Discussion Paper #2010–11

Exact Local Whittle Estimation of Fractionally Cointegrated Systems*

Katsumi Shimotsu
Department of Economics
Hitotsubashi University and Queen's University
shimotsu@econ.hit-u.ac.jp

September 2010

Abstract

Semiparametric estimation of a bivariate fractionally cointegrated system is considered. We propose a two-step procedure that accommodates both (asymptotically) stationary ($\delta < 1/2$) and nonstationary ($\delta \geq 1/2$) stochastic trend and/or equilibrium error. A tapered version of the local Whittle estimator of Robinson (2008) is used as the first-stage estimator, and the second-stage estimator employs the exact local Whittle approach of Shimotsu and Phillips (2005). The consistency and asymptotic distribution of the two-step estimator are derived. The estimator of the memory parameters has the same Gaussian asymptotic distribution in both the stationary and nonstationary case. The convergence rate and the asymptotic distribution of the estimator of the cointegrating vector are affected by the difference between the memory parameters. Further, the estimator has a Gaussian asymptotic distribution when the difference between the memory parameters is less than 1/2.

JEL Classification: C22

Key words and phrases: discrete Fourier transform; fractional cointegration; long memory; nonstationarity; semiparametric estimation; Whittle likelihood

*The author is grateful to the co-editors and two anonymous referees whose comments greatly improved the paper. For helpful and constructive comments, the author thanks Richard Baillie, Benoit Perron, Peter Phillips, and the seminar participants at Michigan State University, Rochester University, University of Montreal, and the Conference in Honour of Peter Phillips. The financial support from the ESRC under Grant R000223986 and the SSHRC is gratefully acknowledged.

1 Introduction

The analysis of the long-run equilibrium relationship between economic variables is now a common task in empirical econometric modeling. The concept of cointegration (Engle and Granger, 1987) has provided powerful tools for the analysis of these issues. Two random processes are said to be cointegrated if they have the same memory parameter but their linear combination has a smaller memory parameter. Cointegrated random processes form a long-run equilibrium relationship, in which the cointegrated processes are driven by a common stochastic trend and the equilibrium error has less persistence than the stochastic trend.

The fractional cointegration analysis generalizes the conventional $I(0)/I(1)$ cointegration analysis by allowing the memory parameter of the variables to be any real number. The system is driven by an $I(\delta_2)$ common stochastic trend and accompanied by an $I(\delta_1)$ equilibrium error. It provides a more flexible apparatus for analyzing long-run relationships between economic time series. For instance, consider the following two cases:

- Two time series have the same memory parameter $\delta_2 < 1$, and the equilibrium error has a memory parameter $\delta_1 < \delta_2$.
- Two time series are $I(1)$, but the equilibrium error is $I(\delta)$, where $\delta \in (0, 1)$.

Clearly, the two time series form a long-run equilibrium in the above two cases, but the conventional $I(0)/I(1)$ cointegration cannot accommodate them. When empirical researchers apply the $I(0)/I(1)$ cointegration to such data, it leads to either (i) a false rejection of the existence of an equilibrium relationship, or (ii) misspecification of the degree of persistence of the stochastic trend and/or the equilibrium error.

Empirical relevance of fractional cointegration has been long recognized, and fractional cointegration has been applied in many areas in economics and social science, including exchange rate dynamics (Cheung and Lai, 1993; Baillie and Bollerslev, 1994), interest rate dynamics (Dueker and Startz, 1998), and poll data (Davidson and Peel, 2006). More recently, fractional cointegration has been shown to be useful in modeling financial volatility series. See, for example, Brunetti and Gilbert (2000), Bandi and Perron (2006), Christensen and Nielsen (2006), and Cassola and Morana (2010).

Because of its attractiveness and relevance, several attempts have been made to develop a semiparametric estimator of fractionally cointegrated systems, but technical difficulties have hampered its development until recently. Robinson (2008) derives the consistency and asymptotic normality of the local Whittle estimator of a stationary fractionally cointegrated system under the assumption $0 \leq \delta_1 < \delta_2 < 1/2$.[1] Hassler et al. (2006) and Velasco (2003) seek to estimate δ_1 by applying semiparametric estimators to the residuals from cointegrating regressions, but they require $\delta_2 - \delta_1 > 1/2$. Nielsen (2007) considers joint estimation of δ_1, δ_2 and the cointegrating vector under the assumption $0 \leq \delta_1, \delta_2 < 1/2$, but derives its asymptotic distribution only under the long-run exogeneity between the stochastic trend and equilibrium error. Nielsen and Frederiksen (2008) consider a fully modified narrow-band least squares (NBLS) estimator that corrects the endogeneity bias of the NBLS estimator.[2]

The above procedures have an additional difficulty: prior to estimation, the researcher needs to know the range of the value of δ_1 and δ_2. Because the semiparametric estimators of δ employed by these procedures have a standard limiting distribution only for $-1/2 < \delta < 3/4$, one needs either to assume $\delta < 3/4$ and use row data or assume $\delta > 1/2$ and use differenced data. This poses problems for the following reasons:

1. Typically, whether $\delta \gtreqless 1/2$ is unknown *a priori*; indeed, often empirical researchers want to *test* whether $\delta \gtreqless 1/2$, because this determines whether the process is stationary (if $\delta > 1/2$) or nonstationary (if $\delta < 1/2$).

2. Because the value of δ of most economic time series lies between 0 and 1, if two economic variables are cointegrated, then the memory parameter of the equilibrium error may take a value larger than or smaller than $1/2$.

3. Because one needs to assume either $\delta < 3/4$ or $\delta > 1/2$, the confidence interval must lie either to the left of $3/4$ or to the right of $1/2$.

This paper develops an estimation and inference method for bivariate fractionally cointegrated systems. The proposed procedure accommodates both stationary and nonstationary processes for the stochastic trend and cointegrating error. We achieve this by a two-step procedure; a tapered and trimmed version of the estimator by Robinson (2008) is used as the first-stage estimator, and the second-stage estimator uses the exact local Whittle (ELW) approach of Shimotsu and Phillips (2005). The second-stage ELW estimator uses neither tapering nor trimming. In a univariate context, Shimotsu and Phillips (2005) prove the consistency and asymptotic normality of the ELW estimator for both

[1]Strictly speaking, the system analyzed by Robinson (2008) is more general and includes a fractionally cointegrated system as a special case.

[2]Some studies focus on testing the null hypothesis of no cointegration by estimating the rank of the (normalized) spectral density matrix at frequency zero. See, for example, Robinson and Yajima (2002), Chen and Hurvich (2002), and Nielsen and Shimotsu (2007).

stationary and nonstationary δ when the mean of the process is known. Shimotsu (2010a) extends it to accommodate an unknown mean and a polynomial time trend.

We derive the asymptotic behavior of the tapered estimator and the second-stage ELW estimator. The ELW estimator of δ_1 and δ_2 has the same Gaussian asymptotic distribution in both the stationary and nonstationary case. The asymptotics of the estimator of the cointegrating vector is affected by the value of $\delta_2 - \delta_1$. Its asymptotic distribution is Gaussian only when $\delta_2 - \delta_1 < 1/2$, and it has a different convergence rate when $\delta_2 - \delta_1 > 1/2$. The first-stage tapered estimator is shown to be consistent for $-1/2 < \delta_1 < \delta_2 < \bar{\delta} < \infty$ and has the same convergence rate in both the stationary and nonstationary case. The estimator imposes an additional restriction that the (pseudo-) spectral density of the processes has no poles outside the origin.

The remainder of the paper is organized as follows. Section 2 briefly reviews the model of fractional cointegration. Section 3 derives the consistency and convergence rate of the tapered local Whittle estimator. Section 4 shows the asymptotic distribution of the second-stage ELW estimator. Section 5 reports some simulation results. Section 6 provides an empirical application that revisits the fractional cointegration analysis between implied and realized volatility by Bandi and Perron (2006). Proofs of the main theorems are collected in Section 7.

2 Preliminaries: a model of fractional cointegration

We consider a model where the observed variables x_t and y_t are fractionally cointegrated. Specifically, x_t and y_t are generated by the model

$$\begin{cases} (1-L)^{\delta_1}(y_t - \beta x_t) = u_{1t} I\{t \geq 1\}, & t = 1, 2, \ldots, \\ (1-L)^{\delta_2} x_t = u_{2t} I\{t \geq 1\}, & t = 1, 2, \ldots, \\ y_t = x_t = 0, & t \leq 0, \end{cases} \tag{1}$$

where $\beta \neq 0$, and $u_t = (u_{1t}, u_{2t})'$ is stationary with zero mean and spectral density matrix $f_u(\lambda)$ with $f_u(0) = \Omega$. We assume $-1/2 < \delta_1 < \delta_2 < \bar{\delta} < \infty$; hence, x_t and y_t are individually $I(\delta_2)$ because their δ_2th differences have a spectral density that is bounded and bounded away from the origin. But their linear combination, $y_t - \beta x_t$, has a memory parameter δ_1 that is smaller than δ_2. We may also write (1) in matrix notation as

$$Bz_t = \begin{bmatrix} (1-L)^{-\delta_1} & 0 \\ 0 & (1-L)^{-\delta_2} \end{bmatrix} u_t I\{t \geq 1\}, \quad B = \begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix}, \quad z_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix}. \tag{2}$$

Expanding the binomial in the second row of (1) gives the form

$$\sum_{k=0}^{t} \frac{\Gamma(k-\delta_2)}{\Gamma(-\delta_2)k!} x_{t-k} = u_{2t} I\{t \geq 1\}, \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function. The model (1) provides a valid data-generating process for any value of (δ_1, δ_2), and accommodates both the nonstationary and (asymptotically) stationary case. When $\delta_2 > 1/2$, x_t is nonstationary, and when $\delta_2 < 1/2$, x_t is asymptotically covariance stationary. Setting $\delta_2 = 1$ and $\delta_1 = 0$ gives the conventional $I(0)/I(1)$ cointegration.

For a vector time series a_t, define the discrete Fourier transform (dft) and the periodogram evaluated at the fundamental frequencies as

$$\begin{aligned} w_a(\lambda_j) &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} a_t e^{it\lambda_j}, \quad \lambda_j = \frac{2\pi j}{n}, \quad j = 1, \ldots, n, \\ I_a(\lambda_j) &= w_a(\lambda_j) \bar{w}_a(\lambda_j), \end{aligned} \tag{4}$$

where $\bar{x}$ denotes the conjugate transpose of x.

3 First-stage estimation: tapered local Whittle estimation

As the first-step estimator, we use the tapered version of the local Whittle estimator of stationary cointegrated systems by Robinson (2008). Robinson (2008) derives the consistency and asymptotic normality of the local Whittle estimator of a stationary bivariate system that includes fractional cointegration as a special case under the assumption $0 \leq \delta_1 < \delta_2 < 1/2$. Our objective is to develop an estimator of $\vartheta = (\beta, \delta')' = (\beta, \delta_1, \delta_2)'$ that does not impose prior restrictions on the stationarity of the processes in the system.

As shown by Velasco (1999) and Lobato and Velasco (2000), tapering allows one to accommodate both stationary and nonstationary processes in local Whittle estimation. We use the taper considered by Velasco (1999). Let h_t denote a pth-order taper generated by Kolmogorov's proposal. Then h_t satisfies the regularity conditions in Velasco (1999) and Robinson (2005), and the tapered estimator is invariant to a polynomial time trend of order $p-1$. Define the tapered dft and periodogram of a_t as $w_a^T(\lambda_j) = (2\pi n)^{-1/2}\sum_{t=1}^{n} h_t z_t e^{it\lambda_j}$ and $I_a^T(\lambda_j) = w_a^T(\lambda_j)\bar{w}_a^T(\lambda_j)$.

We follow notation in Robinson (2008) in most parts of the paper. Let m be some integer less than n, and let $\kappa \in (0,1)$ be an arbitrary small number. Let $\sum_{j(p,\kappa)}^{m}$ denote the sum taken over $j = p, 2p, \ldots, m$ for $j \geq [\kappa m]$. Using κ introduces a trimming of the periodogram ordinates from below. The trimming controls the behavior of the objective function when $\delta_2 - \delta_1 > 1/2$. The tapered local Whittle estimator is defined as (see Shimotsu (2010b) for derivation)

$$
\begin{aligned}
R(\vartheta) &= \log\det\hat{\Omega}^T(\vartheta) - 2(\delta_1+\delta_2)\frac{p}{(1-\kappa)m}\sum_{j(p,\kappa)}^{m}\log\lambda_j, \qquad (5)\\
\hat{\Omega}^T(\vartheta) &= \frac{p}{(1-\kappa)m}\sum_{j(p,\kappa)}^{m}\mathrm{Re}\left[\Psi(\lambda_j;\delta)BI_z^T(\lambda_j)B'\bar{\Psi}(\lambda_j;\delta)\right],
\end{aligned}
$$

where

$$
\Psi(\lambda;\delta) = \mathrm{diag}(\lambda^{\delta_1}, \lambda^{\delta_2}e^{-i(\pi-\lambda_j)(\delta_2-\delta_1)/2}).
$$

We estimate $\vartheta = (\beta, \delta')'$ by $\hat{\vartheta} = \arg\min_{\Theta} R(\vartheta)$. The parameter space is defined as $\Theta = \Theta_\beta \times \Theta_\delta$, where Θ_β is an arbitrary large interval and

$$
\Theta_\delta = \{\delta : -1/2 + \eta_1 \leq \delta_1 \leq \delta_2 - \eta_2 \leq p - 1/2 - \eta_3\}, \qquad (6)
$$

where the η_i's are arbitrary small positive numbers such that $\eta_2 < \eta_3$. The constraint, $\delta_1 \leq \delta_2 - \eta_2$, is also used in Robinson (2008). This constraint imposes that there is cointegration, but this constraint is necessary because β is not identified from the local Whittle-type objective function when $\delta_1 = \delta_2$. Relaxing this restriction remains an important future topic.

Robinson (2008) introduces an additional parameter γ to model the phase between $y_t - \beta x_t$ and x_t flexibly. In place of $\Psi(\lambda;\delta)$, Robinson (2008) uses $\Psi(\lambda;\delta,\gamma) = \mathrm{diag}(\lambda^{\delta_1}, \lambda^{\delta_2}e^{-i\gamma})$ and defines the objective function $R(\cdot)$ in terms of four parameters, $(\beta, \delta_1, \delta_2, \gamma)$. In effect, our parameterization imposes the restriction $\gamma = (\delta_2 - \delta_1)\pi/2$ to the model of Robinson (2008), which is implied by a fractionally cointegrated system (1).

