=1}^m \xrightarrow{d} \left\{\int_{\mathbb{R}} \Delta_p(x + 2\pi j)\,\mathrm{d}Z_{G_0}(x)\right\}_{j=1}^m$$

as $n \to \infty$, *where* $\Lambda_n = diag(n^{-d_x}, n^{-d_u})$ *and* Z_{G_0} *is the bivariate complex Gaussian random measure satisfying the properties given in* (16).

Using Theorem 1 together with the continuous mapping theorem and Eqs. (4) and (16), we can obtain the limit distribution of our estimator $\hat{\beta}_m^{\mathrm{T}}$, where $m \geqslant 1$ is fixed. First, for all positive integers $j, k \in \{1, \ldots, m\}$, we define the nonrandom (2×2) matrices

$$L_1(j, k) = \int_{\mathbb{R}} \Delta_p(x + 2\pi j)\Delta_p(-x + 2\pi k)\,G_0(\mathrm{d}x), \tag{19}$$

$$L_2(j, k) = \int_{\mathbb{R}} \Delta_p(x + 2\pi j)\overline{\Delta_p(x + 2\pi k)}\,G_0(\mathrm{d}x). \tag{20}$$

Given d_x, d_u and $f^\dagger(0)$, each of the above matrices can be obtained by performing four one-dimensional numerical integrations, using Eq. (10) for $G_0(\mathrm{d}x)$. It follows from Eq. (18) that all entries of $L_1(j, k)$ and $L_2(j, k)$ are finite under the conditions of Theorem 1.

Corollary 1. *Under the conditions of Theorem* 1,

$$n^{d_x - d_u}(\hat{\beta}_m^{\mathrm{T}} - \beta) \xrightarrow{d} \frac{\sum_{j=1}^m (A_{x,j} A_{u,j} + B_{x,j} B_{u,j})}{\sum_{j=1}^m (A_{x,j}^2 + B_{x,j}^2)},$$

where $\{A_{x,j}, A_{u,j}, B_{x,j}, B_{u,j}\}_{j=1}^m$ *are jointly normal random variables with zero mean, and covariances determined by*

$$\mathrm{E}[A_j A_k'] = \tfrac{1}{2} Re[L_1(j, k) + L_2(j, k)], \quad \mathrm{E}[B_j B_k'] = \tfrac{1}{2} Re[L_2(j, k) - L_1(j, k)],$$

$$\mathrm{E}[A_j B_k'] = \tfrac{1}{2} Im[L_1(j, k) - L_2(j, k)],$$

with $A_j = (A_{x,j}, A_{u,j})'$, $B_j = (B_{x,j}, B_{u,j})'$.

As an example of the evaluation of some of the above quantities, we have

$$\mathrm{E}[A_{x,j}^2 + B_{x,j}^2] = Re[L_2(j, j)]_{11} = Re \int_{\mathbb{R}} |\Delta_p(x + 2\pi j)|^2 f^\dagger(0)_{11} |x|^{-2d_x}\,\mathrm{d}x.$$

In the case $p = 1$, this reduces to

$$\frac{2}{\pi} f^\dagger(0)_{11} \int_{\mathbb{R}} \frac{\sin^2(x/2)}{(x + 2\pi j)^2} |x|^{-2d_x}\,\mathrm{d}x.$$

It is interesting to note that this expression for the expectation of the limiting distribution of n^{-2d_x} times the periodogram of $\{x_t\}_{t=1}^n$ at frequency λ_j is equivalent to the expression given in Theorem 1 of Hurvich and Beltrao (1993) for the limiting expectation of a normalized periodogram of a univariate series, where the normalization is by the spectral density at λ_j.

In the case where $\{x_t\}$ and $\{u_t\}$ are independent of each other, the cross-spectrum would be identically zero, in violation of our assumption that $\tau_{12}(\cdot)$ and $\tau_{21}(\cdot)$ are positive. Nevertheless, if the off-diagonal entries of $\mathbf{\Psi}(\cdot)$ are taken to be zero in Eq. (6), only a slight modification of our proof of Theorem 1 would be necessary, and the result would be unchanged.

The limiting distribution of $n^{d_x - d_u}(\hat{\beta}_m^{\mathrm{T}} - \beta)$ has $m - 1$ finite moments. A brief demonstration of this result follows. By Hölder's inequality, the expectation of the Lth power of the limit random variable in Corollary 1 is bounded by a constant times $(\mathrm{E}[S_m^{-2mL/(2m-1)}])^{(2m-1)/(2m)}$ where $S_m = \sum_{j=1}^m (A_{x,j}^2 + B_{x,j}^2)$. The quantity S_m can be represented as a linear combination of $2m$ independent χ_1^2 random variables. The covariance matrix of $\{A_{x,j}, B_{x,j}\}_{j=1}^m$ can be shown to be positive definite by an argument similar to that given in Moulines and Soulier (1999, pp. 1437–1438). Thus, S_m is bounded below by a positive multiple of a χ_{2m}^2 random variable, and therefore $\mathrm{E}[S_m^{-2mL/(2m-1)}] < \infty$ for $L = 0, \ldots, m - 1$.

5. Discussion

From Corollary 1, in the case of standard cointegration, we see that our tapered estimator with $p = 2$ is in fact n-consistent, even though it is based on the differenced data. This refutes the widely held belief that estimates of β in standard cointegration should not be based on differences.

It is interesting to compare the rates of convergence of our tapered NBLSE with those attained by OLS and the nontapered NBLSE. Such comparisons should be undertaken with caution, however, since our asymptotic framework for the estimator and definition of the underlying process differ from those taken in Robinson and Marinucci (2001). Our results show that the tapered NBLSE ($\hat{\beta}_m^{\mathrm{T}}$) is n^{d-d_U}-consistent as long as $d \in (-\frac{1}{2}, p - \frac{1}{2})$ and $d_U < d$. The results of Robinson and Marinucci (2001) for OLS and the nontapered NBLSE ($\hat{\beta}_m$) do not require any upper bound on d, but do require that $d \geqslant \frac{1}{2}$ and that deterministic trends be absent. The upper bound on d for the tapered NBLSE is not restrictive, since in practice there is no upper limit on the value of p used by the practitioner. It should be noted that Robinson and Marinucci (2001) did not present a limiting distribution for $\hat{\beta}_m$ in the case $d_U = 0$, d > 1, so our comparisons do not include $\hat{\beta}_m$ for this case.

