<table>
<thead>
<tr>
<th>項目</th>
<th>内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Exact Local Whittle Estimation of Fractionally Cointegrated Systems</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Shimotsu, Katsumi</td>
</tr>
<tr>
<td>Citation</td>
<td></td>
</tr>
<tr>
<td>Issue Date</td>
<td>2010-09</td>
</tr>
<tr>
<td>Type</td>
<td>Technical Report</td>
</tr>
<tr>
<td>Text Version</td>
<td>publisher</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10086/18681">http://hdl.handle.net/10086/18681</a></td>
</tr>
</tbody>
</table>
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 $0 \leq \delta_1 < \delta_2 < 1/2$.\textsuperscript{1} Hassler et al. (2006) and Velasco (2003) seek to estimate $\delta_1$ 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 $\delta_1, \delta_2$ 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.\textsuperscript{2}

The above procedures have an additional difficulty: prior to estimation, the researcher needs to know the range of the value of $\delta_1$ and $\delta_2$. Because the semiparametric estimators of $\delta$ 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 $\delta \gtrless 1/2$ is unknown \textit{a priori}; indeed, often empirical researchers want to test whether $\delta \gtrless 1/2$, because this determines whether the process is stationary (if $\delta > 1/2$) or nonstationary (if $\delta < 1/2$).

2. Because the value of $\delta$ 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

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

\textsuperscript{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 $\delta$ 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 $\delta_1$ and $\delta_2$ 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

\[
\begin{align*}
(1-L)^{\delta_1}(y_t - \beta x_t) &= u_{1t} I\{t \geq 1\}, \quad t = 1, 2, \ldots, \\
(1-L)^{\delta_2}x_t &= u_{2t} I\{t \geq 1\}, \quad t = 1, 2, \ldots, \\
y_t &= x_t = 0, \quad t \leq 0,
\end{align*}
\]

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 $\delta_2$th 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 $\delta_1$ that is smaller than $\delta_2$. We may also write (1) in matrix notation as

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

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

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

where $\Gamma(\cdot)$ is the gamma function. The model (1) provides a valid data-generating process for any value of $(\delta_1, \delta_2)$, 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

$$w_a(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} a_t e^{i\lambda_j t}, \quad \lambda_j = \frac{2\pi j}{n}, \quad j = 1, \ldots, n, \quad (4)$$

$$I_a(\lambda_j) = w_a(\lambda_j) \bar{w}_a(\lambda_j),$$

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 $p$th-order ta-
per 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 $w^T_\alpha(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^n h_t z_t e^{it\lambda_j}$ and $I^T_\alpha(\lambda_j) = w^T_\alpha(\lambda_j)w^T_\alpha(\lambda_j)$.

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 $j \geq \lceil \kappa m \rceil$. Using $\kappa$ 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)

$$ R(\vartheta) = \log \det \Omega^T(\vartheta) - 2(\delta_1 + \delta_2)p \frac{1}{(1 - \kappa)m} \sum_{j(p,\kappa)}^m \log \lambda_j, $$

where

$$ \Omega^T(\vartheta) = \frac{p}{1 - \kappa} \sum_{j(p,\kappa)}^m \text{Re} \left[ \Psi(\lambda_j; \delta) B I^T_\alpha(\lambda_j) B' \Psi(\lambda_j; \delta) \right], $$

and

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

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

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

where the $\eta_i$’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 $\beta$ 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 $\gamma$ to model the phase between $y_t - \beta x_t$ and $\alpha_t$ flexibly. In place of $\Psi(\lambda; \delta)$, Robinson (2008) uses $\Psi(\lambda; \delta, \gamma) = \text{diag}(\lambda^{\delta_1}, \lambda^{\delta_2} e^{-i\gamma})$ and defines the objective function $\hat{R}(\cdot)$ 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 $\Omega_0$ 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**

\[
 u_t - E u_0 = A(L) \varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \|A_j\|^2 < \infty,
\]

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 \( \sigma \)-field generated by \( \varepsilon_s \), \( s \leq t \), and there exists a scalar
random variable \( \varepsilon \) 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 \( \varepsilon_t \) 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,

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

where \( P \) satisfies \( P = PP' = \Omega_0 \) and \( \delta_0 \) 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** \( \vartheta_0 \) is an interior point of \( \Theta \).

**Assumption 5** For any \( C < \infty \),

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

**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 ($\gamma$ in Robinson (2008)) is identified by $\delta$ 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 $\nu_0$ denote its true value. $\nu_0$ 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 $\lambda_m$ and up to scale multiplication:

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

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

$$
I_{\Delta s''z}(\lambda_j; \beta) = w_{\Delta s''z}(\lambda_j; \beta) \bar{w}_{\Delta s''z}(\lambda_j; \beta), \quad w_{\Delta s''z}(\lambda_j; \beta) = \begin{pmatrix} w_{\Delta z_{1}(y-\beta x)}(\lambda_j) \\ w_{\Delta z_{2}x}(\lambda_j) \end{pmatrix}.
$$

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 $w_{Bz}(\lambda_j)$:

$$
w_u(\lambda_j) = w_{\Delta s''z}(\lambda_j; \beta) = \Lambda_n(e^{i\lambda_j}; \delta) v_{Bz}(\lambda_j; \delta),
$$
where
\[
\Lambda_n(e^{i\lambda_j}; \delta) = \begin{pmatrix}
D_n(e^{i\lambda_j}; \delta_1) & 0 \\
0 & D_n(e^{i\lambda_j}; \delta_2)
\end{pmatrix}, \quad v_{Bz}(\lambda_j; \delta) = \begin{pmatrix}
v_{Bz_1}(\lambda_j; \delta_1) \\
v_{Bz_2}(\lambda_j; \delta_2)
\end{pmatrix},
\]
\[
v_{Bz_a}(\lambda_j; \delta_a) = w_{Bz_a}(\lambda_j) - D_n(e^{i\lambda_j}; \delta_a)^{-1}(2\pi n)^{-1/2} \tilde{B}_{z_a,\lambda_j n}(\delta_a), \quad a = 1, 2,
\]
and Phillips (1999) provides the exact definition of \(\tilde{A}_{\lambda_j n}(d)\) 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
\[
f_u(\lambda_j) \sim \Omega \quad \text{and} \quad |D_n(e^{i\lambda_j}; \delta)|^2 \sim \lambda_j^{2\delta_a},
\]
the objective function is simplified to
\[
Q^*_m(\vartheta, \Omega) = \frac{1}{m} \sum_{j=1}^{m} \left\{ \log \det \Omega - 2 \log(\lambda_j^{\delta_1} + \lambda_j^{\delta_2}) + \text{tr} \left[ \Omega^{-1} \Delta_{\delta z}(\lambda_j; \beta) \right] \right\}.
\]

We propose to estimate \((\vartheta, \Omega)\) by minimizing \(Q^*_m(\vartheta, \Omega)\), so that
\[
(\vartheta^*, \Omega^*) = \arg \min_{\Omega \in (0, \infty)^2, \vartheta \in \Theta} Q^*_m(\vartheta, \Omega),
\]
where \(\Theta\) is defined in (6). Concentrating out \(\Omega\) from \(Q^*_m(\vartheta, \Omega)\), we find that \(\vartheta^*\) satisfies
\[
\vartheta^* = \arg \min_{\vartheta \in \Theta} R^*(\vartheta),
\]
where
\[
R^*(\vartheta) = \log \det \tilde{\Omega}^*(\vartheta) - 2(\delta_1 + \delta_2) \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \tilde{\Omega}^*(\vartheta) = \frac{1}{m} \sum_{j=1}^{m} \text{Re}[I_{\Delta_{\delta z}(\lambda_j; \beta)}].
\]

We consider the two-step estimator based on the objective function \(R^*(\vartheta)\). Let \(\hat{\vartheta}\) be the tapered local Whittle estimator of \(\vartheta\). The two-step estimator is defined as
\[
\vartheta^* = \hat{\vartheta} - \left[ (\partial^2/\partial \vartheta \partial \vartheta') R^*(\hat{\vartheta}) \right]^{-1} (\partial/\partial \vartheta) R^*(\hat{\vartheta}).
\]

Iterating the above procedure and updating the estimator by \(\vartheta_{(2)}^* = \vartheta^* - \left[ (\partial^2/\partial \vartheta \partial \vartheta') R^*(\vartheta^*) \right]^{-1} \times (\partial/\partial \vartheta) R^*(\vartheta^*)\) 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 $\sum_{j|j\geq k} \gamma_j = O((\log(k+1))^{-4})$, and $\sum_{j\geq k} A_j = O((\log(k+1))^{-4})$, where $\gamma_j = Eu_tu'_{t-j}$.