We introduce the following assumptions on m and the stationary component u_t in (1). Henceforth, we denote the true parameter values by Ω_0 and $\vartheta_0 = (\beta_0, \delta_0')'$. To simplify the presentation and proof, one set of assumptions is used for both the consistency and the convergence rate of the tapered estimator.

Assumption 1 *$f_u(\lambda) - \Omega_0 = O(\lambda^b)$ as $\lambda \to 0+$ for some $b \in (0, 2]$, and $\Omega_0 = (\omega_{k\ell})$ is real, symmetric, finite, and positive definite.*

Assumption 2

$$u_t - Eu_0 = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} ||A_j||^2 < \infty,$$

where $||\cdot||$ denotes the Euclidean norm and $E(\varepsilon_t|F_{t-1}) = 0$, $E(\varepsilon_t \varepsilon_t'|F_{t-1}) = I_2$ a.s., $t = 0, \pm 1, \ldots$, in which F_t is the σ-field generated by ε_s, $s \leq t$, and there exists a scaler random variable ε such that $E\varepsilon^2 < \infty$ and for all $\eta > 0$ and some $K > 0$, $\Pr(||\varepsilon_t|| > \eta) \leq K \Pr(\varepsilon^2 > \eta)$. Further, the elements of ε_t have a.s. constant third and fourth moment and cross-moments conditional on F_{t-1}.

Assumption 3 *$A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ satisfies, for b defined in Assumption 1,*

$$\Psi(\lambda; \delta_0) A(\lambda) - P = O(\lambda^b) \quad \text{as } \lambda \to 0+,$$

where P satisfies $P = PP' = \Omega_0$ and δ_0 is the true value of $\delta = (\delta_1, \delta_2)'$. Further, $A(\lambda)$ is differentiable in a neighborhood of $\lambda = 0$, and $\partial A(\lambda)/\partial\lambda$ satisfies $\Psi(\lambda; \delta_0)(\partial/\partial\lambda)A(\lambda) = O(\lambda^{-1})$ as $\lambda \to 0+$.

Assumption 4 *ϑ_0 is an interior point of Θ.*

Assumption 5 *For any $C < \infty$,*

$$\frac{1}{m} + \frac{m^{1+2b}(\log m)^2}{n^{2b}} + \frac{(\log n)^C}{m} \to 0 \quad \text{as } n \to \infty.$$

Assumption 6 *$f_u(\lambda)$ is bounded and bounded away from zero for $\lambda \in [0, \pi]$.*

Assumption 7 *The order p of the taper satisfies $p \geq 2$ and $s_{02} < p$, where s_{0i} is the true parameter value of $s_i = [\delta_i + 1/2]$, $i = 1, 2$.*

Assumptions 1-3 are essentially the same as Assumptions B1-B5 and A6 of Robinson (2008), but we impose them in terms of u_t rather than z_t. Assumption A6 of Robinson

(2008) imposes $g_{12} > 0$, but we do not need to assume it because the phase (γ in Robinson (2008)) is identified by δ in our model. Assumption 6 is used in Robinson (2005). This assumption is necessary in approximating the tapered dft of a type-II fractionally integrated process by that of a type-I fractionally integrated process. This assumption excludes the poles outside the origin, but it imposes no additional assumptions in terms of the smoothness of the spectral density beyond Assumptions 1-3.

The following theorem establishes the convergence rate of $\hat{\vartheta}$. Define $\nu = \delta_2 - \delta_1$ and let ν_0 denote its true value. ν_0 affects the convergence rate of $\hat{\beta}$, but it does not affect the convergence rate of $\hat{\delta}$.

Theorem 1 *Suppose z_t is generated by (1) and Assumptions 1-7 hold. Then $\hat{\delta} - \delta_0 = O_p(m^{-1/2})$ and $\hat{\beta} - \beta_0 = O_p(m^{-1/2}(m/n)^{\nu_0})$ as $n \to \infty$.*

4 Exact local Whittle estimation of fractional cointegration

The tapered estimator is consistent for both stationary and nonstationary z_t but is less efficient than the nontapered estimator in the stationary case. In this section, we propose and analyze a two-step estimator that is based on the idea of the exact local Whittle estimation of Shimotsu and Phillips (2005).

We start from the (negative) Whittle likelihood of u_t based on frequencies up to λ_m and up to scale multiplication:

$$\sum_{j=1}^{m} \log(\det f_u(\lambda_j)) + \sum_{j=1}^{m} \operatorname{tr}\left[f_u(\lambda_j)^{-1} I_u(\lambda_j)\right], \tag{7}$$

where m is some integer less than n. Now we transform the likelihood function (7) to be data dependent. Define

$$I_{\Delta^\delta z}(\lambda_j;\beta) = w_{\Delta^\delta z}(\lambda_j;\beta)\bar{w}_{\Delta^\delta z}(\lambda_j;\beta), \quad w_{\Delta^\delta z}(\lambda_j;\beta) = \begin{pmatrix} w_{\Delta^{\delta_1}(y-\beta x)}(\lambda_j) \\ w_{\Delta^{\delta_2} x}(\lambda_j) \end{pmatrix}.$$

Theorem 2.2 of Phillips (1999) (or Lemma 5.1 of Shimotsu and Phillips (2005)) provides an algebraic relationship that connects $w_u(\lambda_j)$ and $w_{Bz}(\lambda_j)$:

$$w_u(\lambda_j) = w_{\Delta^\delta z}(\lambda_j;\beta) = \Lambda_n(e^{i\lambda_j};\delta)v_{Bz}(\lambda_j;\delta), \tag{8}$$

where

$$\Lambda_n(e^{i\lambda_j};\delta) = \begin{pmatrix} D_n(e^{i\lambda_j};\delta_1) & 0 \\ 0 & D_n(e^{i\lambda_j};\delta_2) \end{pmatrix}, \quad v_{Bz}(\lambda_j;\delta) = \begin{pmatrix} v_{Bz_1}(\lambda_j;\delta_1) \\ v_{Bz_2}(\lambda_j;\delta_2) \end{pmatrix},$$
$$v_{Bz_a}(\lambda_j;\delta_a) = w_{Bz_a}(\lambda_j) - D_n(e^{i\lambda_j};\delta_a)^{-1}(2\pi n)^{-1/2}\widetilde{Bz}_{a,\lambda_j n}(\delta_a), \quad a=1,2,$$

and Phillips (1999) provides the exact definition of $\widetilde{A}_{\lambda_j n}(d)$ for a process a_t.

Although $v_{Bz}(\lambda_j;d)$ is not a periodogram of Bz_t, we may view (8) as the frequency domain representation of Bz_t where $\Lambda_n(e^{i\lambda_j};\delta)$ acts as a transfer function. Using (8) in conjunction with the local approximation $f_u(\lambda_j) \sim \Omega$ and $|D_n(e^{i\lambda_j};\delta_a)|^2 \sim \lambda_j^{2\delta_a}$, the objective function is simplified to

$$Q_m^*(\vartheta,\Omega) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\det\Omega - 2\log(\lambda_j^{\delta_1}+\lambda_j^{\delta_2}) + \operatorname{tr}\left[\Omega^{-1}I_{\Delta^\delta z}(\lambda_j;\beta)\right]\right\}.$$

We propose to estimate (ϑ,Ω) by minimizing $Q_m^*(\vartheta,\Omega)$, so that

$$(\vartheta^*,\Omega^*) = \underset{\Omega\in(0,\infty)^2,\vartheta\in\Theta}{\arg\min}\; Q_m^*(\vartheta,\Omega),$$

where Θ is defined in (6). Concentrating out Ω from $Q_m^*(\vartheta,\Omega)$, we find that ϑ^* satisfies

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

where

$$R^*(\vartheta) = \log\det\tilde{\Omega}^*(\vartheta) - 2(\delta_1+\delta_2)\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j, \quad \tilde{\Omega}^*(\vartheta) = \frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}[I_{\Delta^\delta z}(\lambda_j;\beta)]. \tag{9}$$

We consider the two-step estimator based on the objective function $R^*(\vartheta)$. Let $\hat{\vartheta}$ be the tapered local Whittle estimator of ϑ. The two-step estimator is defined as

$$\vartheta^* = \hat{\vartheta} - [(\partial^2/\partial\vartheta\partial\vartheta')R^*(\hat{\vartheta})]^{-1}(\partial/\partial\vartheta)R^*(\hat{\vartheta}). \tag{10}$$

Iterating the above procedure and updating the estimator by $\vartheta^*_{(2)} = \vartheta^* - [(\partial^2/\partial\vartheta\partial\vartheta')R^*(\vartheta^*)]^{-1}\times(\partial/\partial\vartheta)R^*(\vartheta^*)$ and similarly for $\vartheta^*_{(3)}$ do not change the asymptotic distribution of the estimator, but we find that iterating the procedure can improve its finite sample properties.

The following assumption is additionally imposed.

Assumption 8 *$\sum_{|j|\geq k}\gamma_j = O((\log(k+1))^{-4})$, and $\sum_{j\geq k} A_j = O((\log(k+1))^{-4})$, where $\gamma_j = Eu_t u'_{t-j}$.*

Assumption 8 is also used in Phillips and Shimotsu (2004), who analyze the asymptotics of the local Whittle estimator under type-II processes. This assumption is fairly mild and allows for a pole and discontinuity in $f_u(\lambda)$ at $\lambda \neq 0$. For more details, see Phillips and Shimotsu (2004).

The following theorem establishes the asymptotic distribution of the exact local Whittle estimator.[3]

Theorem 2 *Suppose z_t is generated by (1), Assumptions 1-8 hold, and $\nu_0 = \delta_{20} - \delta_{10} \neq 1/2$. Then, as $n \to \infty$,*

(a) When $\nu_0 \in (0, \frac{1}{2})$,

$$m^{1/2}\Delta_n^*(\vartheta^* - \vartheta_0) \to_d N(0, \Xi^{-1}), \quad \Delta_n^* = diag\{\lambda_m^{-\nu_0}, 1, 1\},$$

where $\Xi_{11} = 2\mu\{(1-2\nu_0)^{-1} - (1-\nu_0)^{-2}\cos^2(\gamma_0)\}\omega_{22}/\omega_{11}$, $\Xi_{22} = \Xi_{33} = 4 + (\pi^2/4-1)2\mu\rho^2$, $\Xi_{23} = \Xi_{32} = -(\pi^2/4-1)2\mu\rho^2$, $\Xi_{12} = \Xi_{21} = -2\mu\nu_0(1-\nu_0)^{-2}\cos(\gamma_0)\omega_{12}/\omega_{11} + (\pi/2)2\mu(1-\nu_0)^{-1}\sin(\gamma_0)\omega_{12}/\omega_{11}$, $\Xi_{13} = \Xi_{31} = -\Xi_{12}$, where $\mu = (1-\rho^2)^{-1}$, $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2}$, and $\gamma_0 = \nu_0\pi/2$.

(b) When $\nu_0 \in (\frac{1}{2}, \frac{3}{2})$, assume further that $n^{-b}m^{1-\nu_0+b}$ is bounded and the cumulant spectral density of u_t, $f_u(\lambda, \mu, \omega)$, is continuous at $\lambda = \mu = \omega = 0$ and satisfies $\sup_{\mu,\omega}\int |f_u(\lambda, \mu, \omega)|^2 d\lambda < \infty$. Then

$$m^{1/2}(\delta^* - \delta_0) \to_d N(0, \Xi_\delta^{-1}), \quad \beta^* - \beta_0 = O_p(n^{-\nu_0}),$$

where Ξ_δ is the lower-right (2×2) block of Ξ.

Remark 1 *Because our parameterization and that in Robinson (2008) are related by $(\vartheta_1, \vartheta_2, \vartheta_3) = (\theta_1, \theta_3 - (\pi/2)\theta_2, \theta_4 + (\pi/2)\theta_2)$, the corresponding relation holds between the asymptotic variance of the two estimators.*

---

[3] In the context of univariate ELW estimation, Shimotsu (2010a) shows that the two-step ELW estimator with a mean correction accommodates an unknown mean and has the same asymptotic distribution as the ELW estimator. In the context of our model, suppose the data-generating process is given by $Bz_t - \upsilon = \text{diag}\{(1-L)^{-\delta_1}, (1-L)^{-\delta_2}\}u_t I\{t \geq 1\}$, where $\upsilon = (\upsilon_1, \upsilon_2)'$ is a nonrandom vector. Estimate υ_k by $\hat{\upsilon}_k(\delta_k) = w(\delta_k)m^{-1}\sum_{t=1}^m (Bz)_{kt} + (1 - w(\delta_k))(Bz)_{k1}$, where $(Bz)_{kt}$ is the kth element of Bz_t, and $w(x)$ a weight function used in Shimotsu (2010a). In view of Shimotsu (2010, Theorem 3), the asymptotic distribution of δ_k^* is not affected by υ if $\Delta^{\delta_1}(y_t - \beta x_t)$ and $\Delta^{\delta_2}x_t$ in the objective function are replaced by $\Delta^{\delta_1}(y_t - \beta x_t - \hat{\upsilon}_1(\delta_1))$ and $\Delta^{\delta_2}(x_t - \hat{\upsilon}_2(\delta_2))$. Shimotsu (2010a) also shows that the presence of a polynomial time trend can be dealt with by prior detrending of the data.

Remark 2 *Ξ_δ can be written as $\Xi_\delta = 2[I_2 + \Omega_0 \odot (\Omega_0)^{-1} + (\pi^2/4)(\Omega_0 \odot \Omega_0^{-1} - I_2)]$. Not surprisingly, this is identical to the asymptotic variance of the bivariate local Whittle estimator analyzed by Shimotsu (2007).*

Remark 3 *The additional rate condition on m for $\nu_0 \in (\frac{1}{2}, \frac{3}{2})$ is needed to control the bias from the periodogram of $\Delta^{-\nu_0} x_t$ that appear in the derivatives of the objective function with respect to β. The condition is innocuous when $\nu_0 \geq 1$. When $b = 2$, this condition becomes $n^{-2} m^{3-\nu_0} = O(1)$, which is slightly weaker than $m = O(n^{5/4-\epsilon})$, a condition often used in univariate local Whittle estimation.*

Remark 4 *The convergence rate of β^* is $n^{\nu_0} m^{1/2-\nu_0}$ when $\nu_0 \in (0, 1/2)$ and $n^{\nu 0}$ when $\nu_0 > 1/2$. The two-step estimator converges at the rate of $n^{\nu_0} m^{1/2-\nu_0}$ for $\nu_0 \gtreqless 1/2$. Thus, the convergence rate of β^* is no slower than that of $\hat{\beta}$.*

Remark 5 *The asymptotic distribution of β^* for $\nu_0 > 1/2$ remains an open question. We conjecture that it is not Gaussian, because the part of the Hessian corresponding to β (namely, the $(1, 1)$th element) does not converge to a nonrandom constant.*

A consistent estimate of $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2}$ is necessary to construct a confidence interval for ϑ_0. We can estimate Ω and ρ by $\tilde{\Omega}^*(\vartheta^*) = m^{-1} \sum_{j=1}^{m} \mathrm{Re}[I_{\Delta^{\delta^*} z}(\lambda_j; \beta^*)]$ and $\rho^* = \tilde{\Omega}^*(\vartheta^*)_{12}/(\tilde{\Omega}^*(\vartheta^*)_{11}\tilde{\Omega}^*(\vartheta^*)_{22})^{1/2}$, respectively, which converge to Ω and ρ in probability under the assumptions of Theorem 2. The asymptotic distribution of ρ^* can be derived in the same manner as the proof of Lemma 5 of Nielsen and Shimotsu (2007).