Overall, there is no situation where either $\hat{\beta}_m$ or OLS converges faster than $\hat{\beta}_m^{\mathrm{T}}$. On the other hand, $\hat{\beta}_m^{\mathrm{T}}$ converges faster than both $\hat{\beta}_m$ and OLS in two situations, viz. $\{d_U \geqslant 0, \ d \geqslant \frac{1}{2}, \ d_U + d < 1\}$, and $\{d_U > 0, \ d \geqslant \frac{1}{2}, \ d_U + d = 1\}$. In the first situation, $\hat{\beta}_m$ is $n^{d-d_U} m^{d_U + d - 1}$-consistent, which is slower than the rate for $\hat{\beta}_m^{\mathrm{T}}$ since the theory for $\hat{\beta}_m$ requires $m \to \infty$, while OLS converges at the rate n^{2d-1}, which is

also slower than the rate for $\hat{\beta}_m^{\mathrm{T}}$. In the second situation, $\hat{\beta}_m$ is $n^{2d-1}/\log m$ consistent and OLS is $n^{2d-1}/\log n$ consistent, so that $\hat{\beta}_m^{\mathrm{T}}$ converges faster by a logarithmic term.

Alternative approaches to our method would involve working with the raw data, rather than the differences. In this case besides the approach taken by Robinson and Marinucci (2000) which relies on the presence of trends, there are three different ways in which deterministic polynomial trends could be handled. One way would be to explicitly include regressors to account for these trends. Unfortunately, prior knowledge of existence of the trend would presumably be required here, since if a trend is assumed the asymptotic distribution of the estimator of the cointegrating parameter would be different according to whether the trend is present or absent. Another way of handling deterministic trends in an undifferenced series would be to use Kolmogorov–Zhurbenko tapers on a restricted grid of Fourier frequencies, as was done in Velasco (1999a, b), Lobato and Velasco (2000), and in Velasco and Robinson (2000). This approach with fixed m would presumably lead to an asymptotic distribution theory similar to the one we develop here, though it should be mentioned that it was shown in Hurvich and Chen (2000) that using Kolmogorov–Zhurbenko tapers on undifferenced data leads to asymptotically less efficient semiparametric estimators of the long memory parameter than use of the corresponding Hurvich–Chen taper on the differences. A third way to handle trends in the undifferenced data is to fit and remove these trends, and work with residuals. This approach may be problematic in practice, since the rate of convergence of the estimator of the trend coefficient depends on the memory parameter. (See, e.g., Deo and Hurvich, 1998.)

In the methodology proposed in this paper, it is left up to the practitioner to select the number of differences $p-1$, which in turn determines the particular taper to be used. Ideally, this choice would rely on a priori upper bounds on the order of the polynomial trend and $\lfloor 0.5+d \rfloor$. The number of differences $p-1$ needs to be at least as large as the maximum of these upper bounds. If, for example, it is assumed that $d < 1.5$ and that the deterministic trend is linear, then we can use any $p \geqslant 2$, that is, difference the data at least once. Under the same assumptions on d, if there is a quadratic deterministic trend, then at least two differences would be required, that is, $p \geqslant 3$. Identical considerations apply in the selection of the order p of the Kolmogorov–Zhurbenko tapers, although that methodology does not involve differencing the data.

Robinson and Marinucci (2001, p. 966) conjectured that n^{d-d_U}-consistency could be attained in estimating β assuming only that $d_U < d$, by holding m fixed in the NBLSE, thereby giving an improved rate of convergence in some situations. Our work verifies this conjecture, as long as the tapered NBLSE is used on differenced data. Robinson and Marinucci (2001, p. 966) went on to say that the use of a fixed value of m could lead to high dispersion in the limit distribution of the estimator, particularly in the case $m = 1$. Clearly the limit distribution of our $\hat{\beta}_m^{\mathrm{T}}$ changes discretely with m (held fixed in the asymptotics), so our method, particularly when m is small, has a sensitivity to the choice of m which is of a different character than that incurred when one takes $m \to \infty$. In the sequel, we will explore the performance of our estimator for various choices of m.

6. Analysis of interest rates

One classical example of standard cointegration is long and short term interest rates. According to the expectations hypothesis model (Campbell et al., 1997), the long-short yield spread is stationary. In this case, the cointegration parameter β is 1. In this section, we study cointegration among interest rates of eight maturities, ranging from three months to ten years. Our data are daily yields of US treasury securities from January 1, 1982 to December 31, 1999, weekends and holidays excluded, a total of $n = 4499$ observations for each series. Since yield spreads play an important role in the term structure of interest rates, each pair consisting of the 3 month interest rate and a longer maturity one is investigated. We first estimated d for each process, then estimated the cointegration parameter for each pair of series. Using these estimates of the cointegration parameter, we compute the cointegrating residuals, and finally estimate the d_U of the cointegrating residuals.

Most theoretical results on estimating d assume that the process is stationary and invertible, i.e., $d \in (-0.5, 0.5)$. Prior knowledge of stationarity of the process hence is essential in most existing methods of memory parameter estimation. However, there are conflicting opinions on the stationarity of interest rates. In the context of the theory of expectations hypothesis, interest rates are believed to be nonstationary processes with unit roots. On the other hand, mean reversion is a common assumption in modeling interest rate derivative securities. The tapered Gaussian semiparametric estimator (GSE) proposed by Hurvich and Chen (2000) is advantageous to our analysis since it does not require prior knowledge of stationarity of the process. The tapered GSE is a local Whittle estimator with the periodogram of the data replaced by the tapered periodogram of first differenced data, i.e.

$$\hat{d}^* = \arg \min_{d^* \in \Theta} R(d^*),$$

where

$$R(d^*) = \log G(d^*) - 2d^* \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j,$$

$$G(d^*) = \frac{1}{m} \sum_{j=1}^{m} (\log \lambda_j) I_{\xi\xi,j}^{\mathrm{T}},$$

and $m = o(n^{4/5})$. We can obtain $\hat{d}$ by $\hat{d} = \hat{d}^* + 1$. Table 1 gives the tapered GSE estimates of the memory parameters of the 3-month and 10-year interest rates.

All the estimates indicate that interest rates data are nonstationary. Therefore we can estimate d by applying the nontapered GSE (Künsch, 1987; Robinson, 1995) on the differenced data to gain better efficiency. We record the estimates in Table 2.

Tables 1 and 2 as well as the estimates of d for the other maturities are consistent with the conclusion that the interest rates are nonstationary with unit roots.