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.\textsuperscript{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 $\nu_0 \in (0, 1/2)$,

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

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

(b) When $\nu_0 \in (1/2, 3/2)$, 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 $\sup_{\lambda} \int |f_u(\lambda, \mu, \omega)|^2d\lambda < \infty$. Then

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

where $\Xi_\delta$ is the lower-right $(2 \times 2)$ block of $\Xi$.

**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.

\textsuperscript{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 $B_{z_t} - v = \text{diag}\{(1-L)^{-\delta_3}, (1-L)^{-\delta_2}\}u_1I\{t \geq 1\}$, where $v = (v_1, v_2)'$ is a nonrandom vector. Estimate $v_k$ by $\hat{v}_k = w(\delta_k)m^{-1}\sum_{t=1}^m(Bz_k)_{kt} + (1-w(\delta_k))(Bz)_{k1}$, where $(Bz)_{kt}$ is the $k$th 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 $\delta_k^*$ is not affected by $v$ if $\Delta^{b_1}(y_t - \beta x_t)$ and $\Delta^{b_2}x_t$ in the objective function are replaced by $\Delta^{b_1}(y_t - \beta_x + \hat{v}_1(\delta_1))$ and $\Delta^{b_2}(x_t - \hat{v}_2(\delta_2))$. Shimotsu (2010a) also shows that the presence of a polynomial time trend can be dealt with by prior detrending of the data.
Remark 2 \( \Xi \delta \) 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 \( \beta \). 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 \( \beta^* \) 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 \( \nu_0 \geq 1/2 \). Thus, the convergence rate of \( \beta^* \) is no slower than that of \( \hat{\beta} \).

Remark 5 The asymptotic distribution of \( \beta^* \) for \( \nu_0 > 1/2 \) remains an open question. We conjecture that it is not Gaussian, because the part of the Hessian corresponding to \( \beta \) (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 \( \vartheta_0 \). We can estimate \( \Omega \) and \( \rho \) by \( \hat{\Omega}^*(\hat{\vartheta}^*) = m^{-1}\sum_{j=1}^{m} \text{Re}[I_{\Delta^* z}(\lambda_j; \hat{\beta}^*)] \) and \( \rho^* = \hat{\Omega}^*(\hat{\vartheta}^*)_{12}/(\hat{\Omega}^*(\hat{\vartheta}^*)_{11}\hat{\Omega}^*(\hat{\vartheta}^*)_{22})^{1/2} \), respectively, which converge to \( \Omega \) and \( \rho \) in probability under the assumptions of Theorem 2. The asymptotic distribution of \( \rho^* \) can be derived in the same manner as the proof of Lemma 5 of Nielsen and Shimotsu (2007).

The convergence rate of \( \beta^* \) 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, \( \delta_1 \) and \( \delta_2 \). 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 \( \delta \) 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 \( \mu_t \) 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 \( \Sigma \) are fixed as 1 and the off-diagonal elements of \( \Sigma \), \( \rho \), 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 \( \delta_1 \) is fixed as 0.1. The value of \( \delta_2 \) 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 \( \beta \). The ELW estimator of \( \beta \) is very imprecise when \( \delta_2 - \delta_1 \) is small and appears to stay at a poor initial estimate of \( \beta \); the ELW and tapered estimators of \( \beta \) have almost identical
performance. The stationary LW estimator of $\beta$ 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 $\delta$. When $\delta_2 - \delta_1$ is small, the ELW estimator of $\delta$ is also affected by the poor estimates of $\beta$ and has a slightly larger RMSE than the stationary LW estimator. Interestingly, the ELW estimator of $\delta$ performs better than the tapered estimator even when both have a similar RMSE with respect to $\beta$. 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 $\delta_2$ 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: $\beta$ is imprecisely estimated by the ELW estimator when $\delta_2 - \delta_1$ is small; the ELW estimator of $\delta$ 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 $\beta$ 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 $\beta$, 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 $\rho$ affects the performance of the stationary LW estimator of $\delta_1$: its RMSE deteriorates as $\rho$ increases. Tables 1–3 report the performance of the ELW estimator of $\rho$ in the sixth column. The ELW estimator is unbiased across all the values of $\rho$ and $\delta_2$.

In Table 4, we examine the performance of the estimators when $\delta_1$ 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 $\rho$ is set to 0.4. The results for the other values of $\rho$ are qualitatively similar. Because $\delta_2 - \delta_1$ is smaller than in Table 2, from Theorem 2, we expect that the estimators of $\beta$ perform worse than in Table 2 and that in contrast, the estimators of $\delta_1$ and $\delta_2$ are not affected significantly. When $\delta_2 = 0.4$ and hence $\delta_2 - \delta_1 = 0.1$, the estimators of $\beta$, including the stationary LW estimator, perform very poorly. When $\delta_2 = 0.8$, the performance of all the estimators of $\beta$ improves, but the ELW and tapered estimators of $\beta$ have a large MSE. The performance of the estimators of $\delta_1$
and $\delta_2$ is similar to that in Table 2.

Tables 1–4 show that the estimates of $\beta$ 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 $\beta_{NB}$ is the narrow-band least squares (NBLS) estimator, and $C > 0$ is a constant. In effect, this penalization restricts $\beta$ 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 $\beta_{NB}$ 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 $\beta$ substantially without affecting the estimates of $\delta_1$ and $\delta_2$ negatively. The ELW estimator of $\beta$ 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 $\rho$ 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 regres-
where $\sigma^R_t$ and $\sigma^I_t$ are realized volatility and implied volatility, respectively, and $\epsilon_t$ is the residual term that includes the measurement error in implied volatility and a time-varying volatility risk premium. Note that $\epsilon_t$ 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 $y_t = \sigma^R_t$ and $x_t = \sigma^I_t$, 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, $\sigma^R_t = (n_t^{-1} \sum_{j=1}^{n_t} r_j^2 \times 252)^{1/2}$, where $r_j = \log(S_j/S_{j-1})$, $S_j$ is the closing value of the S&P 500 index on the $j$th 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 $\beta$; 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 $\beta$. 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 $\beta$ and $(\delta_1, \delta_2)$ 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 $\delta$ 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 $n^{\alpha}$, where $\alpha = 0.55, 0.6, 0.65, 0.7, 0.75$. The second row reports the estimates of $\delta$ for the implied volatility $\sigma^I$. Asymptotic 95% confidence intervals are reported in parentheses in the third row. The fourth and sixth rows report the $\delta$ estimates for realized volatility $\sigma^R$ and volatility difference $\sigma^R - \sigma^I$. The estimates of $\delta$ are around 0.6 for implied volatility and around 0.55 for realized volatility. Overall, the two volatility series have similar estimates of $\delta$, whereas the volatility difference $\sigma^R - \sigma^I$ has substantially smaller memory parameter estimates than both $\sigma^I$ and $\sigma^R$, 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 $\beta$ 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 $\beta_{NB}$ is used as the initial value for $\beta$, and the univariate two-step ELW estimators from $y_t - \beta_{NB}x_t$ and $x_t$ are used as the initial values for $\delta_1$ and $\delta_2$, 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 $\delta_1$ and $\delta_2$ are around 0.25 and around 0.65, respectively. The estimate of $\delta_1$ increases as $m$ increases, which may indicate a positive bias in the estimates of $\delta_1$ from short-run dynamics. In many cases, the estimates of $\delta_1$ are smaller than the $\delta$ estimates of $\sigma^R - \sigma^I$ in Table 8, whereas the estimates of $\delta_2$ are similar to the $\delta$ estimates of $\sigma^I$ in Table 8. For all $m$, the confidence intervals of $\delta_1$ and $\delta_2$ do not overlap with each other, which strongly suggests fractional cointegration between $\sigma^R_t$ and $\sigma^I_t$. The point estimates of $\beta$ are very close to one, and the hypothesis $\beta = 1$ is not rejected for all $m$. The estimates of $\rho$ 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 $\beta$. 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 \( 2 \times 2 \) matrix whose \((k, \ell)\)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 \( \hat{\delta} - \delta_0 = O_p((m/n)^{\nu_0}) \) and \( \hat{\beta} - \beta_0 = O_p((m/n)^{\nu_0}) \).