The convergence rate of β^* depends on the difference in the memory parameters, $\nu_0 = \delta_{20} - \delta_{10}$. In the spurious regression wherein a fractionally integrated process is regressed on another unrelated fractionally integrated process, the convergence rate of the slope estimate depends on the memory parameters of the processes involved. Tsay and Chung (2000) analyze this problem and show that the convergence rate depends on either the difference between or the sum of the memory parameter of the regressor and the dependent variable, depending on their stationarity.

Similar to many other semiparametric estimators, the ELW approach estimates only the long-run parameters, δ_1 and δ_2. The estimation of short-run parameters, however, can be critical for evaluation of impulse response weights or forecasts. Baillie and Kapetanios (2009) demonstrate using simulations that when the short-run dynamic of u_t is strong (for example, $AR(1)$ with the autoregressive parameter being 0.8 or 0.95), the univariate local Whittle estimator gives biased estimates of δ and the impulse response weights.

Hence, one must interpret semiparametric estimates carefully when one suspects that the short-run dynamic of u_t is strong.

Using the lag operator $L_b = (1 - (1 - L)^b)$, Johansen (2008) introduces an alternate representation of fractionally integrated processes that is more amenable to economic interpretation. If z_t is defined using Johansen's representation, then z_t is a function of $\{u_t\}_{t=1}^{\infty}$ and another component (denoted by μ_t in Johansen (2008)) that depends on $\{z_t\}_{t=-\infty}^{0}$ (Johansen, 2008, Theorem 8). Consequently, whether the asymptotic results of this paper carry through depends on what is assumed on $\{z_t\}_{t=-\infty}^{0}$. For example, if one conditions on $\{z_t\}_{t=-\infty}^{0}$ and assumes $\{z_t\}_{t=-\infty}^{0}$ is finite, then the asymptotics of the stationary local Whittle estimator and tapered estimator would remain unchanged (c.f. Shimotsu and Phillips, 2006; Shao and Wu, 2007). The effect of Johansen's representation on the ELW estimator needs more careful analysis because the ELW estimator uses fractional differences of z_t. We conjecture that Theorem 2 would still hold conditional on $\{z_t\}_{t=-\infty}^{0}$ if a suitable assumption is imposed on $\{z_t\}_{t=-\infty}^{0}$.

5 Simulations

This section reports some simulations that were conducted to examine the finite sample performance of the developed estimator. We generate a fractionally cointegrated system according to (1) with $\beta = 1$. u_t is generated by $iidN(0, \Sigma)$, where the diagonal elements of Σ are fixed as 1 and the off-diagonal elements of Σ, ρ, are set to $(0.0, 0.4, 0.8)$. The bias, standard deviation, and root mean squared error (RMSE) are computed using 10,000 replications. The sample size (n) and m are chosen as $n = 512$ and $m = n^{0.65} = 57$, respectively. Further, $\kappa = 0.1$ is used in the trimming and yields $[\kappa m] = 5$. The value of δ_1 is fixed as 0.1. The value of δ_2 is set to $(0.4, 0.8, 1.3)$ to analyze three cases: $\nu_0 \in (0, 1/2)$, $\nu_0 \in (1/2, 1)$, and $\nu_0 \in [1, 3/2)$. We compare three estimators: the two-step ELW estimator, the tapered estimator, and the stationary local Whittle (LW) estimator of Robinson (2008). In the two-step estimation, quasi-Newton updating is repeated until convergence. The mean correction by Shimotsu (2010a), discussed in footnote 2, is applied to the ELW estimator because it is found to improve the finite performance of the ELW estimator.

Table 1 shows the simulation results for $\rho = 0$. First, we discuss the estimates of β. The ELW estimator of β is very imprecise when $\delta_2 - \delta_1$ is small and appears to stay at a poor initial estimate of β; the ELW and tapered estimators of β have almost identical performance. The stationary LW estimator of β works well even when $\delta_2 - \delta_1$ is small. When $\delta_2 \geq 0.8$, the performance of the ELW estimator improves and becomes comparable to that of the stationary LW estimator.

We now focus on the estimates of δ. When $\delta_2 - \delta_1$ is small, the ELW estimator of δ is also affected by the poor estimates of β and has a slightly larger RMSE than the stationary LW estimator. Interestingly, the ELW estimator of δ performs better than the tapered estimator even when both have a similar RMSE with respect to β. The stationary LW estimator appears to be consistent even when $\delta_2 = 0.8$. When $\delta_2 = 1.3$, however, the stationary LW estimator of δ_2 converges to 1. This phenomenon is similar to the property of the univariate LW estimator.

Table 2 reports the results when $\rho = 0.4$. The presence of endogeneity improves the performance of all the estimators. This is analogous to the simulation results with bivariate LW estimation in Shimotsu (2007, Tables 2–4). Table 2 is comparable to Table 1 in some aspects: β is imprecisely estimated by the ELW estimator when $\delta_2 - \delta_1$ is small; the ELW estimator of δ is more efficient than the tapered estimator; the stationary LW estimator performs well even when $\delta_2 - \delta_1$ is small but becomes inconsistent when $\delta_2 > 1$. The ELW estimator of β performs poorly when $\delta_2 = 0.8$. We do not know the exact source of this problem, but it was probably caused by a few extremely large or small estimates. In a simulation result not reported here, imposing a bound on β, say $[-10, 10]$, reduced the RMSE substantially.

Table 3 reports the results with $\rho = 0.8$. Stronger endogeneity further improves the RMSE of the estimators. The overall picture is analogous to the case when $\rho = 0.4$. When $\delta_2 > 1$, the value of ρ affects the performance of the stationary LW estimator of δ_1: its RMSE deteriorates as ρ increases. Tables 1–3 report the performance of the ELW estimator of ρ in the sixth column. The ELW estimator is unbiased across all the values of ρ and δ_2.

In Table 4, we examine the performance of the estimators when δ_1 is large and hence $\delta_2 - \delta_1$ is small. We set $\delta_1 = 0.3$; as such, $\delta_2 - \delta_1 = 0.1, 0.5, 1.0$ when $\delta_2 = 0.4, 0.8, 1.3$, respectively. The value of ρ is set to 0.4. The results for the other values of ρ are qualitatively similar. Because $\delta_2 - \delta_1$ is smaller than in Table 2, from Theorem 2, we expect that the estimators of β perform worse than in Table 2 and that in contrast, the estimators of δ_1 and δ_2 are not affected significantly. When $\delta_2 = 0.4$ and hence $\delta_2 - \delta_1 = 0.1$, the estimators of β, including the stationary LW estimator, perform very poorly. When $\delta_2 = 0.8$, the performance of all the estimators of β improves, but the ELW and tapered estimators of β have a large MSE. The performance of the estimators of δ_1

and δ_2 is similar to that in Table 2.

Tables 1–4 show that the estimates of β have a large variance and RMSE when $\delta_2 - \delta_1$ is small. A close examination of the simulation results reveals that this large RMSE is caused by a small number of observations taking extremely large or small values. As such, we consider adding a penalty term $p(\beta, \beta_{NB}) = (\min\{0, \beta - \beta_{NB} + C\})^4 + (\max\{0, \beta - \beta_{NB} - C\})^4$ to the objective function of the ELW estimator $R^*(\vartheta)$, where β_{NB} is the narrow-band least squares (NBLS) estimator, and $C > 0$ is a constant. In effect, this penalization restricts β to the range $[\beta_{NB} - C, \beta_{NB} + C]$. Adding this penalty term (or imposing $\beta \in [\beta_{NB} - C, \beta_{NB} + C]$) does not invalidate the asymptotic results in Theorem 2 because β_{NB} is consistent. We set $C = 50$.

Table 5 reports the simulation results when the penalty term $p(\beta, \beta_{NB})$ is added to the objective function of the ELW, tapered, and stationary LW estimators. The simulation focuses on the case when $\delta_2 - \delta_1$ is small. We set $(\delta_1, \delta_2) = (0.1, 0.4), (0.3, 0.4), (0.3, 0.8)$ and $\rho = 0.4$ so that the results are comparable to the first panel of Table 2 and the first and second panels of Table 4. As can be seen, adding the penalty term improves the performance of the estimate of β substantially without affecting the estimates of δ_1 and δ_2 negatively. The ELW estimator of β rarely lies outside $[\beta_{NB}-50, \beta_{NB}+50]$; this is observed in 0.3%, 7.6%, and 0.0% of the replications when $(\delta_1, \delta_2) = (0.1, 0.4), (0.3, 0.4), (0.3, 0.8)$, respectively.

Fractionally integrated processes are often used to model financial time series. In such cases, the sensitivity of our semiparametric estimator to heavy-tailedness becomes a concern because many financial time series have heavy-tailed distribution reflecting the extent of outlier activity. We examine this issue by generating u_t from a bivariate t-distribution with parameter $(\Sigma, (0, 0)', 2)$. This is a multivariate extension of t-distribution with two degrees of freedom, and u_t has a finite mean but its variance is infinity. Table 6 reports the results for $\rho = 0.4$. The results for the other values ρ are similar and available from the author upon request. In most cases, neither the variance nor the MSE appears to increase.

6 Empirical application

As an empirical application, we revisit Bandi and Perron (2006, henceforth BP), who analyze the fractional cointegration relationship between monthly implied volatility and realized volatility of the S&P 100 index from January 1988 to October 2003. The regression model that BP estimate is

$$\sigma_t^R = \alpha + \beta \sigma_t^I + \epsilon_t, \tag{11}$$

where σ_t^R and σ_t^I are realized volatility and implied volatility, respectively, and ϵ_t is the residual term that includes the measurement error in implied volatility and a time-varying volatility risk premium. Note that ϵ_t may have long memory. Implied volatility is an unbiased forecast of future realized volatility if $\alpha = 0$ and $\beta = 1$. The regression model (11) can be expressed in terms of model (1) by defining $y_t = \sigma_t^R$ and $x_t = \sigma_t^I$, and adding a constant term to $y_t - \beta x_t$ and x_t. The dataset is constructed following BP. We use the S&P 500 index and the implied volatility of S&P 500 index options because S&P 500 options are more liquid than S&P 100 options. The sample period is from January 1990 to December 2009; the number of observations is 240. The data of implied volatility are the monthly observations of VIX (the CBOE Market Volatility index). As in BP, we use the closing value of each month and multiply the VIX data by $(252/365)^{1/2}$ to account for the difference between the numbers of trading days and calendar days in a year. The realized volatility of the S&P 500 index for each month is constructed by taking the average of the daily square return using the closing value of each day. Namely, $\sigma_t^R = (n_t^{-1} \sum_{j=1}^{n_t} r_j^2 \times 252)^{1/2}$, where $r_j = \log(S_j/S_{j-1})$, S_j is the closing value of the S&P 500 index on the jth trading day of month t, and n_t is the number of trading days in month t. See Section 1 of BP for more details.

BP use the NBLS estimator to estimate β; however, the NBLS estimator has different limiting distributions depending on whether $\delta_2 < 1/2$ (Christensen and Nielsen, 2006) or $\delta_2 > 1/2$ (Robinson and Marinucci, 2001). Further, when $\delta_2 < 1/2$, the asymptotic normality of the NBLS estimator is established only when $\delta_1 + \delta_2 < 1/2$ and $\rho = 0$. Consequently, BP use subsampling to construct asymptotic confidence intervals for β. However, subsampling confidence intervals depend on the size of subsamples, and the validity of subsampling is questionable when the asymptotic distribution theory is not available, namely when $\delta_2 < 1/2$ but $\delta_1 + \delta_2 > 1/2$ and/or $\rho \neq 0$. On the other hand, the proposed ELW estimator allows us to construct asymptotic confidence intervals for both β and (δ_1, δ_2) for any value of $(\delta_1, \delta_2) \in \Theta_\delta$ as long as $\delta_2 - \delta_1 < 1/2$. As we shall see below, this condition is satisfied in all the cases we consider.

Table 7 reports the descriptive statistics of the two volatility measures and corresponds to Table 1 of BP. The means are comparable to those in BP. The standard deviations, skewness, and kurtosis are higher than in BP because our sample includes the period of the recent financial crisis.

Table 8 reports the estimates of δ of implied volatility, realized volatility, and their differences using the univariate two-step ELW estimator of Shimotsu (2010a). This table corresponds to Table 2 of BP. The number of Fourier frequencies used is equal to the integer part of n^{α}, where $\alpha = 0.55, 0.6, 0.65, 0.7, 0.75$. The second row reports the estimates of δ for the implied volatility σ^I. Asymptotic 95% confidence intervals are reported in parentheses in the third row. The fourth and sixth rows report the δ estimates for realized volatility σ^R and volatility difference $\sigma^R - \sigma^I$. The estimates of δ are around 0.6 for implied volatility and around 0.55 for realized volatility. Overall, the two volatility series have similar estimates of δ, whereas the volatility difference $\sigma^R - \sigma^I$ has substantially smaller memory parameter estimates than both σ^I and σ^R, suggesting fractional cointegration between implied volatility and realized volatility. In general, our results are in accordance with those in Table 2 in BP, although in many cases, our estimates are larger than those in BP.

Table 9 reports the system ELW (ELW-FCI) estimates of $(\delta_1, \delta_2, \beta, \rho)$ defined by (10), and the NBLS estimates of β for the same values of m as in Table 8. This table corresponds to Table 7 in BP. For the tapered estimator, the NBLS estimator β_{NB} is used as the initial value for β, and the univariate two-step ELW estimators from $y_t - \beta_{NB}x_t$ and x_t are used as the initial values for δ_1 and δ_2, respectively. The same value of m is used in the computing of the NBLS estimator, tapered estimator, and ELW-FCI estimator. The ELW-FCI estimates are computed by repeating quasi-Newton updates from the tapered estimator until convergence. The figures in the parentheses report the confidence intervals for $(\delta_1, \delta_2, \beta)$ constructed using the asymptotic distribution in Theorem 2(a). The ELW-FCI estimate satisfies the condition of Theorem 2(a), i.e., $\nu = \delta_2 - \delta_1 < 0.5$, for all the cases.