To compare our methodology with that of the standard cointegration framework, we first report the unit root test statistic (augmented Dickey–Fuller test with four

Table 1
Tapered GSE estimates of d (standard errors in parentheses)

Maturity	Bandwidth (m)			
	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$
3 months	1.2589	1.2123	1.1737	1.0147
	(0.1492)	(0.0870)	(0.0540)	(0.0341)
10 years	0.9914	1.0151	1.0789	1.0284
	(0.1492)	(0.0870)	(0.0540)	(0.0341)

Table 2
Nontapered GSE estimates of d (standard errors in parentheses)

Maturity	Bandwidth (m)			
	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$
3 months	1.1922	1.1448	1.1011	1.0003
	(0.1115)	(0.0672)	(0.0426)	(0.0273)
10 years	0.9914	1.0151	1.0789	1.0284
	(0.1115)	(0.0672)	(0.0426)	(0.0273)

Table 3
Augmented Dickey–Fuller test of cointegrating residuals (critical values: −2.86 at 5%, −3.12 at 2.5%)

Maturities	Statitics
3 month & 10 year	−1.9759
3 month & 7 year	−2.2546
3 month & 5 year	−2.3784
3 month & 3 year	−3.0511
3 month & 2 year	−3.5698
3 month & 1 year	−4.7890
3 month & 6 month	−7.0306

lags) of each cointegrating residual process computed by using OLS estimates of the cointegration parameter β. The results are given in Table 3.

Note that the statistic becomes more significant as the difference of the maturities decreases. We fail to reject the unit root null at both levels when the longer maturity is at least five years, reject at 5% level for the 3 month–3 year series and reject at both levels for the remaining pairs. To investigate the strength of the fractional cointegration among interest rates thoroughly, we apply both OLS and the tapered NBLS for β and three estimators for d_U the GSE, the GPH (Geweke and Porter-Hudak, 1983) and the FEXP (Moulines and Soulier, 1999; Hurvich and Brodsky, 2001). The GPH estimator

Table 4
Estimates of d_U of residuals on 3 month and 10 year interest rates (standard errors in parentheses)

		Bandwidth			
		$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$
$\hat{\beta}_{OLS}$ 0.646	GSE	1.030 (0.111)	0.971 (0.067)	1.032 (0.043)	0.992 (0.027)
	GPH	0.964 (0.143)	1.029 (0.086)	1.060 (0.055)	1.001 (0.035)
	FEXP		1.039 (0.042)		
$\hat{\beta}_{20}^{T}$ 0.416	GSE	0.954 (0.111)	0.947 (0.067)	1.034 (0.043)	1.033 (0.027)
	GPH	0.740 (0.143)	0.825 (0.086)	0.973 (0.055)	0.962 (0.035)
	FEXP		0.988 (0.042)		

is the least squares estimator of d in the regression model,

$$\log I_{\xi\xi,j} = C_1 - 2d \log \sin(\lambda_j/2) + \varepsilon_j, \quad j = 1, 2, \ldots, m,$$

where $m = o(n^{4/5})$. The FEXP (fractional exponential) estimator is a log-periodogram regression as well, except that a multiple regression is used,

$$\log I_{\xi\xi,j} = -2d \log \lambda_j + \sum_{k=0}^{h} g_k \cos(k\lambda_j) + \varepsilon_j, \quad j = 1, 2, \ldots, n,$$

where $h = O(\log n)$. There are two advantages of the FEXP estimator. An effective automatic model selection method of choosing h, local Mallows' C_L (Hurvich, 2001), is available. Also the selected model is amenable to goodness-of-fit testing since all the Fourier frequencies are used to fit a model to the periodogram. We report the estimates of β and d_U for the 3 month–10 year pair in Table 4.

The two estimates for β are quite different, and they are both far from one. Only two estimates of d_U show evidence of fractional cointegration: they are the GPH estimates with bandwidths $n^{0.4}$ and $n^{0.5}$ on the cointegration residuals using $\hat{\beta}_{20}^{T}$. We see that the choice of the estimator of β is important. To further investigate how well these estimators of β perform, we carried out a simulation of 5000 replications with sample size $n = 4500$, $d = 1$ and various values of d_U ranging from 0 to 1. Fig. 1 gives plots of the log MSE ratios of the nontapered NBLS versus the tapered NBLS with the same value of m and the OLS versus the tapered NBLS. The tapered NBLS is superior to the other two estimators in the case of $d = 1$ and d_U close to one and the superiority becomes more significant as the bandwidth increases.

The case of 3 month–10 year interest rates is a case where the ADF test does not reject the null of no standard cointegration ($d_U = 1$). Our findings agree with that of the ADF test. Table 5 gives the estimates of d_U from the residuals of 3 month–1 year pair, a case where the ADF rejects the null and in favor of the alternative of $d_U = 0$. The OLS
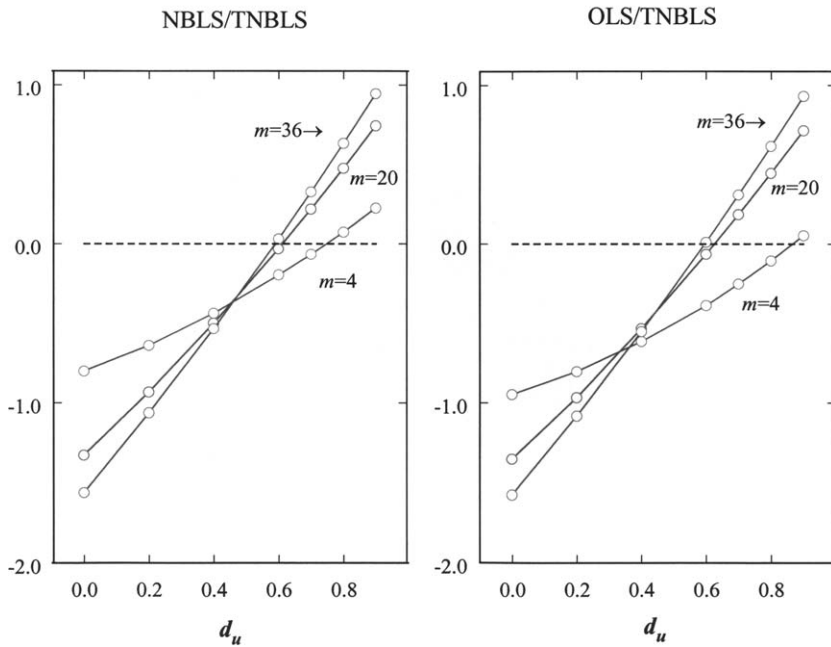
NBLS/TNBLS   OLS/TNBLS

Fig. 1. The log-ratios of MSE $d = 1$, $n = 4500$, 5000 replications.

Table 5
Estimates of d_U on residuals of 3 month and 1 year interest rates (standard errors in parentheses)

	Bandwidth			
	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$
GSE	0.580	0.811	0.927	0.882
	(0.111)	(0.067)	(0.043)	(0.027)
GPH	0.544	0.774	0.895	0.874
	(0.143)	(0.086)	(0.055)	(0.035)
FEXP		0.415 (0.143)		

estimator and the tapered NBLS of β are almost identical ($\hat{\beta} = 0.9670$, $\hat{\beta}_{20}^{\mathrm{T}} = 0.9681$), hence the estimates of d_U are also identical.