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 \( \mathcal{N}_\delta(c) = \{ \delta : |\delta - \delta_0| < c \} \) and \( \mathcal{N}_\beta(c) = \{ \beta : |\beta - \beta_0| < c \} \). Fix \( \varepsilon > 0 \) and define \( \mathcal{N}(\varepsilon) = \mathcal{N}_\beta(\varepsilon^{-1}(m/n)^{\nu_0}) \times \mathcal{N}_\delta(\varepsilon) \), and \( \bar{\mathcal{N}}(\varepsilon) = \Theta \setminus \mathcal{N}(\varepsilon) \). Define \( \zeta_i = \delta_i - \delta_0 \). We split the parameter space \( \Theta^\delta \) 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 \( \Theta^{\delta_2} = \Theta^\delta \setminus \Theta^{\delta_1} \). Since \( \Pr(\hat{\theta} \in \bar{\mathcal{N}}(\varepsilon)) \leq \Pr(\inf_{\mathcal{N}(\varepsilon)} \{ R(\hat{\theta}) - R(\vartheta_0) \} \leq 0) \), the consistency of \( \hat{\theta} \) follows if we show

\[
\Pr \left( \inf_{\mathcal{N}(\varepsilon) \cap \Theta^{\delta_1}} \{ R(\hat{\theta}) - R(\vartheta_0) \} \leq 0 \right) \to 0, \quad \text{as } n \to \infty, \quad (12)
\]

\[
\Pr \left( \inf_{\mathcal{N}(\varepsilon) \cap \Theta^{\delta_2}} \{ R(\hat{\theta}) - R(\vartheta_0) \} \leq 0 \right) \to 0, \quad \text{as } n \to \infty. \quad (13)
\]

First, we show (12). As in equation (7.1) in R08, rewrite \( R(\hat{\theta}) - R(\vartheta_0) \) as

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

where we use \( \hat{\Omega}^T(\hat{\theta}) \) 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

$$
\xi_t = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} = B_0 z_t = \begin{bmatrix} (1 - L)^{-\delta_01} u_{1t} I \{ t \geq 1 \} \\ (1 - L)^{-\delta_02} u_{2t} I \{ t \geq 1 \} \end{bmatrix}, \quad B_0 = \begin{bmatrix} 1 - \beta_0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{14}
$$

Define, analogously to R08 p. 2523, $H_j = (h_{kk}) = \Psi(\lambda_j; \delta_0)I_z^T \Psi(\lambda_j; \delta_0)$, and $\hat{\delta}(\delta) = (\hat{g}_{kk}^{(1)})$, where $\hat{g}_{kk}^{(1)} = p(1 - \kappa)^{-1} m^{-1} \sum_{j(p,\kappa)} (j/m)^{2\kappa} h_{kkj}$, and $\hat{g}_{12}^{(1)} = \hat{g}_{21}^{(1)} = p(1 - \kappa)^{-1} (2m)^{-1} \sum_{j(p,\kappa)} (j/m)^{\zeta_1 + \zeta_2} (e^{i(\pi - \lambda_j)}(\zeta_2 - \zeta_1)/2h_{12j} + e^{-i(\pi - \lambda_j)}(\zeta_2 - \zeta_1)/2h_{21j})$.

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

$$
U_\delta(\delta) = \log \det \{ \Upsilon(\delta) \hat{\delta}^{(1)}(\delta) \Upsilon(\delta) \hat{\delta}^{(1)}(\delta_0)^{-1} \} + \phi_1(\delta, \kappa) + u(\delta) - \phi_2(\delta, \kappa),
$$

$$
U_\beta(\vartheta) = \log \det \{ \hat{\Omega}^{T*}(\vartheta) \hat{\delta}^{(1)}(\delta)^{-1} \} - \phi_1(\delta, \kappa) + \phi_2(\delta, \kappa),
$$

where $\Upsilon(\delta) = \text{diag}((2\zeta_1 + 1)^{1/2}, (2\zeta_2 + 1)^{1/2})$, $\hat{\Omega}^{T*}(\vartheta) = \Xi(\vartheta) \hat{\Omega}^{T}(\vartheta) \Xi(\vartheta)$, $\Xi(\vartheta) = \text{diag}(\lambda_m^{-\zeta_1}, \lambda_m^{-\zeta_2})$, $\phi_1(\delta, \kappa) = \log[(1 - \kappa^2(1 - \kappa^{2\zeta_1 + 1})^{-1}(1 - \kappa^{2\zeta_2 + 1})^{-1}]$, $\phi_2(\delta, \kappa) = 2(\zeta_1 + \zeta_2)(1 - \kappa)^{-1} \kappa \log \kappa$, $u(\delta) = \sum_{i=1}^{2} [2\zeta_i - \log(2\zeta_i + 1) + 2\zeta_i(\log m - p(1 - \kappa)^{-1} m^{-1} \sum_{j(p,\kappa)} \log j - 1)]$. 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 $\hat{\delta}^{(1)}(\delta)$ with the tapered periodograms and $p(1 - \kappa)^{-1} m^{-1} \sum_{j(p,\kappa)}$, and (ii) we use $U_\delta(\delta)$ instead of $U_\alpha(\alpha)$ in R08 because our model does not have the parameter $\gamma$.

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

$$
\Pr \left( \inf_{\mathcal{X}_{\delta}(\epsilon) \cap \Theta_{\delta}} U_\delta(\delta) \leq 0 \right) \to 0, \tag{15}
$$

$$
\Pr \left( \inf_{\mathcal{X}_{\delta}(\epsilon) \cap \Theta_{\delta}} U_\beta(\vartheta) \leq 0 \right) \to 0. \tag{16}
$$

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

$$\sup_{\Theta_{s1}} \left\| \Upsilon(\delta)[\hat{G}^{(1)}(\delta) - G^{(1)}(\delta)]\Upsilon(\delta) \right\| \rightarrow_p 0, \quad (17)$$

$$\sup_{\Theta_{s1}} \left\| \{\Upsilon(\delta)G^{(1)}(\delta)\Upsilon(\delta)\}^{-1} \right\| < \infty, \quad (18)$$

$$\inf_{\hat{N}_3(\varepsilon) \cap \Theta_{s1}} \{\log \det \{\Upsilon(\delta)G^{(1)}(\delta)\Upsilon(\delta)G^{(1)}(\delta_0)^{-1}\} + \phi_1(\vartheta, \kappa)\} \geq 0, \quad (19)$$

$$\lim_{n \to \infty} \inf_{\hat{N}_3(\varepsilon) \cap \Theta_{s1}} [u(\delta) - \phi_2(\delta, \kappa)] > 0. \quad (20)$$

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 $\Omega_0$ using Lemmas 1(b) and 2(c) of Shimotsu (2010b). For (18) and (19), direct calculation gives $\det\{\Upsilon(\delta)G^{(1)}(\delta)\Upsilon(\delta)\} = (2\zeta_1 + 1)(2\zeta_2 + 1)(1 - \kappa)^{-2}\int_\kappa^1 x^{2\zeta_1}dx \int_\kappa^1 x^{2\zeta_2}dx \omega_{11}\omega_{22} - \omega_{12}^2c(\delta)\cos^2\tau$, where $c(\delta) = (\int_\kappa^1 x^{\zeta_1+\zeta_2}dx)^2(\int_\kappa^1 x^{2\zeta_1}dx \int_\kappa^1 x^{2\zeta_2}dx)^{-1}$. Since $0 < c(\delta) \leq 1$ from the Cauchy-Schwarz inequality and $|\cos x| \leq 1$, the right hand side is no smaller than $(2\zeta_1 + 1)(2\zeta_2 + 1)(1 - \kappa)^{-2}\int_\kappa^1 x^{2\zeta_1}dx \int_\kappa^1 x^{2\zeta_2}dx \det \Omega_0 > 0$, 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 $\hat{g}_{kl}^{(i)}$ similarly to R08 p. 2523 but using $p(1 - \kappa)^{-1}m^{-1}\sum_{j=1}^m \zeta_j^{(p, \kappa)}$ and setting $\tau = (\zeta_2 - \zeta_1)\pi/2$ and $\gamma_0 = (\delta_{02} - \delta_{01})\pi/2$. As in R08 p. 2524, define $\hat{a}_1 = (\hat{g}_{12}^{(2)} - \hat{g}_{12}^{(3)} - 2\hat{g}_{12}^{(1)})/\det\{\hat{G}^{(1)}(\delta)\}$, and $\hat{a}_2 = (\hat{g}_{11}^{(3)}\hat{g}_{12}^{(3)} - \hat{g}_{12}^{(2)})/\det\{\hat{G}^{(1)}(\delta)\}$. Define $g_{kl}^{(i)}$, the population counterpart of $\hat{g}_{kl}^{(i)}$, analogously to $g_{kl}^{(1)}$: for example, $g_{21}^{(2)} = (1 - \kappa)^{-1}\omega_{22}\gamma\int_\kappa^1 x^{2\zeta_1-\delta_{02}+\zeta_2}dx$, and $g_{11}^{(3)} = (1 - \kappa)^{-1}\omega_{22}\int_\kappa^1 x^{2\zeta_1-\delta_{02}}dx$, where $\gamma = (\delta_2 - \delta_1)\pi/2$. Using summation by parts and Lemma 1(b) of Shimotsu (2010b), we obtain $\sup_{\Theta_s} \left| g_{kl}^{(i)} - g_{kl}^{(i)} \right| \rightarrow_p 0$ for $i = 1, 2, 3, k, \ell = 1, 2$ as $n \to \infty$.