The estimates of δ_1 and δ_2 are around 0.25 and around 0.65, respectively. The estimate of δ_1 increases as m increases, which may indicate a positive bias in the estimates of δ_1 from short-run dynamics. In many cases, the estimates of δ_1 are smaller than the δ estimates of $\sigma^R - \sigma^I$ in Table 8, whereas the estimates of δ_2 are similar to the δ estimates of σ^I in Table 8. For all m, the confidence intervals of δ_1 and δ_2 do not overlap with each other, which strongly suggests fractional cointegration between σ_t^R and σ_t^I. The point estimates of β are very close to one, and the hypothesis $\beta = 1$ is not rejected for all m. The estimates of ρ are positive and take values between 0.4 and 0.7. This suggests that $\rho > 0$ and that implied volatility and risk premium may be correlated even in the long-run. The last row reports the NBLS estimates of β. Reflecting $\rho > 0$, the NBLS estimates are upwardly biased for all m.

7 Appendix: Proof

In this and the following sections, C denotes a generic constant such that $C \in (1, \infty)$ unless specified otherwise; $E_{k\ell}$ denotes a 2×2 matrix whose (k, ℓ)th element is one and the other elements are zero; I_{xj} denotes $I_x(\lambda_j)$, w_{uj} denotes $w_u(\lambda_j)$, and similarly for other dft's and periodograms. Auxiliary lemmas and their proofs are collected in the supplementary appendix (Shimotsu (2010b)).

7.1 Proof of Theorem 1

The proof is divided into two parts. Part 1 shows $\hat{\delta} \to_p \delta_0$ and $\hat{\beta} - \beta_0 = O_p((m/n)^{\nu_0})$, which serves as a prerequisite for deriving the convergence rate in the theorem. Part 2 strengthens the convergence rate of part 1 to $\hat{\delta} - \delta_0 = O_p(m^{-1/2})$ and $\hat{\beta} - \beta_0 = O_p(m^{-1/2}(m/n)^{\nu_0})$.

7.1.1 Part 1: Proof of $\hat{\delta} \to_p \delta_0$ and $\hat{\beta} - \beta_0 = O_p((m/n)^{\nu_0})$

The proof closely follows the proof of Theorem 3 of Robinson (2008; henceforth R08). For any $c > 0$, define neighborhoods $\mathcal{N}_\beta(c) = \{\beta : |\beta - \beta_0| < c\}$ and $\mathcal{N}_\delta(c) = \{\delta : ||\delta - \delta_0||| < c\}$. Fix $\varepsilon > 0$ and define $\mathcal{N}(\varepsilon) = \mathcal{N}_\beta(\varepsilon^{-1}(m/n)^{\nu_0}) \times \mathcal{N}_\delta(\varepsilon)$, and $\bar{\mathcal{N}}(\varepsilon) = \Theta \backslash \mathcal{N}(\varepsilon)$. Define $\zeta_i = \delta_i - \delta_{0i}$. We split the parameter space Θ_δ into two. For a constant $0 < \Delta \leq 1/8$, define $\Theta_{\delta 1} = \{\delta \in \Theta_\delta : \zeta_1 \geq -1/2 + \Delta, \zeta_2 \geq -1/2 + \Delta\}$ and $\Theta_{\delta 2} = \Theta_\delta \backslash \Theta_{\delta 1}$. Since $\Pr(\hat{\vartheta} \in \bar{\mathcal{N}}(\varepsilon)) \leq \Pr(\inf_{\bar{\mathcal{N}}(\varepsilon)} \{R(\vartheta) - R(\vartheta_0)\} \leq 0)$, the consistency of $\hat{\vartheta}$ follows if we show

$$\Pr\left(\inf_{\bar{\mathcal{N}}(\varepsilon) \cap \{\Theta_\beta \times \Theta_{\delta 1}\}} \{R(\vartheta) - R(\vartheta_0)\} \leq 0 \right) \;\to\; 0, \quad \text{as } n \to \infty, \tag{12}$$

$$\Pr\left(\inf_{\bar{\mathcal{N}}(\varepsilon) \cap \{\Theta_\beta \times \Theta_{\delta 2}\}} \{R(\vartheta) - R(\vartheta_0)\} \leq 0 \right) \;\to\; 0, \quad \text{as } n \to \infty. \tag{13}$$

First, we show (12). As in equation (7.1) in R08, rewrite $R(\vartheta) - R(\vartheta_0)$ as

$$R(\vartheta) - R(\vartheta_0) = \log\det\{\hat{\Omega}^T(\vartheta)\hat{\Omega}^T(\vartheta_0)^{-1}\} - 2(\zeta_1 + \zeta_2)\frac{p}{(1-\kappa)m}\sum_{j(p,\kappa)}^{m} \log\lambda_j,$$

where we use $\hat{\Omega}^T(\vartheta)$ in place of $\hat{\Omega}(\theta)$ in R08. Define a vector type-II $I(\delta_{01}, \delta_{02})$ process that corresponds to a type-II version of u_t in R08 as

$$\xi_t = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} = B_0 z_t = \begin{bmatrix} (1-L)^{-\delta_{01}} u_{1t} I\{t \geq 1\} \\ (1-L)^{-\delta_{02}} u_{2t} I\{t \geq 1\} \end{bmatrix}, \quad B_0 = \begin{bmatrix} 1 & -\beta_0 \\ 0 & 1 \end{bmatrix}. \tag{14}$$

Define, analogously to R08 p. 2523, $H_j = (h_{k\ell j}) = \Psi(\lambda_j;\delta_0) I_{\xi j}^T \bar{\Psi}(\lambda_j;\delta_0)$, and $\hat{G}^{(1)}(\delta) = (\hat{g}_{k\ell}^{(1)})$, where $\hat{g}_{kk}^{(1)} = p(1-\kappa)^{-1} m^{-1} \sum_{j(p,\kappa)}^{m} (j/m)^{2\zeta_k} h_{kkj}$, and $\hat{g}_{12}^{(1)} = \hat{g}_{21}^{(1)} = p(1-\kappa)^{-1}(2m)^{-1} \sum_{j(p,\kappa)}^{m} (j/m)^{\zeta_1+\zeta_2} (e^{i(\pi-\lambda_j)(\zeta_2-\zeta_1)/2} h_{12j} + e^{-i(\pi-\lambda_j)(\zeta_2-\zeta_1)/2} h_{21j})$. Proceeding in the same manner as in Robinson (2008, p. 2523), we obtain

$$R(\vartheta) - R(\vartheta_0) = U_\delta(\delta) + U_\beta(\vartheta),$$

where

$$\begin{aligned} U_\delta(\delta) &= \log\det\{\Upsilon(\delta)\hat{G}^{(1)}(\delta)\Upsilon(\delta)\hat{G}^{(1)}(\delta_0)^{-1}\} + \phi_1(\delta,\kappa) + u(\delta) - \phi_2(\delta,\kappa), \\ U_\beta(\vartheta) &= \log\det\{\hat{\Omega}^{T*}(\vartheta)\hat{G}^{(1)}(\delta)^{-1}\} - \phi_1(\delta,\kappa) + \phi_2(\delta,\kappa), \end{aligned}$$

where $\Upsilon(\delta) = \mathrm{diag}((2\zeta_1+1)^{1/2}, (2\zeta_2+1)^{1/2})$, $\hat{\Omega}^{T*}(\vartheta) = \Xi(\vartheta)\hat{\Omega}^T(\vartheta)\Xi(\vartheta)$, $\Xi(\vartheta) = \mathrm{diag}(\lambda_m^{-\zeta_1}, \lambda_m^{-\zeta_2})$, $\phi_1(\delta,\kappa) = \log[(1-\kappa)^2(1-\kappa^{2\zeta_1+1})^{-1}(1-\kappa^{2\zeta_2+1})^{-1}]$, $\phi_2(\delta,\kappa) = 2(\zeta_1+\zeta_2)(1-\kappa)^{-1}\kappa\log\kappa$, $u(\delta) = \sum_{i=1}^{2}[2\zeta_i - \log(2\zeta_i+1) + 2\zeta_i(\log m - p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^{m}\log j - 1)]$. The functions $\phi_1(\delta,\kappa)$, and $\phi_2(\delta,\kappa)$ control the effect of taking summations from $[\kappa m]$; see Lemma 2(a) of Shimotsu (2010b). Other major differences from R08 are that (i) we define $\hat{G}^{(1)}(\delta)$ with the tapered periodograms and $p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^{m}$, and (ii) we use $U_\delta(\delta)$ instead of $U_\alpha(\alpha)$ in R08 because our model does not have the parameter γ.

Then (12) follows if we show that, as $n \to \infty$,

$$\Pr\left(\inf_{\bar{\mathcal{N}}_\delta(\varepsilon)\cap\Theta_{\delta 1}} U_\delta(\delta) \leq 0\right) \to 0, \tag{15}$$

$$\Pr\left(\inf_{\bar{\mathcal{N}}_\beta\left(\frac{1}{\varepsilon}\left(\frac{n}{m}\right)^{\nu_0}\right)\times\Theta_\delta} U_\beta(\vartheta) \leq 0\right) \to 0. \tag{16}$$

The proof of (15) is essentially the same as in R08. Define the population analogue of $\hat{g}_{k\ell}^{(1)}$ as $\hat{g}_{kk}^{(1)} = \omega_{kk}(1-\kappa)^{-1}\int_\kappa^1 x^{2\zeta_k}dx$ and $g_{12}^{(1)} = g_{21}^{(1)} = \omega_{12}(1-\kappa)^{-1}\int_\kappa^1 x^{\zeta_1+\zeta_2}dx\cos\tau$, where $\tau = (\zeta_2 - \zeta_1)\pi/2$. Then, (15) holds if

$$\begin{aligned}
\sup_{\Theta_{\delta 1}} \left\| \Upsilon(\delta)[\hat{G}^{(1)}(\delta) - G^{(1)}(\delta)]\Upsilon(\delta) \right\| &\to_p 0, & (17)\\
\sup_{\Theta_{\delta 1}} \left\| \{\Upsilon(\delta) G^{(1)}(\delta)\Upsilon(\delta)\}^{-1} \right\| &< \infty, & (18)\\
\inf_{\bar{\mathcal{N}}_\delta(\varepsilon)\cap\Theta_{\delta 1}} \{\log\det\{\Upsilon(\delta)G^{(1)}(\delta)\Upsilon(\delta)G^{(1)}(\delta_0)^{-1}\} + \phi_1(\vartheta,\kappa)\} &\geq 0, & (19)\\
\underline{\lim}_{n\to\infty} \inf_{\bar{\mathcal{N}}_\delta(\varepsilon)\cap\Theta_{\delta 1}} [u(\delta) - \phi_2(\delta,\kappa)] &> 0. & (20)
\end{aligned}$$

These conditions correspond to (7.5)-(7.8) of R08. (19) is weaker than (7.7) of R08 in that the inequality is not strict, but this does not affect (15) as long as (20) holds. The proof of (17) follows from using Lemma 1(b) of Shimotsu (2010b) in conjunction with the arguments in the proof of Theorem 1 of Robinson (1995). Use the summation by parts as in Robinson (1995) to deal with the uniformity, and approximate H_j by Ω_0 using Lemmas 1(b) and 2(c) of Shimotsu (2010b). For (18) and (19), direct calculation gives $\det\{\Upsilon(\delta)G^{(1)}(\delta)\Upsilon(\delta)\} = (2\zeta_1+1)(2\zeta_2+1)(1-\kappa)^{-2}\int_\kappa^1 x^{2\zeta_1}dx \int_\kappa^1 x^{2\zeta_1}dx[\omega_{11}\omega_{22} - \omega_{12}^2 c(\delta)\cos^2\tau]$, where $c(\delta) = (\int_\kappa^1 x^{\zeta_1+\zeta_2}dx)^2(\int_\kappa^1 x^{2\zeta_1}dx\int_\kappa^1 x^{2\zeta_1}dx)^{-1}$. Since $0 < c(\delta) \leq 1$ from the Cauchy-Schwartz inequality and $|\cos x| \leq 1$, the right hand side is no smaller than $(2\zeta_1+1)(2\zeta_2+1)(1-\kappa)^{-2}\int_\kappa^1 x^{2\zeta_1}dx\int_\kappa^1 x^{2\zeta_1}dx \det\Omega_0 > 0$, giving (18). (19) follows from $\log\det\{\Upsilon(\delta)G^{(1)}(\delta)\Upsilon(\delta)\} + \phi_1(\vartheta,\kappa) \geq \log\det\Omega_0$ and $G^{(1)}(\delta_0)^{-1} = \Omega_0^{-1}$. For (20), it follows from Lemma 2(a) of Shimotsu (2010b) that $u(\delta) - \phi_2(\delta,\kappa) = \sum_{i=1}^2[2\zeta_i - \log(2\zeta_i+1)] + O(m^{-1}\log m)$. The required result then follows because $\inf_{|x|>\varepsilon}\{x - \log(x+1)\} > \varepsilon^2/6$ (see (7.9) of R08). Therefore, we establish (15).

We proceed to show (16). Define $\hat{g}_{k\ell}^{(i)}$ similarly to R08 p. 2523 but using $p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^m$ and setting $\tau = (\zeta_2-\zeta_1)\pi/2$ and $\gamma_0 = (\delta_{02}-\delta_{01})\pi/2$. As in R08 p. 2524, define $\hat{a}_1 = (\hat{g}_{11}^{(2)}\hat{g}_{22}^{(1)} - 2\hat{g}_{12}^{(1)}\hat{g}_{12}^{(2)})/\det\{\hat{G}^{(1)}(\delta)\}$, and $\hat{a}_2 = (\hat{g}_{11}^{(3)}\hat{g}_{22}^{(1)} - \hat{g}_{12}^{(2)2})/\det\{\hat{G}^{(1)}(\delta)\}$. Define $g_{k\ell}^{(i)}$, the population counterpart of $\hat{g}_{k\ell}^{(i)}$, analogously to $g_{k\ell}^{(1)}$: for example, $g_{12}^{(2)} = g_{21}^{(2)} = (1-\kappa)^{-1}\omega_{22}\cos\gamma\int_\kappa^1 x^{\delta_1-\delta_{02}+\zeta_2}dx$, and $g_{11}^{(3)} = (1-\kappa)^{-1}\omega_{22}\int_\kappa^1 x^{2(\delta_1-\delta_{02})}dx$, where $\gamma = (\delta_2-\delta_1)\pi/2$. Using summation by parts and Lemma 1(b) of Shimotsu (2010b), we obtain $\sup_{\Theta_\delta}|\hat{g}_{k\ell}^{(i)} - g_{k\ell}^{(i)}| \to_p 0$ for $i = 1,2,3$, $k,\ell = 1,2$ as $n\to\infty$.