Since all the estimates of d_U are significantly less than one, there is strong evidence of fractional cointegration. However, since all the estimates are significantly larger than zero, there is no evidence of standard cointegration. Our conclusion contradicts that of the ADF test in this case. The FEXP estimator seems to show the strongest cointegration. In Fig. 2, we plot the log periodogram versus log Fourier frequency with estimated FEXP spectral density superimposed. We see how the FEXP estimator captures the local property of the spectral density. We also tested the adequacy of
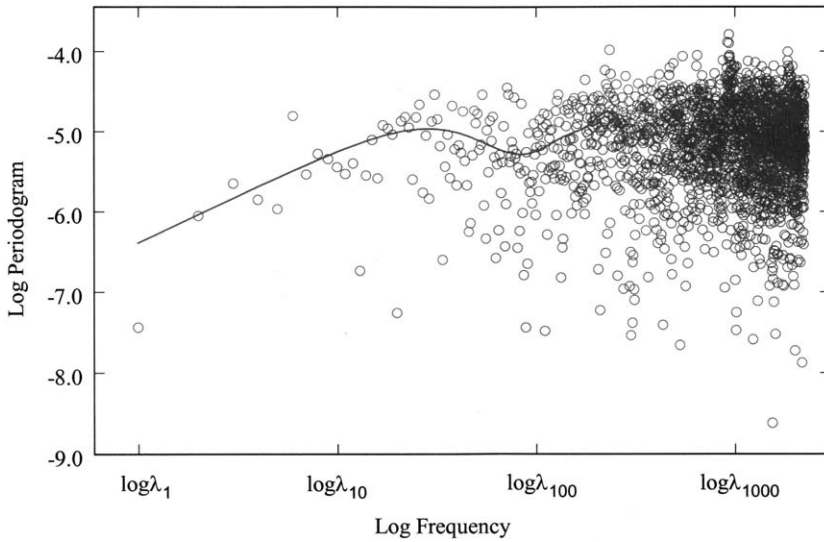
Fig. 2. Log periodogram of the differenced residuals vs. log frequency with FEXP spectrum superimposed ($\hat{d} = -0.5851 + 1$) (3 month and 1 year interest rates).

the FEXP model with a frequency domain generalized portmanteau goodness-of-fit test proposed recently by Chen and Deo (2000). The test statistic T_n is defined as,

$$T_n = \frac{\frac{2\pi}{n} \sum_{j=0}^{n-1} \widetilde{f_e}^2(\hat{\theta}, \lambda_j)}{\{\frac{2\pi}{n} \sum_{j=0}^{n-1} \widetilde{f_e}(\hat{\theta}, \lambda_j)\}^2},$$

where

$$\widetilde{f_e}(\hat{\theta}, \lambda) = \frac{2\pi}{n} \sum_{h=1}^{n-1} W(\lambda - \lambda_h) \frac{I(\lambda_h)}{f_{\hat{\theta}}(\lambda_h)}.$$

The spectral window is given by

$$W(\lambda) = \frac{1}{2\pi} \sum_{\ell=-(n-1)}^{(n-1)} k(\ell/p_n) e^{-i\ell\lambda},$$

where the kernel function $k : R \to [-1, 1]$ satisfies $k(-z) = k(z)$ and $k(0) = 1$, decays smoothly as ℓ increases, the bandwidth $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$. Since the test statistic T_n does not require computation of the residuals from the fitted model, it is easy to use for any long memory model, such as the FEXP. In Chen and Deo (2000), the limiting distribution of T_n was derived for both short and long memory models. This distribution is normal with mean and variance depending on the kernel function k. For the case of 3 month and 1 year interest rates, we tested the goodness of fit of the FEXP model fitted to the cointegration residuals using T_n with the Bartlett

Table 6
Goodness-of-fit test for FEXP model for cointegration residuals of 3 month and 1 year interest rates

Kernel	Bandwidth (p_n)		
	80	100	120
Tukey	1.1553	1.2391	1.3704
Bartlett	1.2200	1.3493	1.5522

Table 7
Evidence of fractional cointegration of 3 month and longer maturity interest rates based on the residuals from the $\hat{\beta}_{\text{OLS}}$ estimates

	GSE				GPH				FEXP	ADF
	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$		
10 yr										
7 yr										
5 yr										
3 yr	*				*				*	*
2 yr	*				*	*			*	*
1 yr	*	*	*	*	*	*	*	*	*	*
6 mo	*	*	*	*	*	*	*	*	*	*

Table 8
Evidence of fractional cointegration of 3 month and longer maturity interest rates based on the residuals from the $\hat{\beta}_{20}^{\text{T}}$ estimates

	GSE				GPH				FEXP	ADF
	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$		
10 yr					*	*				
7 yr					*	*			*	
5 yr	*				*				*	
3 yr	*				*	*			*	*
2 yr	*				*	*			*	*
1 yr	*	*	*	*	*	*	*	*	*	*
6 mo	*	*	*	*	*	*	*	*	*	*

kernel ($k(z) = 1 - |z|$, $z \leqslant 1$) and the Tukey kernel ($k(z) = \frac{1}{2}(1 - \cos(\pi z)$, $z \leqslant 1$). We report the test statistics in Table 6.

Comparing these statistics with the critical values from standard normal distribution, we accept the null that the model is adequate. We summarize our findings Tables 7 and 8. The symbol * indicates that the estimator of d_U is significantly less than 1 at 5% level.

From Tables 7 and 8, we conclude that (a) The strength of cointegration decreases as the longer maturity increases. (b) There is more evidence of cointegration when the tapered NBLS estimator is used for the cointegration parameter than when the OLS estimator is used. (c) For all the evidence of cointegration, none of the tests suggests that the cointegrating residual process has $d_U = 0$. Therefore, we found no evidence of standard cointegration in interest rates. (d) The goodness-of-fit test T_n suggests that FEXP model fits well. The FEXP model seems to show the most evidence of cointegration.