Rewrite $U_\beta(\vartheta) = \log Q(b_n(\beta)) - \phi(\delta, \kappa) + \phi_2(\delta, \kappa)$, where $Q(s) = 1 + \hat{a}_1s + \hat{a}_2s^2$ and $b_n(\beta) = \lambda_n^{(0)}(\beta_0 - \beta)$. Define $a_1 = (g_{12}^{(2)}g_{22}^{(1)} - 2g_{12}^{(1)}g_{12}^{(2)})/\det\{G^{(1)}(\delta)\}$ and $a_2 = (g_{12}^{(3)}g_{22}^{(1)} - g_{12}^{(2)})/\det\{G^{(1)}(\delta)\}$. Following R08 p. 2525, the probability in (16) is bounded
by, with \( \rho = \sup_{\Theta_\delta} |\phi_1(\delta, \kappa) - \phi_2(\delta, \kappa)| < \infty, \)

\[
\Pr \left( \log \left\{ 1 - \sup_{\Theta_\delta} \frac{\hat{\alpha}_1}{\varepsilon} + \inf_{\Theta_\delta} \frac{\hat{\alpha}_2}{\varepsilon^2} \right\} \leq \rho \right) + \Pr \left( \sup_{\Theta_\delta} \frac{\hat{\alpha}_1}{\varepsilon} > \frac{1}{\varepsilon} \right) \leq 2 \Pr \left( \sup_{\Theta_\delta} |\hat{\alpha}_1 - a_1| + \frac{2}{\varepsilon} \sup_{\Theta_\delta} |\hat{\alpha}_2 - a_2| + \varepsilon \rho \geq \frac{1}{\varepsilon} \inf_{\Theta_\delta} |a_2 - \sup_{\Theta_\delta} |a_1| \right), \tag{21} \]

which has an additional term \( \varepsilon \rho \) compared with (7.13) of R08. Since \( \sup_{\Theta_\delta} |\hat{\alpha}_i - a_i| \to \rho 0 \) 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 \( \varepsilon \) can be arbitrarily small, the probability in (21) tends to zero if \( \inf_{\Theta_\delta} a_2 > 0 \). Recall

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

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

It remains to show (13). Write \( R(\vartheta) - R(\vartheta_0) = U_\delta^*(\vartheta) + U_\beta^*(\vartheta) \), where \( U_\delta^*(\vartheta) = \log \det \{ \Xi(\vartheta) \hat{G}^{(1)}(\vartheta) \Xi(\vartheta) \hat{G}^{(1)}(\vartheta)^{-1} \} - 2(\zeta_1 + \zeta_2) p(1 - \kappa)^{-1} m^{-1} \sum_{j(p, \kappa)} \log \lambda_j \) and \( U_\beta^*(\vartheta) = \log \det \{ \hat{\Omega}^T(\vartheta) \hat{G}^{(1)}(\vartheta)^{-1} \} = U_\beta(\vartheta) + \phi_1(\delta, \kappa) - \phi_2(\delta, \kappa) \). Then \( \Pr(\inf_{\hat{\Theta}_\delta} U_\delta^*(\vartheta) \leq 0) \to 0 \) follows from the proof of (16), so it suffices to show \( \Pr(\inf_{\hat{\Theta}_\delta} U_\delta^*(\vartheta) \leq 0) \to 0 \).

Rewrite \( U_\delta^*(\vartheta) \) as (see Shimotsu, 2007, p. 293)

\[
U_\delta^*(\vartheta) = \log \det \hat{D}(\vartheta) - \log \det \hat{D}(\vartheta_0),
\]

where

\[
\hat{D}(\vartheta) = \frac{p}{(1 - \kappa)m} \sum_{j(p, \kappa)} \left[ (j/q)^{2\zeta_1} h_{11j} (j/q)^{\zeta_1 + \zeta_2} \Re \{ e^{i(\pi - \lambda_j)(\zeta_2 - \zeta_1)/2} h_{12j} \} (j/q)^{2\zeta_2} h_{22j} \right],
\]

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

Define \( K(\vartheta) \) as \( \hat{D}(\vartheta) \) but \( h_{kj} \) is replaced with \( \omega_{kj} \). Note that \( \hat{D}(\vartheta) \) is identical to \( \hat{D}(\vartheta) \) in Shimotsu (2007, p. 294) except that \( \{ m^{-1} \sum_{j=[m]} \lambda_j, \theta_k, p, I_j \} \) in Shimotsu (2007) is replaced with \( \{ p(1 - \kappa)^{-1} m^{-1} \sum_{j(p, \kappa)} \zeta_k, q, H_j \} \). Therefore, \( \sup_{\Theta_\delta} |D(\vartheta) - K(\vartheta)| \sim p \) 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
\[ \inf_{\theta_2} \det K(\delta) \geq (1 + \varepsilon) \det G_0 + o(1). \] 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 \( p(1 - \kappa)^{-1}m^{-1} \sum_{j(p,\kappa)}^{m} e^{1+(1-\kappa)^{-1}k\log \kappa} \) as long as \( \kappa \) 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 \( \vartheta = (\beta, \gamma, \delta_1, \delta_2) = (\theta_1, \theta_2, \theta_3, \theta_4) \), 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 \( \vartheta = (\beta, \gamma, \delta_1, \delta_2) \), and write \( \hat{\Omega}^T \) as \( \hat{\Omega}^T(\theta) \). Similar to R08, define \( s^T(\theta) = (\partial/\partial \theta) R^T(\theta), \) \( S^T(\theta) = (\partial/\partial \theta') s^T(\theta) \), \( \Delta_n = \text{diag}\{\lambda_0^{\nu_0}, 1, 1, 1\} \), and denote by \( \tilde{S}^T \) the matrix \( S^T(\theta) \) whose elements are evaluated at a point between \( \theta_0 \) and \( \hat{\theta} \). The required result follows if we show, for a finite matrix \( \Sigma_\kappa \),

\[
\begin{align*}
    m^{1/2} \Delta_n^{-1} s^T(\theta_0) &= O_p(1), \\
    \Delta_n^{-1} \tilde{S}^T \Delta_n^{-1} &\rightarrow_p \Sigma_\kappa.
\end{align*}
\]