Rewrite $U_\beta(\vartheta) = \log Q(b_n(\beta)) - \phi_1(\delta,\kappa) + \phi_2(\delta,\kappa)$, where $Q(s) = 1 + \hat{a}_1 s + \hat{a}_2 s^2$ and $b_n(\beta) = \lambda_m^{-\nu_0}(\beta_0 - \beta)$. Define $a_1 = (g_{11}^{(2)}g_{22}^{(1)} - 2g_{12}^{(1)}g_{12}^{(2)})/\det\{G^{(1)}(\delta)\}$ and $a_2 = (g_{11}^{(3)}g_{22}^{(1)} - g_{12}^{(2)2})/\det\{G^{(1)}(\delta)\}$. Following R08 p. 2525, the probability in (16) is bounded by, with $\rho = \sup_{\Theta_\delta} |\phi_1(\delta,\kappa) - \phi_2(\delta,\kappa)| < \infty$,

$$\begin{aligned} & \Pr\left(\log\left\{1 - \sup_{\Theta_\delta} \frac{|\hat{a}_1|}{\varepsilon} + \inf_{\Theta_\delta} \frac{|\hat{a}_2|}{\varepsilon^2}\right\} \le \rho\right) + \Pr\left(\sup_{\Theta_\delta} \frac{|\hat{a}_1|}{2|\hat{a}_2|} > \frac{1}{\varepsilon}\right) \\ \le\ & 2\Pr\left(\sup_{\Theta_\delta} |\hat{a}_1 - a_1| + \frac{2}{\varepsilon} \sup_{\Theta_\delta} |\hat{a}_2 - a_2| + \varepsilon\rho \ge \frac{1}{\varepsilon} \inf_{\Theta_\delta} a_2 - \sup_{\Theta_\delta} |a_1|\right), \end{aligned} \tag{21}$$

which has an additional term $\varepsilon\rho$ compared with (7.13) of R08. Since $\sup_{\Theta_\delta} |\hat{a}_i - a_i| \to_p 0$ for $i = 1, 2$ as $n \to \infty$ and $\inf_{\Theta_\delta} \det\{G^{(1)}(\delta)\} > 0$, we have $\sup_{\Theta_\delta} |a_1| < \infty$. Because ε can be arbitrarily small, the probability in (21) tends to zero if $\inf_{\Theta_\delta} a_2 > 0$. Recall

$$g_{11}^{(3)} g_{22}^{(1)} - g_{12}^{(2)2} \ge (1-\kappa)^{-2} \omega_{22}^2 \left(\int_\kappa^1 x^{2(\delta_1 - \delta_{02})} dx \int_\kappa^1 x^{\zeta_2} dx - \left(\int_\kappa^1 x^{\delta_1 - \delta_{02} + \zeta_2} dx \right)^2 \right).$$

Note that $\zeta_2 = \delta_2 - \delta_{02}$. The right hand side is strictly positive because (i) $\int_\kappa^1 x^{2(\delta_1-\delta_{02})} dx \int_\kappa^1 x^{\zeta_2} dx - (\int_\kappa^1 x^{\delta_1 - \delta_{02} + \zeta_2} dx)^2 > 0$ if $\delta_1 \neq \delta_2$ from the Cauchy-Schwartz inequality, and (ii) $\delta_2 - \delta_1 \ge \eta_2 > 0$ in $\delta \in \Theta_\delta$. Consequently, we have $\inf_{\Theta_\delta} a_2 > 0$, and (16) follows.

It remains to show (13). Write $R(\vartheta) - R(\vartheta_0) = U_\delta^*(\delta) + U_\beta^*(\vartheta)$, where $U_\delta^*(\delta) = \log\det\{\Xi(\delta)\hat{G}^{(1)}(\delta)\Xi(\delta)\hat{G}^{(1)}(\delta_0)^{-1}\} - 2(\zeta_1+\zeta_2)p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^m \log\lambda_j$ and $U_\beta^*(\vartheta) = \log\det\{\hat{\Omega}^{T*}(\vartheta)\hat{G}^{(1)}(\delta)^{-1}\} = U_\beta(\vartheta) + \phi_1(\delta,\kappa) - \phi_2(\delta,\kappa)$. Then $\Pr(\inf_{\bar{\mathcal{N}}_\beta(\varepsilon^{-1}(n/m)^{\nu_0})\times\Theta_\delta} U_\beta^*(\vartheta) \le 0) \to 0$ follows from the proof of (16), so it suffices to show $\Pr(\inf_{\Theta_{\delta 2}} U_\delta^*(\delta) \le 0) \to 0$. Rewrite $U_\delta^*(\delta)$ as (see Shimotsu, 2007, p. 293)

$$U_\delta^*(\delta) = \log\det \hat{D}(\delta) - \log\det\hat{D}(\delta_0),$$

where

$$\hat{D}(\delta) = \frac{p}{(1-\kappa)m} \sum_{j(p,\kappa)}^m \begin{bmatrix} (j/q)^{2\zeta_1} h_{11j} & (j/q)^{\zeta_1+\zeta_2} \mathrm{Re}\{e^{i(\pi-\lambda_j)(\zeta_2-\zeta_1)/2} h_{12j}\} \\ (j/q)^{\zeta_1+\zeta_2} \mathrm{Re}\{e^{i(\pi-\lambda_j)(\zeta_2-\zeta_1)/2} h_{12j}\} & (j/q)^{2\zeta_2} h_{22j} \end{bmatrix},$$

and $q = \exp(p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^m \log j) \sim m/e^{1+(1-\kappa)^{-1}\kappa\log\kappa}$.

Define $K(\delta)$ as $\hat{D}(\delta)$ but $h_{k\ell j}$ is replaced with $\omega_{k\ell}$. Note that $\hat{D}(\delta)$ is identical to $\hat{D}_\kappa(d)$ in Shimotsu (2007, p. 294) except that $\{m^{-1}\sum_{j=[\kappa m]}^m, \theta_k, p, I_j\}$ in Shimotsu (2007) is replaced with $\{p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^m, \zeta_k, q, H_j\}$. Therefore, $\sup_{\Theta_{\delta 2}} |D(\delta) - K(\delta)| \to_p 0$ follows from using Lemma 1 of Shimotsu (2010b) and proceeding as in the proof of Theorem 1 of Shimotsu (2007, p. 294). Further, we can use the argument in Shimotsu (2007, pp. 294-95) to show that there exist $\varepsilon \in (0, 0.1)$ and $\kappa \in (0, 1/4)$ such that

$\inf_{\Theta_{\delta 2}} \det K(\delta) \geq (1+\varepsilon)\det G_0 + o(1)$. This is because Lemma 2 of Shimotsu (2007) and Lemma 5.5 of Shimotsu and Phillips (2005) hold even if $m^{-1}\sum_{j=[\kappa m]}^{m}$ and e are replaced with $p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^{m}$ and $e^{1+(1-\kappa)^{-1}\kappa\log\kappa}$ as long as κ is sufficiently small, since $\lim_{\kappa\to 0}\kappa\log\kappa = 0$. Therefore, $\det \hat{D}(\delta) \geq (1+\varepsilon)\det G_0 + o_p(1)$. Since $\det \hat{D}(\delta_0) = \det \hat{G}^{(1)}(\delta_0) = \det \Omega_0 + o_p(1)$ from (17), we establish (13). $\square$

7.1.2 Part 2: Proof of $\hat{\delta} - \delta_0 = O_p(m^{-1/2})$ and $\hat{\beta} - \beta_0 = O_p(m^{-1/2}(m/n)^{\nu_0})$

The proof closely follows the proof of Theorem 4 of R08. R08 uses the parameterization $\theta = (\beta, \gamma, \delta_1, \delta_2) = (\theta_1, \theta_2, \theta_3, \theta_4)$, whereas our parameterization is $\vartheta = (\beta, \delta_1, \delta_2)$. Because our parameterization implies $\gamma = (\delta_2 - \delta_1)\pi/2$, the derivatives $(\partial/\partial\beta)$, $(\partial/\partial\delta_1)$ and $(\partial/\partial\delta_2)$ in our model correspond to $(\partial/\partial\theta_1)$, $-\pi/2(\partial/\partial\theta_2) + (\partial/\partial\theta_3)$, and $\pi/2(\partial/\partial\theta_2) + (\partial/\partial\theta_4)$ in R08, respectively.

For the clarity of the proof, we reparameterize our objective function with the parameterization of R08, namely $\theta = (\beta, \gamma, \delta_1, \delta_2)$, and write $\hat{\Omega}^T$ as $\hat{\Omega}^T(\theta)$. Similar to R08, define $s^T(\theta) = (\partial/\partial\theta)R^T(\theta)$, $S^T(\theta) = (\partial/\partial\theta')s^T(\theta)$, $\Delta_n =$diag$\{\lambda_m^{-\nu_0}, 1, 1, 1\}$, and denote by $\tilde{S}^T$ the matrix $S^T(\theta)$ whose elements are evaluated at a point between θ_0 and $\hat{\theta}$. The required result follows if we show, for a finite matrix Σ_κ,

$$
\begin{aligned}
m^{1/2}\Delta_n^{-1}s^T(\theta_0) &= O_p(1), && (22)\\
\Delta_n^{-1}\tilde{S}^T\Delta_n^{-1} &\to_p \Sigma_\kappa. && (23)
\end{aligned}
$$

We show (22) first. The elements of $s^T(\theta)$ and $S^T(\theta)$ admit the same expression as equations (8.3) and (8.4) in R08 but in terms of $\hat{\Omega}^T(\theta)$ and its derivatives. Define $A_{0j}^T = \Psi(\lambda_j;\delta_0)B_0 I_{zj}^T B_0' \bar{\Psi}(\lambda_j;\delta_0)$. Define the score vectors $s_1^T(\theta_0), \dots, s_4^T(\theta_0)$ as $s_1(\theta), \dots, s_4(\theta)$ in R08 p. 2527 but replacing A_{0j} and $m^{-1}\sum_j$ in R08 with A_{0j}^T and $p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^{m}$.

First, we analyze the score vector. From Lemma 1(a) of Shimotsu (2010b), we obtain $m^{1/2}\Delta_n^{-1}s^T(\theta_0) = m^{1/2}\Delta_n^{-1}s^{T*}(\theta_0) + o_p(1)$ as in R08, where $s^{T*}(\theta_0)$ has kth element $s_k^{T*} = 2p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^{m}\mathrm{tr}(U_{Rkj}\mathrm{Re}\{I_{\varepsilon j}^T\} + U_{Ikj}\mathrm{Im}\{I_{\varepsilon j}^T\})$, where the coefficients U_{Rkj}, and U_{Ikj} are defined similarly to R08 p. 2527 by replacing $m^{-1}\sum_j$ and $I_{\varepsilon j}$ in R08 with $p(1-\kappa)^{-1}m^{-1}\sum_{j(p,\kappa)}^{m}$ and $I_{\varepsilon j}^T$. We do not provide the explicit formula for U_{Rkj}, and U_{Ikj} here. Because $Es^{T*}(\theta_0) = 0$, equation (22) follows if we show $\mathrm{var}(m^{1/2}\Delta_n^{-1}s^*(\theta_0)) = O(1)$. Note that $\lambda_j^{-\nu_0} \leq C([\kappa m]/m)^{-\nu_0} < \infty$ for all $j \geq [\kappa m]$. Consequently, $|U_{Rkj}|, |U_{Ikj}| \leq C\log m$ for $k = 1, 2, 3$, where $\log m$ term is from $U_{R,2+k,j}$. Therefore, in view of Lemma 2(b) of Shimotsu (2010b), $\mathrm{var}(m^{1/2}\Delta_n^{-1}s^{T*}(\theta_0))$ is bounded by $O(m^{-1}\sum_{j(p,\kappa)}^{m}\mathrm{var}(I_{\varepsilon j}^T)) +$

$O((\log m)^2 m^{-1}\sum_{j(p,\kappa),j\neq k}^m \sum_{k(p,\kappa)}^m \text{cov}(I_{\varepsilon j}^T, I_{\varepsilon k}^T))$. From Velasco (1999, p. 114), $\text{var}(I_{\varepsilon j}^T) = V + O(n^{-1})$, where V does not depend on j, and $\text{cov}(I_{\varepsilon j}^T, I_{\varepsilon k}^T) = O(|j-k|^{-2p} + |j+k|^{-2p} + n^{-1})$ for $j \neq k$. Therefore, $\text{var}(m^{1/2}\Delta_n^{-1} s^{T*}(\theta_0)) = O(1)$ follows, giving (22).

It remains to show (23). It suffices to show the following results, which correspond to (8.6) and (8.7) in R08:

$$\begin{aligned}\Delta_n^{-1}\{\tilde{S}^T - S^T(\theta_0)\}\Delta_n^{-1} &\to_p 0, \quad (24)\\ \frac{1}{2}\Delta_n^{-1} S^T(\theta_0)\Delta_n^{-1} &\to_p \Sigma_k. \quad (25)\end{aligned}$$

For (24), repeating the argument in R08 p. 2528 gives $(\log n)^C(\hat{\delta}_k - \delta_{0k}) \to_p 0$ for $k = 1, 2$ for any $C < \infty$. Then (24) follows from $m^\theta - 1 = O((\log n)^{-C+1})$ if $|\theta| \leq (\log n)^{-C}$.