Acknowledgements

Appendix A. Proof of Lemma 1

From (14) and the properties (11), (12) of the spectral representation, we conclude that $Z_n(\Delta_j) = \overline{Z}_n(-\Delta_j)$ for $j = 1, \ldots, M$ and

$$E[Re\, Z_n(\Delta_j)\, Re\, Z'_n(\Delta_k)] = E[Re\, Z_n(\Delta_j)\, Im\, Z'_n(\Delta_k)] = E[Im\, Z_n(\Delta_j)\, Re\, Z'_n(\Delta_k)]$$

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

It therefore suffices to prove that $Z_n(\Delta) \xrightarrow{d} Z_{G_0}(\Delta)$ 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 $\alpha, \beta \in \mathbb{R}^2$, $\alpha'\, Re\, Z_n(\Delta) + \beta'\, Im\, Z_n(\Delta) \xrightarrow{d} \alpha'\, Re\, Z_{G_0}(\Delta) + \beta'\, Im\, Z_{G_0}(\Delta)$. Note that $E[Z_n(\Delta) Z_n^*(\Delta)] = G_n(\Delta) \to G_0(\Delta)$ 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

$$Var[\alpha'\, Re\, Z_n(\Delta) + \beta'\, Im\, Z_n(\Delta)]$$

$$= [\alpha'/2 + \beta'/(2i)]G_n(\Delta)[\alpha/2 - \beta/(2i)]$$

$$\quad + [\alpha'/2 - \beta'/(2i)]\overline{G_n(\Delta)}[\alpha/2 + \beta/(2i)]$$

$$\to [\alpha'/2 + \beta'/(2i)]G_0(\Delta)[\alpha/2 - \beta/(2i)]$$

$$\quad + [\alpha'/2 - \beta'/(2i)]\overline{G_0(\Delta)}[\alpha/2 + \beta/(2i)]$$

$$= Var[\alpha'\, Re\, Z_{G_0}(\Delta) + \beta'\, Im\, Z_{G_0}(\Delta)] := \sigma_0^2 > 0$$

since $G_0(\Delta)$ and $\overline{G_0(\Delta)}$ are both Hermitian and positive definite.

We will require bounds on the entries of $a_\Delta(s,n)$, 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 start with the $(1,1)$ entry, $a_\Delta(s,n)_{11}$. Using integration by parts, we have

$$a_\Delta(s,n)_{11} = \frac{1}{2\pi} \int_{A/n}^{B/n} e^{-isx} \mathbf{\Psi}_{11}(x)\, dx = \frac{1}{2\pi} \frac{1}{-is} e^{-isx} \mathbf{\Psi}_{11}(x)\Big|_{A/n}^{B/n}$$

$$- \frac{1}{2\pi} \frac{1}{-is} \int_{A/n}^{B/n} e^{-isx} \mathbf{\Psi}'_{11}(x)\, dx.$$

In the sequel, we use C to denote a generic constant. From Eqs. (6) and (8) we have, for all sufficiently small $x > 0$, $|\mathbf{\Psi}_{11}(x)| < Cx^{-d_x}$, and $|\mathbf{\Psi}'_{11}(x)| < Cx^{-d_x-1}$. We obtain for all $s \neq 0$

$$|a_\Delta(s,n)_{11}| \leqslant \frac{C}{|s|} [(A/n)^{-d_x} + (B/n)^{-d_x}] + \frac{C}{|s|} \frac{B-A}{n} [(A/n)^{-d_x-1}$$

$$+ (B/n)^{-d_x-1}] \leqslant Cn^{d_x} \frac{1}{|s|}. \tag{A.1}$$

Similarly, we have

$$|a_\Delta(s,n)_{12}| \leqslant Cn^{d_{12}}/|s|, \qquad |a_\Delta(s,n)_{21}| \leqslant Cn^{d_{21}}/|s|,$$

$$|a_\Delta(s,n)_{22}| \leqslant Cn^{d_u}/|s|. \tag{A.2}$$

We can write

$$\alpha' \, Re\, Z_n(\Delta) + \beta' \, Im\, Z_n(\Delta)$$

$$= \alpha' n^{1/2} \Lambda_n \sum_{s=-\infty}^{\infty} Re\, a_\Delta(s)\varepsilon_s + \beta' n^{1/2} \Lambda_n \sum_{s=-\infty}^{\infty} Im\, a_\Delta(s)\varepsilon_s$$

$$= n^{1/2-d_x} \sum_{s=-\infty}^{\infty} \{\alpha_1[Re\, a_\Delta(s)_{11}\varepsilon_{s,1} + Re\, a_\Delta(s)_{12}\varepsilon_{s,2}]$$

$$+ \beta_1[Im\, a_\Delta(s)_{11}\varepsilon_{s,1} + Im\, a_\Delta(s)_{12}\varepsilon_{s,2}]\}$$

$$+ n^{1/2-d_u} \sum_{s=-\infty}^{\infty} \{\alpha_2[Re\, a_\Delta(s)_{21}\varepsilon_{s,1} + Re\, a_\Delta(s)_{22}\varepsilon_{s,2}]$$

$$+ \beta_2[Im\, a_\Delta(s)_{21}\varepsilon_{s,1} + Im\, a_\Delta(s)_{22}\varepsilon_{s,2}]\}$$

$$= \sum_{s=-\infty}^{\infty} W_{ns}$$

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

$$
\begin{aligned}
W_{ns} = &\{n^{1/2-d_x}[\alpha_1 \, Re \, a_\Delta(s)_{11} + \beta_1 \, Im \, a_\Delta(s)_{11}] \\
&+ n^{1/2-d_u}[\alpha_2 \, Re \, a_\Delta(s)_{21} + \beta_2 \, Im \, a_\Delta(s)_{21}]\} \, \varepsilon_{s,1} \\
&+ \{n^{1/2-d_x}[\alpha_1 \, Re \, a_\Delta(s)_{12} + \beta_1 \, Im \, a_\Delta(s)_{12}] \\
&+ n^{1/2-d_u}[\alpha_2 \, Re \, a_\Delta(s)_{22} + \beta_2 \, Im \, a_\Delta(s)_{22}]\} \, \varepsilon_{s,2}.
\end{aligned}
\tag{A.3}
$$

Since $|Re \, a_\Delta(s)_{jk}| \leqslant |a_\Delta(s)_{jk}|$ and $|Im \, a_\Delta(s)_{jk}| \leqslant |a_\Delta(s)_{jk}|$ for $j,k \in \{1,2\}$, we conclude that for $s \neq 0$

$$
\begin{aligned}
E[W_{ns}^2] \leqslant &\, C\{n^{1-2d_x}[|a_\Delta(s)_{11}|^2 + |a_\Delta(s)_{12}|^2] + n^{1-2d_u}[|a_\Delta(s)_{21}|^2 \\
&+ |a_\Delta(s)_{22}|^2]\} \leqslant Cn/s^2,
\end{aligned}
\tag{A.4}
$$

where the final inequality follows from the bounds (A.1), (A.2) for the entries of $a_\Delta(s,n)$.