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 \( A_{ij}^T = \Psi(\lambda_j; \delta_0) B_0 I_{ij}^T B_0^\prime \Psi(\lambda_j; \delta_0) \). Define the score vectors \( s_1^T(\theta_0), \ldots, s_4^T(\theta_0) \) as \( s_1(\theta), \ldots, s_4(\theta) \) 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 \( k \)th element

\[
    s_k^T = 2p(1 - \kappa)^{-1}m^{-1} \sum_{j(p,\kappa)}^{m} \text{tr}(U_{Rkj} \text{Re}\{I_j^T\} + U_{Ikj} \text{Im}\{I_j^T\}) \]

where the coefficients \( U_{Rkj} \) and \( U_{Ikj} \) are defined similarly to R08 p. 2527 by replacing \( m^{-1} \sum_j I_{ij} \) in R08 with \( p(1 - \kappa)^{-1}m^{-1} \sum_{j(p,\kappa)}^{m} I_{ij}^T \). We do not provide the explicit formula for \( U_{Rkj} \) and \( U_{Ikj} \) here. Because \( E s^T(\theta_0) = 0 \), equation (22) follows if we show \( \var(m^{1/2} \Delta_n^{-1} s^T(\theta_0)) = O(1) \). 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), \( \var(m^{1/2} \Delta_n^{-1} s^T(\theta_0)) \) is bounded by \( O(m^{-1} \sum_{j(p,\kappa)}^{m} \var(I_j^T)) \).
\(O((\log m)^2m^{-1}\sum_{j(p,k),j\neq k}^{m}\sum_{k(p,k)}^{m}\text{cov}(I_{kj}, I_{lk}))\). From Velasco (1999, p. 114), \(\text{var}(I_{kj}^T) = V + O(n^{-1})\), where \(V\) does not depend on \(j\), and \(\text{cov}(I_{kj}^T, I_{lk}^T) = O(|j-k|^{-2p} + |j+k|^{-2p} + n^{-1})\) for \(j \neq k\). Therefore, \(\text{var}(m^{1/2}\Delta_n^{-1}sT^*(\theta_0)) = O(1)\) 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:

\[
\Delta_n^{-1}\{\bar{S}^T - S^T(\theta_0)\}\Delta_n^{-1} \to_p 0, \tag{24}
\]

\[
\frac{1}{2}\Delta_n^{-1}S^T(\theta_0)\Delta_n^{-1} \to_p \Sigma_k. \tag{25}
\]

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 \(\Omega_0\). First, from Lemma 1 of Lobato and Velasco (2000) and Lemma 1 of Shimotsu (2010b), we have \(\sum_{j(p,k)}^{s}(A_{0j}^T - \Omega_0) = O_p(s^\beta n^{-\beta} + \log n + s^{1/2})\) for any \(1 \leq s \leq m\). Combining it with Lemma 2(c) of Shimotsu (2010b) and summation by parts, we obtain \(\sum_{j(p,k)}^{s}(A_{0j}^{T(k,\ell)} - \Omega_0^{(k,\ell)}) = O_p(s^\beta n^{-\beta} + \log n + s^{1/2})\) at \(\theta = \theta_0\), and (25) follows. \(\square\)

### 7.2 Proof of Theorem 2 (a)

The required result follows if we show, for any \(\bar{\theta}\) such that \(\bar{\theta} - \theta_0 = O_p(m^{-1/2})\),

\[
m^{1/2}(\Delta_n^*)^{-1}(\partial/\partial\bar{\theta})R^*(\theta_0) \to_d N(0, \Xi), \quad \text{and} \quad (\Delta_n^*)^{-1}(\partial^2/\partial\bar{\theta}\partial\bar{\theta}')R^*(\bar{\theta})(\Delta_n^*)^{-1} \to_p \Xi. \tag{26}
\]

#### 7.2.1 Score vector with respect to \(\delta\)

From the proof of Theorem 4 of R08, \((\partial/\partial\delta_k)R^*(\theta_0)\) satisfies the first result in (26) if

\[
(\partial/\partial\delta_k)R^*(\theta_0) = \text{tr}((\partial/\partial\delta_k)\hat{\Omega}^*(\theta_0)\hat{\Omega}^*(\theta_0)^{-1}) - 2m^{-1}\sum_{j=1}^{m}\log \lambda_j
\]

\[
= s_1^* + (-1)^k(\pi/2)s_2^* + o_p(m^{-1/2}), \tag{27}
\]

where \(s_1^*\) and \(s_2^*\) are defined in (8.5) of R08. Define \(w_{\Delta lj} = (w_{\log(1-L)uj1}, w_{ujj})'\), then \((\partial/\partial\delta_l)\hat{\Omega}^*(\theta_0) = E_{11}m^{-1}\sum_{j=1}^{m}\text{Re}\{w_{\Delta lj}w_{ujj}\} + m^{-1}\sum_{j=1}^{m}\text{Re}\{w_{ujj}\}E_{11}\). Because \(w_{\Delta lj}\) is premultiplied by \(E_{11}\), we only need to analyze the first row of \(m^{-1}\sum_{j=1}^{m}\text{Re}\{w_{\Delta lj}w_{ujj}\}\). Define the \((1,1)\)th element of \(m^{-1}\sum_{j=1}^{m}\text{Re}\{w_{\Delta lj}w_{ujj}\}\) as \(s_1 = m^{-1}\sum_{j=1}^{m}w_{\log(1-L)uj1}w_{ujj}\).
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:

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

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 $(\partial/\partial \delta) \tilde{\Omega}^*(\vartheta_0) = E_{11} \Phi + \Phi E_{11} + o_p(m^{-1/2})$, 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 $A(\lambda_j) I_{uj} \bar{A}(\lambda_j)$ by Lemma 1(b1) of Shimotsu (2007) and then by $PI_{ej} P$ using Assumptions 1-3, we obtain $\Phi = m^{-1} \sum_{j=1}^m \log \lambda_j P \text{Re} \{I_{ej}\} P' + (\pi/2) m^{-1} \sum_{j=1}^m P \text{Im} \{I_{ej}\} P' + o_p(m^{-1/2})$, in which the terms with $\lambda_j \text{Im} \{I_{ej}\}$ reduce to $o_p(m^{-1/2})$ by $E \text{Im} \{I_{ej}\} = 0$, $\sum_{j=1}^m (I_{ej} - I_2) = O(r^{1/2})$ (Lobato, 1999, p. 145, (C.3)), and summation by parts. It follows that

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

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

$$(\partial/\partial \delta_k) R^*(\vartheta_0) = \frac{2}{m} \sum_{j=1}^m \text{tr} \left\{ \left( \log j - \frac{1}{m} \sum_{j=1}^m \log j \right) P' \Omega_0^{-1} E_{kk} P \text{Re} \{I_{ej}\} \right\}$$

$$+ (-1)^k \frac{\pi}{2} \sum_{j=1}^m \text{tr} \left\{ P' \Omega_0^{-1} E_{22} P \text{Im} \{I_{ej}\} \right\} + o_p(m^{-1/2}).$$

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 $\beta$

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

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

(28)
where $s_1(\theta_0)$ is defined similarly to p. 2527 of R08 but $A_{0j}$ in R08 is replaced with $I_{u_j}$.

Define $w_{\Delta2j} = (0, w_{\Delta\epsilon_{1u_{2j}}}')$ so that $E_{12} w_{\Delta2j} = (w_{\Delta\epsilon_{1u_{2j}}}, 0)'$. Then

$$\frac{\partial}{\partial \beta} \tilde{\Omega}^*(\vartheta) = -E_{12} m \sum_{j=1}^{m} \text{Re}\{w_{\Delta2j} \bar{w}_{\Delta\epsilon_{2j}}(\lambda_j; \beta)\} - m \sum_{j=1}^{m} \text{Re}\{w_{\Delta\epsilon_{2j}}(\lambda_j; \beta) \bar{w}_{\Delta2j}\} E_{21}. \quad (29)$$

Define $w_{\Delta2j}$ evaluated at $\vartheta_0$ as $w_{\Delta\varphi_{2j}} = (0, w_{\Delta-\nu_0u_{2j}}')$, then we have

$$\frac{\partial}{\partial \beta} \tilde{\Omega}^*(\vartheta_0) = -E_{12} m^{-1} \sum_{j=1}^{m} \text{Re}\{w_{\Delta\varphi_{2j}} \bar{w}_{u_{2j}}\} - m^{-1} \sum_{j=1}^{m} \text{Re}\{w_{u_{2j}} \bar{w}_{\Delta\varphi_{2j}}\} E_{21}.$$

We approximate $w_{\Delta-\nu_0u_{2j}} \bar{w}_{u_{2j}}$ by $e^{i\pi \nu_0/2} \lambda_{-\nu_0} w_{u_{2j}} \bar{w}_{u_{2j}}$ 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

$$w_{\Delta-\nu_0 u_{2j}} = D_n(e^{i\lambda_j}; -\nu_0) u_{2j} \left(2\pi n\right)^{-1/2} \lambda_j^{\nu_0} \tilde{U}_{2,j,n}(-\nu_0).$$

Define $D_{nj}(\nu_0) = \lambda^{\nu_0} D_n(e^{i\lambda_j}; -\nu_0)$ as on p. 226 of SP06. Then

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

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

$$e^{i\pi \nu_0/2} m^{-1} \sum_{j=1}^{m} \left(\frac{j}{m}\right)^{-\nu_0} I_{u_{2j}} + o_p(m^{-1/2}),$$

by approximating $I_{u_{2j}}$ by $A(\lambda_j) I_{e_j} \bar{A}(\lambda_j)$ first (see Lemma 1(b1) of Shimotsu (2007)) and then using the order of the covariance between $I_{\epsilon_j}$ and $I_{ek}$.