The result (25) is obtained by following the argument of R08. Define $\Omega_0^{(k)}$ and $\Omega_0^{(k,\ell)}$ by replacing A_j in the definition of $A_j^{(k)}$ and $A_j^{(k,\ell)}$ in R08 with Ω_0. First, from Lemma 1 of Lobato and Velasco (2000) and Lemma 1 of Shimotsu (2010b), we have $\sum_{j(p,\kappa)}^s (A_{0j}^T - \Omega_0) = O_p(s^{\beta+1}n^{-\beta} + \log n + s^{1/2})$ for any $1 \leq s \leq m$. Combining it with Lemma 2(c) of Shimotsu (2010b) and summation by parts, we obtain $\sum_{j(p,\kappa)}^s (A_{0j}^{T(k,\ell)} - \Omega_0^{(k,\ell)}) = O_p(s^{\beta+1}n^{-\beta} + \log n + s^{1/2})$ at $\theta = \theta_0$, and (25) follows. □

7.2 Proof of Theorem 2 (a)

The required result follows if we show, for any $\bar{\vartheta}$ such that $\bar{\vartheta} - \vartheta_0 = O_p(m^{-1/2})$,

$$m^{1/2}(\Delta_n^*)^{-1}(\partial/\partial\vartheta)R^*(\vartheta_0) \to_d N(0, \Xi), \text{ and } (\Delta_n^*)^{-1}(\partial^2/\partial\vartheta\partial\vartheta')R^*(\bar{\vartheta})(\Delta_n^*)^{-1} \to_p \Xi. \quad (26)$$

7.2.1 Score vector with respect to δ

From the proof of Theorem 4 of R08, $(\partial/\partial\delta_k)R^*(\vartheta_0)$ satisfies the first result in (26) if

$$\begin{aligned}(\partial/\partial\delta_k)R^*(\vartheta_0) &= \text{tr}((\partial/\partial\delta_k)\tilde{\Omega}^*(\vartheta_0)\tilde{\Omega}^*(\vartheta_0)^{-1}) - 2m^{-1}\sum_{j=1}^m \log\lambda_j \\ &= s_1^* + (-1)^k(\pi/2)s_2^* + o_p(m^{-1/2}), \quad (27)\end{aligned}$$

where s_1^* and s_2^* are defined in (8.5) of R08. Define $w_{\Delta 1j} = (w_{\log(1-L)u_1 j}, w_{u_2 j})'$, then $(\partial/\partial\delta_1)\tilde{\Omega}^*(\vartheta_0) = E_{11}m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta 1j}\bar{w}_{uj}\} + m^{-1}\sum_{j=1}^m \text{Re}\{w_{uj}\bar{w}_{\Delta 1j}\}E_{11}$. Because $w_{\Delta 1j}$ is premultiplied by E_{11}, we only need to analyze the first row of $m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta 1j}\bar{w}_{uj}\}$. Define the $(1,1)$th element of $m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta 1j}\bar{w}_{uj}\}$ as $s_1 = m^{-1}\sum_{j=1}^m w_{\log(1-L)u_1 j}w_{u_1 j}$.

Observe that s_1 is identical to 1/2 times $\hat{G}_1(d_0)$ that is defined on p. 1912 of Shimotsu and Phillips (2005; henceforth SP). SP derive the limit of $m^{1/2}\hat{G}_1(d_0)$ on pp. 1916-18. Because their argument uses only Lemmas 5.8 and 5.9 of SP, we can obtain a matrix-valued version of SP, pp. 1916-18 using Lemma 5 of Shimotsu (2010b). Specifically, the following matrix-valued version of line 3, p. 1917 of SP holds for our model:

$$-w_{\text{diag}\{\log(1-L)\}uj}\bar{w}_{uj} = \text{diag}\{J_n(e^{i\lambda_j})\}I_{uj} - (2\pi n)^{-1/2}\text{diag}\{\tilde{J}_{n\lambda_j}(e^{-i\lambda_j}L)\}A(0)\varepsilon_n\bar{w}_{\varepsilon j}\bar{A}(\lambda_j) + R_{nj},$$

where R_{nj} has the same order of magnitude as specified in SP. Taking its average over $j = 1, \ldots, m$ and repeating SP pp. 1916-8, we obtain $-m^{-1}\sum_{j=1}^m w_{\log(1-L)uj}\bar{w}_{uj} = m^{-1}\sum_{j=1}^m \text{diag}\{J_n(e^{i\lambda_j})\}I_{uj} + o_p(m^{-1/2})$. Using the approximation of $J_n(e^{i\lambda_j})$ at the end of Lemma 5 of Shimotsu (2010b) gives $(\partial/\partial\delta_1)\tilde{\Omega}^*(\vartheta_0) = E_{11}\Phi + \Phi E_{11} + o_p(m^{-1/2})$, where $\Phi = m^{-1}\sum_{j=1}^m \log\lambda_j \text{Re}\{I_{uj}\} - m^{-1}\sum_{j=1}^m (\lambda_j - \pi)\text{Im}\{I_{uj}\}/2$. Approximating I_{uj} by $A(\lambda_j)I_{\varepsilon j}\bar{A}(\lambda_j)$ by Lemma 1(b1) of Shimotsu (2007) and then by $PI_{\varepsilon j}P$ using Assumptions 1-3, we obtain $\Phi = m^{-1}\sum_{j=1}^m \log\lambda_j P\text{Re}\{I_{\varepsilon j}\}P' + (\pi/2)m^{-1}\sum_{j=1}^m P\text{Im}\{I_{\varepsilon j}\}P' + o_p(m^{-1/2})$, in which the terms with $\lambda_j\text{Im}\{I_{\varepsilon j}\}$ reduce to $o_p(m^{-1/2})$ by $E\text{Im}\{I_{\varepsilon j}\} = 0$, $\sum_{j=1}^r (I_{\varepsilon j} - I_2) = O(r^{1/2})$ (Lobato, 1999, p. 145, (C.3)), and summation by parts. It follows that

$$\text{tr}\left(\frac{\partial}{\partial\delta_1}\tilde{\Omega}^*(\vartheta_0)\Omega_0^{-1}\right) = \frac{2}{m}\sum_{j=1}^m \log\lambda_j \text{tr}(E_{11}P\text{Re}\{I_{\varepsilon j}\}P'\Omega_0^{-1}) - \frac{\pi}{2}\frac{2}{m}\sum_{j=1}^m \text{tr}(E_{22}P\text{Im}\{I_{\varepsilon j}\}P'\Omega_0^{-1}),$$

where the negative sign appears because $\text{tr}(E_{11}\text{Im}\{I_{\varepsilon j}\}) = -\text{tr}(E_{22}\text{Im}\{I_{\varepsilon j}\})$. Using the fact that $\tilde{\Omega}^*(\vartheta_0) = m^{-1}\sum_{j=1}^m P\text{Re}\{I_{\varepsilon j}\}P' + o_p(m^{-1/2})$ and evaluating $(\partial/\partial\delta_2)\tilde{\Omega}^*(\vartheta_0)$ similarly, we find

$$\begin{aligned}(\partial/\partial\delta_k)R^*(\vartheta_0) &= \frac{2}{m}\sum_{j=1}^m \text{tr}\left\{\left(\log j - \frac{1}{m}\sum_{j=1}^m \log j\right)P'\Omega_0^{-1}E_{kk}P\text{Re}\{I_{\varepsilon j}\}\right\} \\ &\quad + (-1)^k\frac{\pi}{2}\frac{2}{m}\sum_{j=1}^m \text{tr}\left\{P'\Omega_0^{-1}E_{22}P\text{Im}\{I_{\varepsilon j}\}\right\} + o_p(m^{-1/2}).\end{aligned}$$

The first two terms on the right are the same as s_1^* and s_2^*. Hence, we establish (27).

7.2.2 Score vector with respect to β

From the proof of Theorem 4 of R08, $(\partial/\partial\beta)R^*(\vartheta_0)$ satisfies the first result in (26) if

$$(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0) = s_1(\theta_0) + o_p(\lambda_m^{-\nu_0}m^{-1/2}), \tag{28}$$

where $s_1(\theta_0)$ is defined similarly to p. 2527 of R08 but A_{0j} in R08 is replaced with I_{uj}.

Define $w_{\Delta 2j} = (0, w_{\Delta^{\delta_1} u_{2j}})'$ so that $E_{12} w_{\Delta 2j} = (w_{\Delta^{\delta_1} u_{2j}}, 0)'$. Then

$$\frac{\partial}{\partial \beta}\tilde{\Omega}^*(\vartheta) = -E_{12}\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\{w_{\Delta 2j}\bar{w}_{\Delta^\delta z}(\lambda_j;\beta)\} - \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\{w_{\Delta^\delta z}(\lambda_j;\beta)\bar{w}_{\Delta 2j}\}E_{21}. \quad (29)$$

Define $w_{\Delta 2j}$ evaluated at ϑ_0 as $w_{\Delta^0 2j} = (0, w_{\Delta^{-\nu_0} u_{2j}})'$, then we have $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0) = -E_{12} m^{-1}\sum_{j=1}^m \mathrm{Re}\{w_{\Delta^0 2j}\bar{w}_{uj}\} - m^{-1}\sum_{j=1}^m \mathrm{Re}\{w_{uj}\bar{w}_{\Delta^0 2j}\}E_{21}$.

We approximate $w_{\Delta^{-\nu_0} u_{2j}}\bar{w}_{uj}$ by $e^{i\pi\nu_0/2}\lambda_j^{-\nu_0} w_{u_{2j}}\bar{w}_{uj}$ by applying the results from Phillips and Shimotsu (2004) and Shimotsu and Phillips (2006, henceforth SP06). First, replace $(X_t - X_0, u_t)$ in equation (26) of SP06 with $(\Delta^{-\nu_0}u_{2t}, u_{2t})$ to obtain $w_{\Delta^{-\nu_0} u_{2j}} = D_n(e^{i\lambda_j}; -\nu_0) w_{u_{2j}} - (2\pi n)^{-1/2}\lambda_j^{\nu_0}\tilde{U}_{2,\lambda_j n}(-\nu_0)$. Define $D_{nj}(\nu_0) = \lambda_j^{\nu_0} D_n(e^{i\lambda_j}; -\nu_0)$ as on p. 226 of SP06. Then

$$\begin{aligned} &\lambda_m^{\nu_0}\frac{1}{m}\sum_{j=1}^{m} w_{\Delta^{-\nu_0} u_{2j}}\bar{w}_{u_{2j}} \\ = \ &\frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{-\nu_0} D_{nj}(\nu_0) I_{u_{2j}} + \frac{1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{-\nu_0}\frac{1}{\sqrt{2\pi n}}\lambda_j^{\nu_0}\tilde{U}_{2,\lambda_j n}(-\nu_0)\,\bar{w}_{u_{2j}}. \quad (30)\end{aligned}$$

Since $D_{nj}(\nu_0) = e^{i\pi\nu_0/2} + O(\lambda_j) + O(j^{-\nu_0-1})$ from (27) of SP06, we can write the first term in (30) as $e^{i\pi\nu_0/2} m^{-1}\sum_{j=1}^m (j/m)^{-\nu_0} I_{u_{2j}} + o_p(m^{-1/2})$, by approximating I_{uj} by $A(\lambda_j) I_{\varepsilon j}\bar{A}(\lambda_j)$ first (see Lemma 1(b1) of Shimotsu (2007)) and then using the order of the covariance between $I_{\varepsilon j}$ and $I_{\varepsilon k}$.

For the second term in (30), note that a vector version of Lemma A.5(b) of Phillips and Shimotsu (2004) holds for a vector process u_t. Namely, $E||\tilde{U}_{\lambda_j n}(-\nu_0) - A(0)\tilde{\varepsilon}_{\lambda_j n}(-\nu_0)||^2 = O(n^{1+2\nu_0} s^{-2\nu_0-1}(\log n)^{-4} + n^{1+2\nu_0}s^{-2})$, where $\tilde{U}_{\lambda n}(\nu_0)$ and $\tilde{\varepsilon}_{\lambda n}(\nu_0)$ are 2×1 and defined exactly in same manner as in Phillips and Shimotsu (2004, p. 667). Combining it with Lemma 3(a) of Shimotsu (2010b) and the order of $\tilde{U}_{\lambda_j n}(-\nu_0)$, we may write the second term in (30) as $U_n + o_p(m^{-1/2})$, where $U_n = m^{-1}\sum_{j=1}^m (j/m)^{-\nu_0}(2\pi n)^{-1/2}\lambda_j^{\nu_0} A_2(0)\tilde{\varepsilon}_{\lambda_j n}(-\nu_0)\,\bar{w}_{\varepsilon j}\bar{A}_2(\lambda_j)$, where $A_2(\lambda)$ denotes the second row of $A(\lambda)$. Observe that mU_n is closely related to T_n' that is defined on p. 231 of SP06. The major differences between mU_n and T_n' are that U_n is constructed from vector-valued ε_t, and (d_0, ν_j) in T_n' corresponds to $(\nu_0, (j/m)^{-\nu_0})$ in mU_n. Observe that, if we replace ν_j in T_n' with $(j/m)^{-\nu_0}$ in the derivations on pp. 231-32 of SP06, the order of T_n' does not change and $E|T_n'|^2 = o(m)$ still holds. Therefore, $U_n = o_p(m^{-1/2})$. A similar analysis applies to $\lambda_m^{\nu_0} m^{-1}\sum_{j=1}^m w_{\Delta^{-\nu_0} u_{2j}}\bar{w}_{u_1 j}$, and, consequently, $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0) = -m^{-1}\sum_{j=1}^m \lambda_j^{-\nu_0}(E_{12}I_{uj}e^{i\pi\nu_0/2}+I_{uj}E_{21}e^{-i\pi\nu_0/2})+o_p(\lambda_m^{-\nu_0}m^{-1/2})$, giving (28).

7.2.3 Hessian approximation

We prove the required result by approximating the derivatives of $\tilde{\Omega}^*(\vartheta_0) = m^{-1}\sum_{j=1}^m \mathrm{Re}\{I_{\Delta^\delta z}(\lambda_j;\beta)\}$ by the derivatives of the counterpart in R08 p. 2512, $\hat{\Omega}(\theta_0) = m^{-1}\sum_{j=1}^m \mathrm{Re}\{A(\lambda_j;\theta)\}$. First, replace A_j in the definition of $A_j^{(k)}$ and $A_j^{(k,\ell)}$ in R08 p. 2527 with I_{uj}. This does not change the limit of the derivatives of $\hat{\Omega}(\theta_0)$, because both A_{0j} and I_{uj} are approximated by $PI_{\varepsilon j}P'$. Define $I_j^{(k)}$ and $I_j^{(k,\ell)}$ similarly to $A_j^{(k)}$ and $A_j^{(k,\ell)}$in R08 but using I_{uj} in place of A_{0j}. We proceed to show that the derivatives of $I_{\Delta^\delta z}(\lambda_j;\beta)$ at $\vartheta = \vartheta_0$ are equal to linear combinations of $I_j^{(k)}$ and $I_j^{(k,\ell)}$ up to an negligible term. First, $w_{\Delta^{\delta_0}z}(\lambda_j;\beta_0) = w_{uj}$ from the definition. Second, for the derivative with respect to δ_k, it follows from the derivation of the score approximation above that, for $k=1,2$, $(\partial/\partial\delta_k)\tilde{\Omega}(\vartheta_0) = [\partial/\partial\theta_{2+k} + (-1)^k(\pi/2)(\partial/\partial\theta_2)]\hat{\Omega}(\theta_0) + o_p((\log n)^{-C})$. For the derivative with respect to β, (28) implies $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0) = (\partial/\partial\beta)\Omega^*(\theta_0) + o_p(\lambda_m^{-\nu_0}(\log n)^{-C})$. Similarly, we can use Lemma 4 of Shimotsu (2010b) to express the other derivatives of $\tilde{\Omega}^*(\vartheta_0)$ in terms of the derivatives of $\hat{\Omega}(\theta)$ in R08 such as $(\partial^2/\partial\delta_k\partial\beta)\tilde{\Omega}^*(\vartheta_0) = [\partial/\partial\theta_{2+k} + (-1)^k(\pi/2)(\partial/\partial\theta_2)](\partial/\partial\beta)\hat{\Omega}(\theta_0) + o_p(\lambda_m^{-\nu_0}(\log n)^{-C})$. We suppress the obvious formula for $(\partial^2/\partial\delta_k\partial\delta_\ell)\tilde{\Omega}^*(\vartheta_0)$. Therefore, it follows from the relation between ϑ and θ that $(\Delta_n^*)^{-1}(\partial^2/\partial\vartheta\partial\vartheta')R^*(\bar{\vartheta})(\Delta_n^*)^{-1} \to_p \Xi$.