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

$$
\sum_{s=-\infty}^{\infty} W_{ns} = \sum_{|s| \leqslant V_0(n)} W_{ns} + \sum_{|s| > V_0(n)} W_{ns}.
$$

Using (A.4), we have

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

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

$$
\sum_{|s| \leqslant V_0(n)} W_{ns} \xrightarrow{d} N(0, \sigma_0^2).
$$

By the Lyapounov condition (see, e.g., Billingsley, 1986, p. 371) it suffices to show that

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

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

Since $E[\varepsilon_{s,1}^4] < \infty$ and $E[\varepsilon_{s,2}^4] < \infty$, we have from (A.3)

$$
\begin{aligned}
E[W_{ns}^4] \leqslant &\, C[|n^{1/2-d_x} a_\Delta(s,n)_{11}|^4 + |n^{1/2-d_x} a_\Delta(s,n)_{12}|^4 + |n^{1/2-d_u} a_\Delta(s,n)_{21}|^4 \\
&+ |n^{1/2-d_u} a_\Delta(s,n)_{22}|^4].
\end{aligned}
$$

Regarding the first term, we have from the Cauchy–Schwarz inequality

$$\max_s |n^{1/2-d_x} a_\Delta(s,n)_{11}| \leqslant \frac{1}{2\pi} n^{1/2-d_x} \left(\int_{A/n}^{B/n} |\mathbf{\Psi}_{11}(x)|^2 \, \mathrm{d}x \right)^{1/2} \left(\int_{A/n}^{B/n} (1) \, \mathrm{d}x \right)^{1/2}$$

$$\leqslant C[n^{1-2d_x}(B/n)^{-2d_x+1} + n^{1-2d_x}(A/n)^{-2d_x+1}]^{1/2}$$

$$\times \left(\frac{B-A}{n} \right)^{1/2} = Cn^{-1/2}.$$

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

$$\sum_{|s| \leqslant V_0(n)} \mathrm{E}[W_{ns}^4] \leqslant CV_0(n)/n^2.$$

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

Appendix B. 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 $2m$ complex-valued random variables contained in $\{\Lambda_n w_j^{\mathrm{T}}\}_{j=1}^m$, with fixed complex-valued coefficients, converges in distribution to the corresponding linear combination of the limit distribution given in Theorem 1. The initial linear combination can be expressed as

$$Y_n = n^{-d_x} \sum_{j=1}^m a_j w_{x,j}^{\mathrm{T}} + n^{-d_u} \sum_{j=1}^m b_j w_{u,j}^{\mathrm{T}} = (1,0)\Lambda_n a' \begin{pmatrix} w_1^{\mathrm{T}} \\ \vdots \\ w_m^{\mathrm{T}} \end{pmatrix} + (0,1)\Lambda_n b' \begin{pmatrix} w_1^{\mathrm{T}} \\ \vdots \\ w_m^{\mathrm{T}} \end{pmatrix}$$

where $a = (a_1,\ldots,a_m)'$ and $b = (b_1,\ldots,b_m)'$ are 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

$$n^{-d_x} \sum_{j=1}^m a_j w_{x,j}^{\mathrm{T}}$$

$$= (1,0) \frac{\Lambda_n}{\sqrt{2\pi n c_p}} \sum_{j=1}^m a_j \sum_{t=1}^n h_t^{p-1} \sum_{k=-\infty}^\infty \psi_k \varepsilon_{t-k} \exp(\mathrm{i}\lambda_j t)$$

$$= (1,0) \frac{\Lambda_n}{\sqrt{2\pi n c_p}} \sum_{j=1}^m a_j \sum_{t=1}^n h_t^{p-1} \sum_{k=-\infty}^\infty \frac{1}{2\pi} \int_{-\pi}^\pi \mathrm{e}^{\mathrm{i}k\lambda} \mathbf{\Psi}(\lambda) \, \mathrm{d}\lambda \varepsilon_{t-k} \exp(\mathrm{i}\lambda_j t)$$

$$= (1,0) \frac{\Lambda_n}{\sqrt{2\pi n c_p}} \sum_{j=1}^{m} a_j \sum_{t=1}^{n} h_t^{p-1} \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(t-s)\lambda} \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s \exp(i\lambda_j t)$$

$$= (1,0) \frac{\Lambda_n}{\sqrt{2\pi n c_p}} \sum_{s=-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-is\lambda} \sum_{j=1}^{m} a_j \sum_{t=1}^{n} h_t^{p-1} \exp(it(\lambda + \lambda_j) \mathbf{\Psi}(\lambda) d\lambda \right\} \varepsilon_s.$$

Using a similar argument to that given above and defining

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

$$h_{2n}(\lambda) = \frac{1}{\sqrt{2\pi n c_p}} \sum_{j=1}^{m} b_j \sum_{t=1}^{n} h_t^{p-1} \exp(it(\lambda + \lambda_j)),$$

we conclude that

$$Y_n = (1,0)\Lambda_n \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-is\lambda} h_{1n}(\lambda) \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s + (0,1)\Lambda_n$$

$$\times \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-is\lambda} h_{2n}(\lambda) \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s. \tag{B.1}$$

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

$$Y_n^A = (1,0)\Lambda_n \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-is\lambda} h_{1n}(\lambda) \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s + (0,1)\Lambda_n$$

$$\times \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-is\lambda} h_{2n}(\lambda) \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s, \tag{B.2}$$

$$R_n = (1,0)\Lambda_n \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{[-\pi,\pi]\setminus[-A/n,A/n]} e^{-is\lambda} h_{1n}(\lambda) \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s$$

$$+ (0,1)\Lambda_n \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{[-\pi,\pi]\setminus[-A/n,A/n]} e^{-is\lambda} h_{2n}(\lambda) \mathbf{\Psi}(\lambda) \, d\lambda \varepsilon_s. \tag{B.3}$$

By an argument similar to that given in the proof of Proposition 2 of Terrin and Hurvich (1994), it can be shown that $K_{1n}(x) := n^{-1/2} h_{1n}(x/n) \to K_{1,0}(x)$ and $K_{2n}(x) := n^{-1/2} h_{2n}(x/n) \to K_{2,0}(x)$ uniformly on $[-A, A]$, where

$$K_{1,0}(x) = \sum_{j=1}^{m} a_j \Lambda_p(x + 2\pi j), \quad K_{2,0}(x) = \sum_{j=1}^{m} b_j \Lambda_p(x + 2\pi j).$$

Similarly, it can be shown that for $-p + \frac{1}{2} < d_u < d_x < \frac{1}{2}$,

$$\lim_{n \to \infty} \left[n^{-2d_x} \int_{[-\pi,\pi]} |h_{1n}(x)|^2 (1,0) \, dG(x)(1,0)' \right.$$

$$+ n^{-2d_u} \int_{[-\pi,\pi]} |h_{2n}(x)|^2 (0,1) \, dG(x)(0,1)' \Bigg]$$

$$= \int_{\mathbb{R}} |K_{1,0}(x)|^2 (1,0) \, dG_0(x)(1,0)'$$

$$+ \int_{\mathbb{R}} |K_{2,0}(x)|^2 (0,1) \, dG_0(x)(0,1)' < \infty, \tag{B.4}$$

and

$$\lim_{A \to \infty} \left[\int_{\mathbb{R} \setminus [-A,A]} |K_{1n}(x)|^2 (1,0) \, dG_n(x)(1,0)' \right.$$

$$+ \int_{\mathbb{R} \setminus [-A,A]} |K_{2n}(x)|^2 (0,1) \, dG_n(x)(0,1)' \Bigg] = 0, \tag{B.5}$$