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, $E||\tilde{U}_{\lambda,n}(-\nu_0) - A(0) \tilde{\epsilon}_{\lambda,n}(-\nu_0)||^2 = O(n^{1+2\nu_0}s^{-2\nu_0-1}(\log n)^{-4} + n^{1+2\nu_0}s^{-2})$, where $\tilde{U}_{\lambda,n}(\nu_0)$ and $\tilde{\epsilon}_{\lambda,n}(\nu_0)$ are $2 \times 1$ 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 $\tilde{U}_{\lambda,n}(-\nu_0)$, we may write the second term in (30) as $U_n + o_p(m^{-1/2})$, where $U_n = m^{-1} \sum_{j=1}^{m} \left(\frac{j}{m}\right)^{-\nu_0} (2\pi n)^{-1/2} \lambda_j^{\nu_0} A_2(0) \tilde{\epsilon}_{\lambda,n}(-\nu_0) \bar{w}_{\epsilon_j} \bar{A}_2(\lambda_j)$, 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 $\epsilon_t$, and $(d_0, \nu_j)$ in $T_n'$ corresponds to $(\nu_j, (j/m)^{-\nu_0})$ in $mU_n$. Observe that, if we replace $\nu_j$ 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 $\lambda^{\nu_0} m^{-1} \sum_{j=1}^{m} w_{\Delta-\nu_0 u_{2j}} \bar{w}_{u_{2j}}$, and conse-
quently, \((\partial / \partial \beta) \tilde{\Omega}^*(\vartheta_0) = -m^{-1} \sum_{j=1}^{m} \lambda_j^{\nu_0}(E_{12} I_{uj} e^{i\nu_0/2} + I_{uj} E_{21} e^{-i\nu_0/2}) + o_p(\lambda_m^{\nu_0} m^{-1/2}),\) giving (28).

### 7.2.3 Hessian approximation

We prove the required result by approximating the derivatives of \(\tilde{\Omega}^*(\vartheta_0) = m^{-1} \sum_{j=1}^{m} \text{Re}\{I_{\Delta^*_2}(\lambda_j; \beta)\}\) by the derivatives of the counterpart in R08 p. 2512, \(\hat{\Omega}(\theta_0) = m^{-1} \sum_{j=1}^{m} \text{Re}\{A(\lambda_j; \theta)\}\).

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 \(PI_{ej}P'\). 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 \(I_{\Delta^*_2}(\lambda_j; \beta)\) at \(\vartheta = \vartheta_0\) are equal to linear combinations of \(I_j^{(k)}\) and \(I_j^{(k,\ell)}\) up to an negligible term.

First, \(w_{\Delta^*_2}(\lambda_j; \beta_0) = w_{uj}\) from the definition. Second, for the derivative with respect to \(\delta_k\), it follows from the derivation of the score approximation above that, for \(k = 1, 2,\)

\((\partial / \partial \delta_k) \tilde{\Omega}(\vartheta_0) = [\partial / \partial \theta_{2+k} + (-1)^k(\pi / 2)(\partial / \partial \theta_2)] \tilde{\Omega}(\vartheta_0) + o_p(\log n)]^C\). For the derivative with respect to \(\beta\), (28) implies \((\partial / \partial \beta) \tilde{\Omega}^*(\vartheta_0) = (\partial / \partial \beta) \hat{\Omega}^*(\vartheta_0) + o_p(\lambda_m^{\nu_0} \log n)]^C\). Similarly, we can use Lemma 4 of Shimotsu (2010b) to express the other derivatives of \(\hat{\Omega}^*(\vartheta_0)\) in terms of the derivatives of \(\hat{\Omega}(\theta)\) in R08 such as \((\partial^2 / \partial \delta_k \partial \beta) \tilde{\Omega}^*(\vartheta_0) = [\partial / \partial \theta_{2+k} + (-1)^k(\pi / 2)(\partial / \partial \theta_2)](\partial / \partial \beta) \hat{\Omega}(\vartheta_0) + o_p(\lambda_m^{\nu_0} \log n)]^C\). We suppress the obvious formula for \((\partial^2 / \partial \delta_k \partial \delta_l) \tilde{\Omega}^*(\vartheta_0)\). Therefore, it follows from the relation between \(\tilde{\vartheta}\) and \(\theta\) that \((\Delta_n^*)^{-1}(\partial^2 / \partial \tilde{\vartheta} \partial \tilde{\vartheta}') R^*(\tilde{\vartheta}) (\Delta_n^*)^{-1} \to_p \Xi.\)

Finally, the proof of \((\Delta_n^*)^{-1}[\partial^2 / \partial \tilde{\vartheta} \partial \tilde{\vartheta}'] R^*(\tilde{\vartheta}) - (\partial^2 / \partial \tilde{\vartheta} \partial \tilde{\vartheta}') R^*(\vartheta_0)](\Delta_n^*)^{-1} = o_p(1)\) follows from the root–m consistency of \(\tilde{\vartheta}\) and \(\lambda_j^0 - 1 = O(\log n)\) for any finite \(\alpha\). \(\square\)

### 7.3 Proof of Theorem 2 (b)

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

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

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

\[H_{11,n} = O_p(1), \quad \text{Pr}(|H_{11,n}| < \varepsilon) \to 0 \text{ as } \varepsilon \to 0, \quad H_{12,n}, H_{21,n} \to_p 0, \quad H_{22} \to_p \Xi_{\delta}, \quad (31)\]
and (iii) \((\Delta_n^{**})^{-1}(\partial^2/\partial \vartheta \partial \vartheta')[R^* (\vartheta) - R^*(\vartheta_0)](\Delta_n^{**})^{-1} \to p 0\), for any \(\vartheta\) such that \(\Delta_n^*(\vartheta - \vartheta) = O_p(m^{-1/2})\). We omit a tedious but straightforward proof of \((\Delta_m^{**})^{-1}(\partial^2/\partial \vartheta \partial \vartheta')[R^* (\vartheta) - R^*(\vartheta_0)](\Delta_n^{**})^{-1} \to p 0\).

### 7.3.1 Score vector approximation

The score with respect to \(\delta_k\) remains unchanged, because it does not depend on the value of \(\nu_0\). We need only to analyze the score with respect to \(\beta\). 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 \((\partial/\partial \beta)\tilde{\Omega}^*(\vartheta_0) = -E_{12}m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} - m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} E_{21}\), hence \((\partial/\partial \beta)R^*(\vartheta_0) = -2 \text{tr} [E_{12}m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} \tilde{\Omega}^*(\vartheta_0) - \Omega^*(\vartheta_0) - \Omega_0^*]\). Define \(c_n = m^{-1}\sum_{j=1}^m \text{Re}\{(1 - e^{i\lambda_j})^{-\nu_0}\}, \) which satisfies \(c_n = O(n^{\nu_0}m^{-1/2} + n^{\nu_0}m^{-1}\log m) = O(n^{\nu_0}m^{-1/2-n})\) for some \(\eta > 0\). Using \(\text{tr}(E_{12}) = 0\), rewrite \((\partial/\partial \beta)R^*(\vartheta_0)\) further as

\[-2 \text{tr} [E_{12}m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} (\tilde{\Omega}^*(\vartheta_0) - \Omega_0^*) - 2 \text{tr}[E_{12}(m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} - c_n\Omega_0]\Omega_0^{-1}]\].