Finally, the proof of $(\Delta_n^*)^{-1}[(\partial^2/\partial\vartheta\partial\vartheta')R^*(\bar{\vartheta}) - (\partial^2/\partial\vartheta\partial\vartheta')R^*(\vartheta_0)](\Delta_n^*)^{-1} = o_p(1)$ follows from the root–m consistency of $\bar{\vartheta}$ and $\lambda_j^\alpha - 1 = O(\log n)$ for any finite α. □

7.3 Proof of Theorem 2 (b)

Define Δ_n^{**} =diag$(n^{\nu_0}m^{-1/2},1,1)$. The stated result follows if we show (i) $m^{1/2}(\Delta_n^{**})^{-1}(\partial/\partial\vartheta)R^*(\vartheta_0) = (r_1,r_2,r_3)$, where $r_1 = O_p(1)$, $(r_2,r_3)' \to_d N(0,\Xi_\delta)$,

$$(ii) \quad (\Delta_n^{**})^{-1}(\partial^2/\partial\vartheta\partial\vartheta')R^*(\vartheta_0)(\Delta_n^{**})^{-1} = H_n = \begin{bmatrix} H_{11,n} & H_{12,n} \\ H_{21,n} & H_{22,n} \end{bmatrix},$$

where $H_{11,n}$ is (1×1), $H_{22,n}$ is (2×2), and $H_{k\ell,n}$ satisfies

$$H_{11,n} = O_p(1), \quad \Pr(|H_{11,n}| < \varepsilon) \to 0 \text{ as } \varepsilon \to 0, \quad H_{12,n}, H_{21,n} \to_p 0, \quad H_{22} \to_p \Xi_\delta, \quad (31)$$

and (iii) $(\Delta_n^{**})^{-1}(\partial^2/\partial\vartheta\partial\vartheta')[R^*(\bar\vartheta)-R^*(\vartheta_0)](\Delta_n^{**})^{-1}\to_p 0$, for any $\bar\vartheta$ such that $\Delta_n^*(\bar\vartheta-\vartheta)=O_p(m^{-1/2})$. We omit a tedious but straightforward proof of $(\Delta_n^{**})^{-1}(\partial^2/\partial\vartheta\partial\vartheta')[R^*(\bar\vartheta)-R^*(\vartheta_0)](\Delta_n^{**})^{-1}\to_p 0$.

7.3.1 Score vector approximation

The score with respect to δ_k remains unchanged, because it does not depend on the value of ν_0. We need only to analyze the score with respect to β. The required result is

$$(\partial/\partial\beta)R^*(\vartheta_0)=O_p(n^{\nu_0}m^{-1}). \tag{32}$$

First, as in the proof for $\nu_0\in(0,1/2)$, we have $(\partial/\partial\beta)\tilde\Omega^*(\vartheta_0)=-E_{12}m^{-1}\sum_{j=1}^m \mathrm{Re}\{w_{\Delta^{\nu_0}2j}\bar w_{uj}\}-m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{uj}\bar w_{\Delta^{\nu_0}2j}\}E_{21}$, hence $(\partial/\partial\beta)R^*(\vartheta_0)=-2\mathrm{tr}[E_{12}m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{\Delta^{\nu_0}2j}\bar w_{uj}\}\tilde\Omega^*(\vartheta_0)^{-1}]$. Define $c_n=m^{-1}\sum_{j=1}^m\mathrm{Re}\{(1-e^{i\lambda_j})^{-\nu_0}\}$, which satisfies $c_n=O(n^{\nu_0}m^{-\nu_0}+n^{\nu_0}m^{-1}\log m)=O(n^{\nu_0}m^{-1/2-\eta})$ for some $\eta>0$. Using $\mathrm{tr}(E_{12})=0$, rewrite $(\partial/\partial\beta)R^*(\vartheta_0)$ further as $-2\mathrm{tr}[E_{12}m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{\Delta^{\nu_0}2j}\bar w_{uj}\}(\tilde\Omega^*(\vartheta_0)^{-1}-\Omega_0^{-1})]-2\mathrm{tr}[E_{12}(m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{\Delta^{\nu_0}2j}\bar w_{uj}\}-c_n\Omega_0)\Omega_0^{-1}]$. Then (32) follows if we show

$$E_{12}\left(m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{\Delta^{\nu_0}2j}\bar w_{uj}\}-c_n\Omega_0\right)=O_p(n^{\nu_0}m^{-1}). \tag{33}$$

We prove (33) only for $m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{\Delta^{-\nu_0}u_2j}\bar w_{u_2j}\}-c_n\omega_{22}$. The other elements are analyzed similarly. First, it follows from Lemma A.1 of Phillips and Shimotsu (2004) that

$$(1-e^{i\lambda_j})w_{\Delta^{-\nu_0}u_2j}=D_n(e^{\lambda_j};1-\nu_0)w_{u_2j}-(2\pi n)^{-1/2}\widetilde U_{2,\lambda_j n}(1-\nu_0)-(2\pi n)^{-1/2}e^{i\lambda_j}Y_n, \tag{34}$$

where $Y_n=\Delta^{-\nu_0}u_{2n}I\{t\ge 1\}$. Hence, $m^{-1}\sum_{j=1}^m\mathrm{Re}\{w_{\Delta^{-\nu_0}u_2j}\bar w_{u_2j}\}-c_n\omega_{22}$ is written as $\mathrm{Re}\{T_{1n}+T_{2n}+T_{3n}\}$, where

$$\begin{aligned}
T_{1n} &= m^{-1}\sum_{j=1}^m(1-e^{i\lambda_j})^{-1}D_n(e^{\lambda_j};1-\nu_0)I_{u_2j}-c_n\omega_{22},\\
T_{2n} &= m^{-1}\sum_{j=1}^m(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}\widetilde U_{2,\lambda_j n}(1-\nu_0)\bar w_{u_2j},\\
T_{3n} &= m^{-1}\sum_{j=1}^m(1-e^{i\lambda_j})^{-1}(2\pi n)^{-1/2}e^{i\lambda_j}Y_n\bar w_{u_2j}.
\end{aligned}$$

For $\mathrm{Re}\{T_{1n}\}$, noting that $D_n(e^{\lambda_j}; 1-\nu_0) = (1-e^{i\lambda_j})^{1-\nu_0} + (n^{\nu_0 - 1} j^{-1})$ from Lemma A.2 of Phillips and Shimotsu (2004) and using the definition of c_n, we have $\mathrm{Re}\{T_{1n}\} = m^{-1}\sum_{j=1}^m \mathrm{Re}\{(1-e^{i\lambda_j})^{-\nu_0}\}(I_{u_2 j} - \omega_{22}) + O_p(n^{\nu_0} m^{-1})$.

Therefore, using Lemma 3(b) of Shimotsu (2010b), summation by parts, and $(1-e^{i\lambda_j})^{-\nu_0} - (1-e^{i\lambda_{j+1}})^{-\nu_0} = O(n^{\nu_0} j^{-\nu_0 - 1})$, and noting $\nu_0 > 1/2$, we obtain $\mathrm{Re}\{T_{1n}\} = m^{-1}\sum_{j=1}^m \mathrm{Re}\{(1-e^{i\lambda_j})^{-\nu_0}\}(f_{22}(\lambda_j) - \omega_{22}) + O_p(n^{\nu_0} m^{-1})$. The first term on the right has the order $m^{-1}\sum_{j=1}^m j^{-\nu_0+b} n^{\nu_0 - b} = n^{\nu_0} m^{-1}(n^{-b}\sum_{j=1}^m j^{-\nu_0+b})$. When $\nu_0 \geq 1$, $(n^{-b}\sum_{j=1}^m j^{-\nu_0+b})$ is $o(1)$. When $\nu_0 \in (1/2, 1)$, it is $O(n^{-b} m^{1-\nu_0+b})$, thus $\mathrm{Re}\{T_{1n}\} = O_p(n^{\nu_0} m^{-1})$.

It remains to show the order of T_{2n} and T_{3n}. For T_{2n}, it easily follows from the order of $\widetilde{U}_{2,\lambda_j n}(1-\nu_0)$ provided in Lemma A.5 of Phillips and Shimotsu (2004) that $T_{2n} = O_p(m^{-1}\sum_{j=1}^m j^{-1} n^{1/2} n^{\nu_0 - 1/2} j^{1/2-\nu_0}) = O_p(n^{\nu_0} m^{-1})$. For T_{3n}, it follows from Lemma 3(a) of Shimotsu (2010b) that $T_{3n} = m^{-1} n^{-1/2} Y_{3n}(\sum_{j=1}^m (1-e^{i\lambda_j})^{-1} e^{i\lambda_j} \bar{w}_{\varepsilon j} \bar{A}_2(\lambda_j) + O_p(n))$. This is $O_p(n^{\nu_0} m^{-1})$ because $w_{\varepsilon j}$ and $w_{\varepsilon k}$ are uncorrelated for $j \neq k$ and $n^{1/2-\nu_0} Y_{3n} \to_d N(0, \sigma^2)$ with $\sigma^2 < \infty$ from Lemma A.5(a2) of Phillips and Shimotsu (2004) and a standard MDS-CLT. This establishes (33).

7.3.2 Hessian approximation

Again, we need only to consider the terms involving the derivatives with respect to β. Note that

$$\begin{aligned}
\frac{\partial^2 R^*(\vartheta)}{\partial\beta\partial\delta_k} &= \mathrm{tr}\left\{ \frac{\partial^2 \tilde{\Omega}^*(\vartheta)}{\partial\beta\partial\delta_k}\tilde{\Omega}^*(\vartheta)^{-1} - \frac{\partial\tilde{\Omega}^*(\vartheta)}{\partial\beta}\tilde{\Omega}^*(\vartheta)^{-1}\frac{\partial\tilde{\Omega}^*(\vartheta)}{\partial\delta_k}\tilde{\Omega}^*(\vartheta)^{-1} \right\}, \\
\frac{\partial^2 R^*(\vartheta)}{\partial\beta^2} &= \mathrm{tr}\left\{ \frac{\partial^2 \tilde{\Omega}^*(\vartheta)}{\partial\beta^2}\tilde{\Omega}^*(\vartheta)^{-1} - \frac{\partial\tilde{\Omega}^*(\vartheta)}{\partial\beta}\tilde{\Omega}^*(\vartheta)^{-1}\frac{\partial\tilde{\Omega}^*(\vartheta)}{\partial\beta}\tilde{\Omega}^*(\vartheta)^{-1} \right\}.
\end{aligned}$$

The required result follows if

$$(\partial^2/\partial\beta\partial\delta_k) R^*(\vartheta_0) = o_p(n^{\nu_0} m^{-1/2}), \tag{35}$$

and $n^{-2\nu_0} m(\partial^2/\partial\beta^2) R^*(\vartheta_0)$ satisfies the condition (31) of $H_{11,n}$. We analyze $(\partial^2/\partial\beta\partial\delta_k) R^*(\vartheta_0)$ first. $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0)\tilde{\Omega}^*(\vartheta_0)^{-1} = o_p(n^{\nu_0} m^{-1/2})$ holds because $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0) = O(c_n) + O_p(n^{\nu_0} m^{-1}) = O_p(n^{\nu_0} m^{-1/2-\eta})$ from (33) and $\tilde{\Omega}^*(\vartheta_0)^{-1} = O_p(1)$. Note that $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta)$ consists of $w_{\Delta 2j}$ and $w_{\Delta^\delta z}(\lambda_j; \beta)$, and that $(\partial^2/\partial\beta\partial\delta_k)\tilde{\Omega}^*(\vartheta)$ consists of $w_{\Delta 2j}$ and $w_{\Delta^\delta z}(\lambda_j; \beta)$ and their derivatives with respect to δ_k. From Lemma 4 of Shimotsu (2010b), the leading term of these derivatives is $(\log n)^s$, $s = 1, 2$, times $w_{\Delta 2j}$ and $w_{\Delta^\delta z}(\lambda_j; \beta)$. Therefore, the order of magnitude of $(\partial^2/\partial\beta\partial\delta_k)\tilde{\Omega}^*(\vartheta)$ is no larger than $(\log n)^2$ times that of $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta)$. Thus, $(\partial^2/\partial\beta\partial\delta_k)\tilde{\Omega}^*(\vartheta_0) = o_p(n^{\nu_0}m^{-1/2})$ follows, and (35) follows.

The proof completes by showing the behavior of $n^{-2\nu_0}m(\partial^2/\partial\beta^2)R^*(\vartheta_0)$. Taking a derivative of (29) gives $(\partial^2/\partial\beta^2)\tilde{\Omega}^*(\vartheta_0) = 2E_{12}m^{-1}\sum_{j=1}^m \mathrm{Re}\{w_{\Delta^0 2j}\bar{w}_{\Delta^0 2j}\}E_{21}$. The only non-zero element of this matrix is its (1,1)th element, $2m^{-1}\sum_{j=1}^m I_{\Delta^{-\nu_0}u_2 j}$. From Theorems 4.5 and 5.1 of Robinson and Marinucci (2001), we have $\lim_{n\to\infty} E[n^{-2\nu_0}\sum_{j=1}^m I_{\Delta^{-\nu_0}u_2 j}] < \infty$ and $\lim_{n\to\infty}\mathrm{var}[n^{-2\nu_0}\sum_{j=1}^m I_{\Delta^{-\nu_0}u_2 j}] = V \in (0,\infty)$. The required result then follows from $(\partial/\partial\beta)\tilde{\Omega}^*(\vartheta_0)\tilde{\Omega}^*(\vartheta_0)^{-1} = o_p(n^{\nu_0}m^{-1/2})$ and $\tilde{\Omega}^*(\vartheta_0) \to_p \Omega_0$. □

References

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

Baillie, R. T. and Kapetanios, G. (2009). Semi parametric estimation of long memory: comparisons and some attractive alternatives. Paper presented at the conference at Singapore Management University in honor of Peter Phillips' sixtieth birthday.