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

It follows from the properties given above that we can approximate K_{1n} and K_{2n} on $[-A, A]$ by step functions of form

$$g_{1,A}(x) = \sum_{\ell=-L}^{L} g_{1,\Delta_\ell} 1_{\Delta_\ell}(x), \qquad g_{2,A}(x) = \sum_{\ell=-L}^{L} g_{2,\Delta_\ell} 1_{\Delta_\ell}(x),$$

where $\Delta_{-L}, \ldots, \Delta_L$ partitions $[-A, A]$ into equal subintervals, and $g_{1,\Delta_0} = g_{2,\Delta_0} = 0$. Specifically, we have

Lemma 2. *There exist step functions $\{g_{1,A}\}_{A>0}$ and $\{g_{2,A}\}_{A>0}$ as above, such that for any $\varepsilon > 0$*

$$\int_{-A}^{A} |K_{1n} - g_{1,A}|^2 (1,0) \, dG_n(1,0)' + \int_{-A}^{A} |K_{2n} - g_{2,A}|^2 (0,1) \, dG_n(0,1)' < \varepsilon \quad \text{(B.6)}$$

when $A > A(\varepsilon)$ and $n \geqslant n(\varepsilon)$, and such that

$$\int_{-A}^{A} |K_{1,0} - g_{1,A}|^2 (1,0) \, dG_0(1,0)' + \int_{-A}^{A} |K_{2,0} - g_{2,A}|^2 (0,1) \, dG_0(0,1)' < \varepsilon \quad \text{(B.7)}$$

when $A > A(\varepsilon)$.

Define

$$I_{1n}^A = (1,0)\Lambda_n \sum_{s=-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-isx} n^{1/2} g_{1,A}(nx) \Psi(x) \, dx \right\} \varepsilon_s$$

$$= (1,0)\, n^{1/2} \Lambda_n \sum_{s=-\infty}^{\infty} \sum_{\ell=-L}^{L} g_{1,\Delta_\ell} \frac{1}{2\pi} \int_{\Delta_\ell/n} \mathrm{e}^{-\mathrm{i}sx} \boldsymbol{\Psi}(x)\,\mathrm{d}x \varepsilon_s$$

$$= (1,0) \sum_{\ell=-L}^{L} g_{1,\Delta_\ell} Z_n(\Delta_\ell). \tag{B.8}$$

Similarly, define

$$I_{2n}^A = (0,1)\Lambda_n \sum_{s=-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-A/n}^{A/n} \mathrm{e}^{-\mathrm{i}sx} n^{1/2} g_{2,A}(nx) \boldsymbol{\Psi}(x)\,\mathrm{d}x \right\} \varepsilon_s$$

$$= (0,1) \sum_{\ell=-L}^{L} g_{2,\Delta_\ell} Z_n(\Delta_\ell), \tag{B.9}$$

and

$$I_n^A = I_{1n}^A + I_{2n}^A.$$

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

$$I_n^A \overset{d}{\to} I_0^A := (1,0) \sum_{\ell=-L}^{L} g_{1,\Delta_\ell} Z_{G_0}(\Delta_\ell) + (0,1) \sum_{\ell=-L}^{L} g_{2,\Delta_\ell} Z_{G_0}(\Delta_\ell). \tag{B.10}$$

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

$$Y := (1,0) \int_{-\infty}^{\infty} K_{1,0}(x)\,\mathrm{d}Z_{G_0}(x) + (0,1) \int_{-\infty}^{\infty} K_{2,0}(x)\,\mathrm{d}Z_{G_0}(x).$$

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

$$\lim_{A \to \infty} \limsup_n P[|Y_n - I_n^A| \geqslant \varepsilon] = 0 \tag{B.11}$$

it will follow that $Y_n \overset{d}{\to} Y$ by Theorem 25.5 of Billingsley (1986). We prove Eq. (B.11) in Lemma 3. It remains to show that $I_0^A \overset{d}{\to} Y$ as $A \to \infty$.

We have

$$I_0^A - Y = (1,0) \int_{-\infty}^{\infty} (g_{1,A}(x) - K_{1,0}(x))\,\mathrm{d}Z_{G_0}(x)$$

$$+ (0,1) \int_{-\infty}^{\infty} (g_{2,A}(x) - K_{2,0}(x))\,\mathrm{d}Z_{G_0}(x).$$

By Eq. (16) and Cauchy's inequality,

$$\mathrm{E}|I_0^A - Y|^2 \leqslant 3 \int_{-\infty}^{\infty} |g_{1,A}(x) - K_{1,0}(x)|^2 (1,0) G_0(\mathrm{d}x)(1,0)'$$

$$+ 3 \int_{-\infty}^{\infty} |g_{2,A}(x) - K_{2,0}(x)|^2 (0,1) G_0(\mathrm{d}x)(0,1)'.$$

It follows from (B.7) and (B.4) that $E|I_0^A - Y|^2 \to 0$ and hence that $I_0^A \overset{d}{\to} Y$ as $A \to \infty$. $\square$

Lemma 3. *For every $\varepsilon > 0$,*

$$\lim_{A \to \infty} \limsup_n P[|Y_n - I_n^A| \geqslant \varepsilon] = 0.$$

Proof. Since $Y_n - I_n^A = Y_n^A - I_n^A + R_n$, it suffices to show that

$$\lim_{A \to \infty} \limsup_n E[|Y_n^A - I_n^A|^2] \to 0, \tag{B.12}$$

and

$$\lim_{A \to \infty} \limsup_n E[|R_n|^2] \to 0. \tag{B.13}$$

We start by proving (B.12). We have

$$Y_n^A - I_n^A = (1,0)\Lambda_n \sum_{s=-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-isx}[h_{1n}(x) - n^{1/2}g_{1,A}(nx)]\boldsymbol{\Psi}(x)\,\mathrm{d}x \right\} \varepsilon_s$$

$$+ (0,1)\Lambda_n \sum_{s=-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-A/n}^{A/n} e^{-isx}[h_{2n}(x) - n^{1/2}g_{2,A}(nx)]\boldsymbol{\Psi}(x)\,\mathrm{d}x \right\} \varepsilon_s.$$