Then (32) follows if we show

\[
E_{12}\left(m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} - c_n\Omega_0\right) = O_p(n^{\nu_0}m^{-1}). \tag{33}
\]

We prove (33) only for \(m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} - c_n\omega_{22}\). The other elements are analyzed similarly. First, it follows from Lemma A.1 of Phillips and Shimotsu (2004) that

\[
(1 - e^{i\lambda})w_{\Delta^{0}v\omega_j\nu} = D_n(e^{\lambda}; 1 - \nu_0)w_{u_j} - (2\pi n)^{-1/2}\tilde{U}_{2,\lambda, n}(1 - \nu_0) - (2\pi n)^{-1/2}e^{i\lambda_j}Y_n, \tag{34}
\]

where \(Y_n = \Delta^{0}v_{2n}I\{t \geq 1\}\). Hence, \(m^{-1}\sum_{j=1}^m \text{Re}\{w_{\Delta^{0}v\omega_j\nu} \tilde{w}_{u_j}\} - c_n\omega_{22}\) is written as \(\text{Re}\{T_{1n} + T_{2n} + T_{3n}\}\), where

\[
T_{1n} = m^{-1}\sum_{j=1}^m (1 - e^{i\lambda})^{-1}D_n(e^{\lambda}; 1 - \nu_0)I_{u_j} - c_n\omega_{22},
\]

\[
T_{2n} = m^{-1}\sum_{j=1}^m (1 - e^{i\lambda})^{-1}(2\pi n)^{-1/2}\tilde{U}_{2,\lambda, n}(1 - \nu_0)\tilde{w}_{u_j},
\]

\[
T_{3n} = m^{-1}\sum_{j=1}^m (1 - e^{i\lambda})^{-1}(2\pi n)^{-1/2}e^{i\lambda_j}Y_n\tilde{w}_{u_j}.
\]
For $\text{Re}\{T_{1n}\}$, noting that $D_n(e^{i\lambda_j}; 1 - \nu_0) = (1 - e^{i\lambda_j})^{-1} - (1 - e^{i\lambda_{j+1}})^{-1}$ from Lemma A.2 of Phillips and Shimotsu (2004) and using the definition of $c_n$, we have $\text{Re}\{T_{1n}\} = m^{-1} \sum_{j=1}^{m} \text{Re}\{(1 - e^{i\lambda_j})^{-1}\} \{I_{u_2j} - \omega_{22}\} + O_p(n^{v_0}m^{-1})$.

Therefore, using Lemma 3(b) of Shimotsu (2010b), summation by parts, and $(1 - e^{i\lambda_j})^{-1} = O(n^{v_0}j^{-v_0-1})$, and noting $\nu_0 > 1/2$, we obtain $\text{Re}\{T_{1n}\} = m^{-1} \sum_{j=1}^{m} \text{Re}\{(1 - e^{i\lambda_j})^{-1}\} \{f_{22}(\lambda_j) - \omega_{22}\} + O_p(n^{v_0}m^{-1})$. The first term on the right has the order $m^{-1} \sum_{j=1}^{m} j^{-v_0+b}n^{v_0-b} = n^{v_0}m^{-1}(n^{-b} \sum_{j=1}^{m} j^{-v_0+b})$. When $\nu_0 \geq 1$, $(n^{-b} \sum_{j=1}^{m} j^{-v_0+b})$ is $o(1)$. When $\nu_0 \in (1/2, 1)$, it is $O(n^{-b}m^{-1}+b)$, thus $\text{Re}\{T_{1n}\} = O_p(n^{v_0}m^{-1})$.

It remains to show the order of $T_{2n}$ and $T_{3n}$. For $T_{2n}$, it easily follows from the order of $\tilde{U}_{2,\lambda_n}(1 - \nu_0)$ provided in Lemma A.5 of Phillips and Shimotsu (2004) that $T_{2n} = O_p(m^{-1} \sum_{j=1}^{m} j^{-1/2}n^{v_0-1/2}j^{1/2-v_0}) = O_p(n^{v_0}m^{-1})$. For $T_{3n}$, it follows from Lemma 3(a) of Shimotsu (2010b) that $T_{3n} = m^{-1}n^{-1/2}Y_{3n} \sum_{j=1}^{m} (1 - e^{i\lambda_j})^{-1}e^{i\lambda_j} \bar{w}_{\epsilon j} \bar{A}_2(\lambda_j) + O_p(n))$. This is $O_p(n^{v_0}m^{-1})$ because $w_{\epsilon j}$ and $w_{\delta k}$ are uncorrelated for $j \neq k$ and $n^{1/2-v_0}Y_{3n} \to_d N(0, \sigma^2)$ 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 $\beta$. Note that
\[
\frac{\partial^2 R^*(\vartheta)}{\partial \beta \partial \delta_k} = \text{tr} \left\{ \frac{\partial^2 \tilde{\Omega}^*(\vartheta)}{\partial \beta \partial \delta_k} \tilde{\Omega}^*(\vartheta)^{-1} - \frac{\partial \tilde{\Omega}^*(\vartheta)}{\partial \beta} \frac{\partial \tilde{\Omega}^*(\vartheta)}{\partial \delta_k} \tilde{\Omega}^*(\vartheta) \right\},
\]
\[
\frac{\partial^2 R^*(\vartheta)}{\partial \beta^2} = \text{tr} \left\{ \frac{\partial^2 \tilde{\Omega}^*(\vartheta)}{\partial \beta^2} \tilde{\Omega}^*(\vartheta)^{-1} - \frac{\partial \tilde{\Omega}^*(\vartheta)}{\partial \beta} \frac{\partial \tilde{\Omega}^*(\vartheta)}{\partial \beta} \tilde{\Omega}^*(\vartheta)^{-1} \right\}.
\]

The required result follows if
\[
(\partial^2/\partial \beta \partial \delta_k)R^*(\vartheta_0) = O_p(n^{v_0}m^{-1/2}),
\] (35)

and $n^{-2v_0}m(\partial^2/\partial \beta^2)R^*(\vartheta_0)$ satisfies the condition (31) of $H_{11,n}$. We analyze $(\partial^2/\partial \beta \partial \delta_k)R^*(\vartheta_0)$ first. $(\partial/\partial \beta)\tilde{\Omega}^*(\vartheta_0)\tilde{\Omega}^*(\vartheta_0)^{-1} = O_p(n^{v_0}m^{-1/2})$ holds because $(\partial/\partial \beta)\tilde{\Omega}^*(\vartheta_0) = O(c_n) + O_p(n^{v_0}m^{-1}) = O_p(n^{v_0}m^{-1/2-\eta})$ 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 $w_{\Delta \delta \epsilon} (\lambda_j; \beta)$, and that $(\partial^2/\partial \beta \partial \delta_k)\tilde{\Omega}^*(\vartheta)$ consists of $w_{\Delta 2j}$ and $w_{\Delta \delta \epsilon} (\lambda_j; \beta)$ and their derivatives with respect to $\delta_k$. From Lemma 4 of Shimotsu (2010b), the leading term of these derivatives is $(\log n)^s$, $s = 1, 2$, times $w_{\Delta 2j}$ and $w_{\Delta \delta \epsilon} (\lambda_j; \beta)$.
fore, 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 \( (\partial / \partial \beta) \tilde{\Omega}^*(\vartheta_0) = o_p(n^{\nu_0} m^{-1/2}) \) 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 \( (\partial^2 / \partial \beta^2) \tilde{\Omega}^*(\vartheta_0) = 2 E_{12} m^{-1} \sum_{j=1}^{m} \text{Re} \{ w_{t_0j} w_{t_0j} \} E_{21} \). The only non-zero element of this matrix is its (1,1)th element, \( 2 m^{-1} \sum_{j=1}^{m} I_{\Delta_{t-\nu_0u_j}} \). From Theorems 4.5 and 5.1 of Robinson and Marinucci (2001), we have \( \lim_{n \to \infty} E[n^{-2\nu_0} \sum_{j=1}^{m} I_{\Delta_{t-\nu_0u_j}}] < \infty \) and \( \lim_{n \to \infty} \text{var}[n^{-2\nu_0} \sum_{j=1}^{m} I_{\Delta_{t-\nu_0u_j}]} = V \in (0, \infty) \). 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


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

<table>
<thead>
<tr>
<th>$\delta_2$</th>
<th>ELW</th>
<th>Tapered Estimator</th>
<th>Stationary LW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias</td>
<td>$\delta_1$</td>
<td>$\delta_2$</td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>-0.005</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.086</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.086</td>
<td>0.085</td>
</tr>
<tr>
<td>0.8</td>
<td>bias</td>
<td>-0.015</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.083</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.084</td>
<td>0.076</td>
</tr>
<tr>
<td>1.3</td>
<td>bias</td>
<td>-0.020</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.083</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.085</td>
<td>0.078</td>
</tr>
</tbody>
</table>