Bandi, F. and Perron, B. (2006). Long memory and the relation between implied and realized volatility. Journal of Financial Econometrics 4, 636-670.

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

Cassola, N. and Morana, C. (2010). Comovements in volatility in the euro money market. *Journal of International Money and Finance*, 29, 525-539.

Chen, W. W. and C. M. Hurvich (2002). Semiparametric estimation of multivariate fractional cointegration. Journal of the American Statistical Association 98, 629-642.

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

Christensen, B. J. and Nielsen, M. Ø. (2006). Semiparametric analysis of stationary fractional cointegration and the implied-realized volatility relation." Journal of Econometrics 133, 343-371.

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

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

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

Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. Econometric Theory 24, 651-676.

Lobato, I. N. (1999). A semiparametric two-step estimator in a multivariate long memory model. Journal of Econometrics 90, 129-53.

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

Nielsen, M. Ø. (2007). Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation. Journal of Business & Economic Statistics 25, 427-446.

Nielsen, M. Ø. and P. Frederiksen (2010). Fully modified narrow-band least squares estimation of weak fractional cointegration. Working paper, Queen's University.

Nielsen, M. Ø. and K. Shimotsu (2007). Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. Journal of Econometrics 141, 574-96.

Phillips, P. C. B. (1999). Discrete Fourier transforms of fractional processes. Cowles Foundation Discussion Paper #1243, Yale University.

Phillips, P. C. B. and K. Shimotsu (2004). Local Whittle estimation in nonstationary and unit root cases. Annals of Statistics 32, 656-692.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630-61.

Robinson, P. M. (2005). The distance between rival nonstationary fractional processes. Journal of Econometrics 128, 283-300.

Robinson, P. M. (2008). Multiple local Whittle estimation in stationary systems. Annals of Statistics 36, 2508-2530.

Robinson, P. M. and D. Marinucci (2001). Narrow-band analysis of nonstationary processes. Annals of Statistics 29, 947-86.

Robinson, P. M. and Y. Yajima (2002). Determination of cointegrating rank in fractional systems. Journal of Econometrics 106, 217-41.

Shao, X. and W. B. Wu (2007). Local Whittle estimation of fractional integration for nonlinear processes. Econometric Theory 23, 899-929.

Shimotsu, K. (2007). Gaussian semiparametric estimation of multivariate fractionally integrated processes. Journal of Econometrics 137, 277-310.

Shimotsu, K. (2010a). Exact local Whittle estimation of fractional integration with unknown mean and time trend. Econometric Theory 26, 501-540.

Shimotsu, K. (2010b). Supplement to "Exact local Whittle estimation of fractionally cointegrated systems." Working paper, Hitotsubashi University.

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

Shimotsu, K. and P. C. B. Phillips (2006). Local Whittle estimation of fractional integration and some of its variants. Journal of Econometrics 130, 209-233.

Tsay, W.-J. and Chung, C.-F. (2000). The spurious regression of fractionally integrated processes. Journal of Econometrics 96, 155-182.

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

Velasco, C. (2003). Gaussian semiparametric estimation of fractional cointegration. Journal of Time Series Analysis 24, 345-378.

Table 1: Simulation results with $\delta_1 = 0.1$ and $\rho = 0.0$

		ELW				Tapered Estimator			Stationary LW		
		δ_1	δ_2	β	ρ	δ_1	δ_2	β	δ_1	δ_2	β
	bias	-0.005	-0.002	-3.271	0.005	-0.010	0.008	-3.233	-0.014	-0.005	-0.471
$\delta_2 = 0.4$	s.d.	0.086	0.085	251.1	0.288	0.129	0.123	251.0	0.081	0.078	21.03
	RMSE	0.086	0.085	251.1	0.288	0.130	0.124	251.0	0.082	0.078	21.04
	bias	-0.015	-0.001	0.001	0.000	-0.021	0.014	0.006	-0.020	0.013	-0.000
$\delta_2 = 0.8$	s.d.	0.083	0.076	1.340	0.105	0.133	0.125	0.594	0.082	0.082	0.022
	RMSE	0.084	0.076	1.340	0.105	0.135	0.125	0.594	0.084	0.083	0.022
	bias	-0.020	-0.002	-0.000	0.000	-0.030	0.034	-0.000	-0.024	-0.212	-0.000
$\delta_2 = 1.3$	s.d.	0.083	0.078	0.003	0.098	0.134	0.127	0.008	0.082	0.098	0.003
	RMSE	0.085	0.078	0.003	0.098	0.137	0.131	0.008	0.085	0.234	0.003

Note: The sample size and bandwidth are $n = 512$ and $m = n^{0.65} = 57$, respectively.

Table 2: Simulation results with $\delta_1 = 0.1$ and $\rho = 0.4$

		ELW				Tapered Estimator			Stationary LW		
		δ_1	δ_2	β	ρ	δ_1	δ_2	β	δ_1	δ_2	β
	bias	-0.004	-0.002	-19.14	0.013	-0.009	0.008	-19.45	-0.013	-0.005	-0.470
$\delta_2 = 0.4$	s.d.	0.085	0.085	220.5	0.250	0.127	0.120	220.4	0.079	0.078	10.25
	RMSE	0.085	0.085	221.3	0.250	0.127	0.120	221.3	0.081	0.078	10.26
	bias	-0.013	-0.002	0.029	0.003	-0.014	0.011	-0.005	-0.015	0.012	-0.001
$\delta_2 = 0.8$	s.d.	0.077	0.072	1.512	0.091	0.124	0.116	0.054	0.077	0.078	0.024
	RMSE	0.078	0.072	1.512	0.091	0.125	0.117	0.054	0.078	0.079	0.024
	bias	-0.015	-0.004	-0.000	0.003	-0.015	0.026	0.000	-0.029	-0.211	-0.001
$\delta_2 = 1.3$	s.d.	0.073	0.070	0.002	0.081	0.118	0.113	0.008	0.080	0.097	0.003
	RMSE	0.075	0.070	0.002	0.081	0.119	0.116	0.008	0.085	0.233	0.003

Note: The sample size and bandwidth are $n = 512$ and $m = n^{0.65} = 57$, respectively.

Table 3: Simulation results with $\delta_1 = 0.1$ and $\rho = 0.8$

		ELW				Tapered Estimator			Stationary LW		
		δ_1	δ_2	β	ρ	δ_1	δ_2	β	δ_1	δ_2	β
	bias	0.001	-0.002	-8.737	0.004	-0.003	0.005	-8.821	-0.007	-0.004	-0.140
$\delta_2 = 0.4$	s.d.	0.078	0.078	75.80	0.092	0.118	0.111	75.79	0.073	0.072	2.353
	RMSE	0.078	0.078	76.30	0.092	0.118	0.111	76.30	0.074	0.072	2.357
	bias	-0.009	-0.005	-0.010	0.000	0.001	0.007	-0.020	0.003	0.016	-0.001
$\delta_2 = 0.8$	s.d.	0.064	0.064	1.604	0.041	0.103	0.099	1.512	0.066	0.068	0.021
	RMSE	0.064	0.064	1.605	0.041	0.103	0.099	1.512	0.066	0.070	0.021
	bias	-0.010	-0.007	0.000	0.001	0.011	0.028	0.001	-0.046	-0.213	-0.001
$\delta_2 = 1.3$	s.d.	0.061	0.060	0.002	0.034	0.099	0.099	0.006	0.077	0.091	0.002
	RMSE	0.061	0.060	0.002	0.034	0.100	0.103	0.006	0.089	0.232	0.002

Note: The sample size and bandwidth are $n = 512$ and $m = n^{0.65} = 57$, respectively.

Table 4: Simulation results with $\delta_1 = 0.3$ and $\rho = 0.4$

		ELW				Tapered Estimator			Stationary LW		
		δ_1	δ_2	β	ρ	δ_1	δ_2	β	δ_1	δ_2	β
	bias	0.002	0.001	-224.4	-0.036	-0.007	0.011	-223.6	-0.011	-0.004	-3.257
$\delta_2 = 0.4$	s.d.	0.086	0.082	675.6	0.545	0.125	0.121	675.7	0.079	0.079	80.42
	RMSE	0.086	0.082	711.9	0.546	0.125	0.121	711.7	0.080	0.079	80.49
	bias	-0.011	-0.001	-0.852	0.000	-0.012	0.012	-0.964	-0.014	0.012	-0.003
$\delta_2 = 0.8$	s.d.	0.080	0.076	27.78	0.121	0.127	0.119	27.63	0.079	0.080	0.060
	RMSE	0.080	0.076	27.79	0.121	0.127	0.119	27.65	0.080	0.081	0.060
	bias	-0.014	-0.003	0.000	0.003	-0.014	0.027	-0.000	-0.027	-0.212	-0.001
$\delta_2 = 1.3$	s.d.	0.074	0.070	0.005	0.081	0.120	0.114	0.016	0.080	0.097	0.006
	RMSE	0.075	0.070	0.005	0.081	0.120	0.117	0.016	0.085	0.233	0.006

Note: The sample size and bandwidth are $n = 512$ and $m = n^{0.65} = 57$, respectively.

Table 5: Simulation results when a penalty term is added to the objective function

		ELW				Tapered Estimator			Stationary LW		
		δ_1	δ_2	β	ρ	δ_1	δ_2	β	δ_1	δ_2	β
$(\delta_1, \delta_2) =$	bias	-0.003	-0.001	-1.167	0.013	-0.009	0.008	-1.414	-0.013	-0.005	-0.349
$(0.1, 0.4)$	s.d.	0.085	0.083	9.664	0.250	0.127	0.120	9.068	0.079	0.078	10.25
	RMSE	0.085	0.083	9.735	0.250	0.127	0.120	9.177	0.081	0.078	10.25
$(\delta_1, \delta_2) =$	bias	0.000	-0.002	-8.311	-0.036	-0.007	0.011	-7.618	-0.011	-0.004	-0.723
$(0.3, 0.4)$	s.d.	0.080	0.079	21.11	0.545	0.125	0.121	18.39	0.079	0.079	14.05
	RMSE	0.080	0.079	22.69	0.546	0.125	0.121	19.91	0.080	0.079	14.07
$(\delta_1, \delta_2) =$	bias	-0.011	-0.001	-0.049	0.000	-0.012	0.012	-0.124	-0.014	0.012	-0.003
$(0.3, 0.8)$	s.d.	0.080	0.075	2.859	0.121	0.127	0.119	2.329	0.079	0.080	0.060
	RMSE	0.081	0.075	2.859	0.121	0.127	0.119	2.332	0.080	0.081	0.060

Note: The sample size and bandwidth are $n = 512$ and $m = n^{0.65} = 57$, respectively. ρ is set to 0.4.

Table 6: Simulation results with t–distributed u_t: $u_t \sim t(\Sigma, (0,0)', 2)$

		ELW				Tapered Estimator			Stationary LW		
		δ_1	δ_2	β	ρ	δ_1	δ_2	β	δ_1	δ_2	β
	bias	-0.003	-0.001	-16.25	-0.006	-0.008	0.008	-16.44	-0.012	-0.004	-0.240
$\delta_2 = 0.4$	s.d.	0.082	0.077	201.3	0.337	0.123	0.110	201.3	0.076	0.072	7.844
	RMSE	0.082	0.077	202.0	0.337	0.123	0.110	201.9	0.077	0.072	7.848
	bias	-0.012	-0.002	0.015	-0.013	-0.013	0.011	-0.003	-0.012	0.014	-0.000
$\delta_2 = 0.8$	s.d.	0.073	0.068	0.862	0.264	0.118	0.107	0.055	0.073	0.074	0.026
	RMSE	0.074	0.068	0.862	0.264	0.118	0.107	0.055	0.074	0.075	0.026
	bias	-0.015	-0.004	0.000	-0.012	-0.013	0.029	0.000	-0.030	-0.214	-0.001
$\delta_2 = 1.3$	s.d.	0.070	0.066	0.003	0.261	0.113	0.105	0.009	0.076	0.096	0.003
	RMSE	0.072	0.066	0.003	0.261	0.114	0.109	0.009	0.081	0.234	0.003

Note: The sample size and bandwidth are $n = 512$ and $m = n^{0.65} = 57$, respectively. δ_1 and ρ are set to $\delta_1 = 0.1$ and $\rho = 0.4$.

Table 7: Descriptive statistics

	Mean	Standard Deviation	Skewness	Kurtosis
Implied volatility (σ^I)	0.168	0.066	1.624	7.077
Realized volatility (σ^R)	0.159	0.097	2.807	14.98

Table 8: Univariate two-step ELW estimates of δ

m	$n^{0.55} = 20$	$n^{0.6} = 26$	$n^{0.65} = 35$	$n^{0.7} = 46$	$n^{0.75} = 60$
σ^I	0.572	0.554	0.634	0.628	0.645
	(0.353, 0.792)	(0.362, 0.747)	(0.468, 0.800)	(0.483, 0.772)	(0.518, 0.772)
σ^R	0.512	0.480	0.550	0.561	0.609
	(0.293, 0.731)	(0.288, 0.672)	(0.384, 0.715)	(0.417, 0.706)	(0.482, 0.735)
$\sigma^R - \sigma^I$	0.250	0.246	0.319	0.377	0.457
	(0.031, 0.469)	(0.054, 0.438)	(0.153, 0.484)	(0.233, 0.522)	(0.330, 0.583)

Note: Asymptotic 95% confidence intervals are in parentheses.

Table 9: ELW-FCI estimates of $(\delta_1, \delta_2, \beta, \rho)$ and NBLS estimates of β

m	$n^{0.55} = 20$	$n^{0.6} = 26$	$n^{0.65} = 35$	$n^{0.7} = 46$	$n^{0.75} = 60$
δ_1	0.208	0.220	0.262	0.313	0.381
	(0.015, 0.402)	(0.050, 0.390)	(0.116, 0.407)	(0.182, 0.444)	(0.264, 0.499)
δ_2	0.619	0.675	0.675	0.644	0.642
	(0.425, 0.812)	(0.505, 0.845)	(0.530, 0.821)	(0.513, 0.775)	(0.524, 0.760)
β	1.052	1.107	1.045	1.032	1.039
	(0.925, 1.180)	(1.031, 1.183)	(0.930, 1.160)	(0.851, 1.213)	(0.791, 1.287)
ρ	0.609	0.594	0.621	0.554	0.487
β_{NB}	1.309	1.318	1.325	1.330	1.330

Note: Asymptotic 95% confidence intervals are in parentheses.