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

$$Var[Y_n^A - I_n^A] \leqslant C(1,0)\Lambda_n \sum_{s=-\infty}^{\infty} \int_{x=-A/n}^{A/n} e^{-isx}[h_{1n}(x) - n^{1/2}g_{1,A}(nx)]\boldsymbol{\Psi}(x)\,\mathrm{d}x\,\boldsymbol{\Sigma}$$

$$\times \int_{y=-A/n}^{A/n} e^{isy}\overline{[h_{1n}(y) - n^{1/2}g_{1,A}(ny)]}\boldsymbol{\Psi}^*(y)\,\mathrm{d}y\,\Lambda_n(1,0)'$$

$$+ C(0,1)\Lambda_n \sum_{s=-\infty}^{\infty} \int_{x=-A/n}^{A/n} e^{-isx}[h_{2n}(x) - n^{1/2}g_{2,A}(nx)]\boldsymbol{\Psi}(x)\,\mathrm{d}x\,\boldsymbol{\Sigma}$$

$$\times \int_{y=-A/n}^{A/n} e^{isy}\overline{[h_{2n}(y) - n^{1/2}g_{2,A}(ny)]}\boldsymbol{\Psi}^*(y)\,\mathrm{d}y\,\Lambda_n(0,1)'$$

$$= C(1,0)\Lambda_n \int_{-A/n}^{A/n} |h_{1n}(x) - n^{1/2}g_{1,A}(nx)|^2 f(x)\,\mathrm{d}x\,\Lambda_n(1,0)'$$

$$+ C(0,1)\Lambda_n \int_{-A/n}^{A/n} |h_{2n}(x) - n^{1/2}g_{2,A}(nx)|^2 f(x)\,\mathrm{d}x\,\Lambda_n(0,1)',$$

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

$$Var[Y_n^A - I_n^A] \leqslant C \int_{-A}^{A} |K_{1n}(x) - g_{1,A}(x)|^2 (1,0) \, \mathrm{d}G_n(x)(1,0)'$$

$$+ C \int_{-A}^{A} |K_{2n}(x) - g_{2,A}(x)|^2 (0,1) \, \mathrm{d}G_n(x)(0,1)'.$$

Eq. (B.12) now follows from Lemma 2, Eq. (B.6). We next prove (B.13). Using an argument very similar to that given in proving (B.12), we conclude from (B.3) that

$$Var[R_n] \leqslant C \int_{[-n\pi, n\pi] \backslash [-A, A]} |K_{1n}(x)|^2 (1,0) \, \mathrm{d}G_n(x)(1,0)'$$

$$+ C \int_{[-n\pi, n\pi] \backslash [-A, A]} |K_{2n}(x)|^2 (0,1) \, \mathrm{d}G_n(x)(0,1)'.$$

Thus, (B.13) follows from (B.5).   □

References

Baillie, R.T., Bollerslev, T., 1994. Cointegration, fractional cointegration, and exchange rate dynamics. Journal of Finance 49, 737–745.

Billingsley, P., 1986. Probability and Measure, 2nd Edition. Wiley, New York.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, 2nd Edition. Springer, New York.

Campbell, J.Y., Lo, A.W., Mackinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.

Chen, W.W., Deo, R.S., 2000. A generalized portmanteau goodness-of-fit test for time series models. Working Paper, SOR-2000-6, New York University.

Cheung, Y., Lai, K., 1993. A fractional cointegration analysis of purchasing power parity. Journal of Business and Economic Statistics 11, 103–112.

Deo, R.S., Hurvich, C.M., 1998. Linear trend with fractionally integrated errors. Journal of Time Series Analysis 19, 379–397.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 69, 1057–1072.

Diebold, F.X., Gardeazabal, J., Yilmaz, K., 1994. On cointegration and exchange rate dynamics. Journal of Finance 49, 727–735.

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

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Hart, J.D., 1989. Differencing as an approximate de-trending device. Stochastic Processes and their Applications 31, 251–259.

Hurvich, C.M., 2001. Model selection for broadband semiparametric estimation of long memory in time series. Journal of Time Series Analysis 22, 679–709.

Hurvich, C.M., Beltrao, K., 1993. Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. Journal of Time Series Analysis 14, 455–472.

Hurvich, C.M., Brodsky, J., 2001. Broadband semiparametric estimation of the memory parameter of a long-memory time series using fractional exponential models. Journal of Time Series Analysis 22, 221–249.

Hurvich, C.M., Chen, W.W., 2000. An efficient taper for potentially overdifferenced long-memory time series. Journal of Time Series Analysis 21, 155–180.

Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. Journal of Time Series Analysis 16, 17–41.

Hurvich, C.M., Moulines, E., Soulier, P., 2002. The FEXP estimator for potentially non-stationary linear time series. Stochastic Processes and their Applications 97, 307–340.

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

Künsch, H.R., 1987. Statistical aspects of self-similar processes. In Proceedings of the First World Congress of the Bernoulli Society, Vol. 1, VNU Science Press, Utrecht, pp. 67–74.

Lobato, I.N., 1997. Consistency of the averaged cross-periodogram in long memory series. Journal of Time Series Analysis 18, 137–155.

Lobato, I.N., Velasco, C., 2000. Long memory in stock-market trading volume. Journal of Business and Economic Statistics 18, 410–427.

Lütkepohl, H., Saikkonen, P., 2000. Testing for the cointegrating rank of a VAR process with a time trend. Journal of Econometrics 95, 177–198.

Moulines, E., Soulier, P., 1999. Broadband log-periodogram regression of time series with long-range dependence. Annals of Statistics 27, 1415–1439.

Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. Econometrica 59, 283–306.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Robinson, P.M., 1994. Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515–539.

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

Robinson, P.M., Marinucci, D., 2000. The averaged periodogram for nonstationary vector time series. Statistical Inference for Stochastic Processes 3, 149–160.

Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes. Annals of Statistics 29, 947–986.

Stock, J.H., Watson, M.W., 1988. Testing for common trends. Journal of the American Statistical Association 83, 1097–1107.

Terrin, N., 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.

Terrin, N., 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.

Tukey, J.W., 1967. An introduction to the calculations of numerical spectrum analysis. In: Harris, B. (Ed.), Advanced Seminar on Spectral Analysis of Time Series. Wiley, New York, pp. 25–46.

Velasco, C., 1999a. Non-stationary log-periodogram regression. Journal of Econometrics 91, 325–371.

Velasco, C., 1999b. Gaussian semiparametric estimation of non-stationary time series. Journal of Time Series Analysis 20, 87–127.

Velasco, C., Robinson, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95, 1229–1243.

Zhurbenko, I.G., 1979. On the efficiency of estimates of a spectral density. Scandanavian Journal of Statistics 6, 49–56.