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 $\rho = 0.4$

<table>
<thead>
<tr>
<th>$\delta_2$</th>
<th>ELW</th>
<th>Tapered Estimator</th>
<th>Stationary LW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias</td>
<td>$\delta_1$</td>
<td>$\delta_2$</td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>-0.004</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.085</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.085</td>
<td>0.085</td>
</tr>
<tr>
<td>0.8</td>
<td>bias</td>
<td>-0.013</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.077</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.078</td>
<td>0.072</td>
</tr>
<tr>
<td>1.3</td>
<td>bias</td>
<td>-0.015</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.073</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.075</td>
<td>0.070</td>
</tr>
</tbody>
</table>

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 $\rho = 0.8$

<table>
<thead>
<tr>
<th></th>
<th>ELW</th>
<th>Tapered Estimator</th>
<th>Stationary LW</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2 = 0.4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.001</td>
<td>-0.002</td>
<td>-8.737</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.078</td>
<td>0.078</td>
<td>75.80</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.078</td>
<td>0.078</td>
<td>76.30</td>
</tr>
<tr>
<td>$\delta_2 = 0.8$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>-0.009</td>
<td>-0.005</td>
<td>-0.010</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.064</td>
<td>0.064</td>
<td>1.604</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.064</td>
<td>0.064</td>
<td>1.605</td>
</tr>
<tr>
<td>$\delta_2 = 1.3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>-0.010</td>
<td>-0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.061</td>
<td>0.060</td>
<td>0.002</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.061</td>
<td>0.060</td>
<td>0.002</td>
</tr>
</tbody>
</table>

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 $\rho = 0.4$

<table>
<thead>
<tr>
<th></th>
<th>ELW</th>
<th>Tapered Estimator</th>
<th>Stationary LW</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2 = 0.4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.002</td>
<td>0.001</td>
<td>-224.4</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.086</td>
<td>0.082</td>
<td>675.6</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.086</td>
<td>0.082</td>
<td>711.9</td>
</tr>
<tr>
<td>$\delta_2 = 0.8$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>-0.011</td>
<td>-0.001</td>
<td>-0.852</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.080</td>
<td>0.076</td>
<td>27.28</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.080</td>
<td>0.076</td>
<td>27.79</td>
</tr>
<tr>
<td>$\delta_2 = 1.3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>-0.014</td>
<td>-0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>s.d.</td>
<td>0.074</td>
<td>0.070</td>
<td>0.005</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.075</td>
<td>0.070</td>
<td>0.005</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>ELW</th>
<th></th>
<th>Tapered Estimator</th>
<th></th>
<th>Stationary LW</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\delta_1, \delta_2))</td>
<td>bias</td>
<td>(-0.003)</td>
<td>(-0.001)</td>
<td>(-1.167)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>((0.1, 0.4))</td>
<td>s.d.</td>
<td>0.085</td>
<td>0.083</td>
<td>9.664</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.085</td>
<td>0.083</td>
<td>9.735</td>
<td>0.250</td>
</tr>
<tr>
<td>((\delta_1, \delta_2))</td>
<td>bias</td>
<td>0.000</td>
<td>(-0.002)</td>
<td>(-8.311)</td>
<td>(-0.036)</td>
</tr>
<tr>
<td>((0.3, 0.4))</td>
<td>s.d.</td>
<td>0.080</td>
<td>0.079</td>
<td>21.11</td>
<td>0.545</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.080</td>
<td>0.079</td>
<td>22.69</td>
<td>0.546</td>
</tr>
<tr>
<td>((\delta_1, \delta_2))</td>
<td>bias</td>
<td>(-0.011)</td>
<td>(-0.001)</td>
<td>(-0.049)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>((0.3, 0.8))</td>
<td>s.d.</td>
<td>0.080</td>
<td>0.075</td>
<td>2.859</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.081</td>
<td>0.075</td>
<td>2.859</td>
<td>0.121</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th></th>
<th>ELW</th>
<th></th>
<th>Tapered Estimator</th>
<th></th>
<th>Stationary LW</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta_1, \delta_2) = (\delta_1) = 0.1, (\delta_2) = 0.4</td>
<td>bias</td>
<td>(-0.003)</td>
<td>(-0.001)</td>
<td>(-16.25)</td>
<td>(-0.006)</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.082</td>
<td>0.077</td>
<td>201.3</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.082</td>
<td>0.077</td>
<td>202.0</td>
<td>0.337</td>
</tr>
<tr>
<td>(\delta_1, \delta_2) = (\delta_1) = 0.1, (\delta_2) = 0.8</td>
<td>bias</td>
<td>(-0.012)</td>
<td>(-0.002)</td>
<td>(0.015)</td>
<td>(-0.013)</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.073</td>
<td>0.068</td>
<td>0.862</td>
<td>0.264</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.074</td>
<td>0.068</td>
<td>0.862</td>
<td>0.264</td>
</tr>
<tr>
<td>(\delta_1, \delta_2) = (\delta_1) = 0.1, (\delta_2) = 1.3</td>
<td>bias</td>
<td>(-0.015)</td>
<td>(-0.004)</td>
<td>(0.000)</td>
<td>(-0.012)</td>
</tr>
<tr>
<td></td>
<td>s.d.</td>
<td>0.070</td>
<td>0.066</td>
<td>0.003</td>
<td>0.261</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.072</td>
<td>0.066</td>
<td>0.003</td>
<td>0.261</td>
</tr>
</tbody>
</table>

Note: The sample size and bandwidth are \(n = 512\) and \(m = n^{0.65} = 57\), respectively. \(\delta_1\) and \(\rho\) are set to \(\delta_1 = 0.1\) and \(\rho = 0.4\).
Table 7: Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied volatility ($\sigma^I$)</td>
<td>0.168</td>
<td>0.066</td>
<td>1.624</td>
<td>7.077</td>
</tr>
<tr>
<td>Realized volatility ($\sigma^R$)</td>
<td>0.159</td>
<td>0.097</td>
<td>2.807</td>
<td>14.98</td>
</tr>
</tbody>
</table>

Table 8: Univariate two-step ELW estimates of $\delta$

<table>
<thead>
<tr>
<th>$m$</th>
<th>$n^{0.35} = 20$</th>
<th>$n^{0.6} = 26$</th>
<th>$n^{0.65} = 35$</th>
<th>$n^{0.7} = 46$</th>
<th>$n^{0.75} = 60$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^I$</td>
<td>0.572</td>
<td>0.554</td>
<td>0.634</td>
<td>0.628</td>
<td>0.645</td>
</tr>
<tr>
<td></td>
<td>(0.353, 0.792)</td>
<td>(0.362, 0.747)</td>
<td>(0.468, 0.800)</td>
<td>(0.483, 0.772)</td>
<td>(0.518, 0.772)</td>
</tr>
<tr>
<td>$\sigma^R$</td>
<td>0.512</td>
<td>0.480</td>
<td>0.550</td>
<td>0.561</td>
<td>0.609</td>
</tr>
<tr>
<td></td>
<td>(0.293, 0.731)</td>
<td>(0.288, 0.672)</td>
<td>(0.384, 0.715)</td>
<td>(0.417, 0.706)</td>
<td>(0.482, 0.735)</td>
</tr>
<tr>
<td>$\sigma^R - \sigma^I$</td>
<td>0.250</td>
<td>0.246</td>
<td>0.319</td>
<td>0.377</td>
<td>0.457</td>
</tr>
<tr>
<td></td>
<td>(0.031, 0.469)</td>
<td>(0.054, 0.438)</td>
<td>(0.153, 0.484)</td>
<td>(0.233, 0.522)</td>
<td>(0.330, 0.583)</td>
</tr>
</tbody>
</table>

Note: Asymptotic 95% confidence intervals are in parentheses.

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

<table>
<thead>
<tr>
<th>$m$</th>
<th>$n^{0.35} = 20$</th>
<th>$n^{0.6} = 26$</th>
<th>$n^{0.65} = 35$</th>
<th>$n^{0.7} = 46$</th>
<th>$n^{0.75} = 60$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_1$</td>
<td>0.208</td>
<td>0.220</td>
<td>0.262</td>
<td>0.313</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>(0.015, 0.402)</td>
<td>(0.050, 0.390)</td>
<td>(0.116, 0.407)</td>
<td>(0.182, 0.444)</td>
<td>(0.264, 0.499)</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.619</td>
<td>0.675</td>
<td>0.675</td>
<td>0.644</td>
<td>0.642</td>
</tr>
<tr>
<td></td>
<td>(0.425, 0.812)</td>
<td>(0.505, 0.845)</td>
<td>(0.530, 0.821)</td>
<td>(0.513, 0.775)</td>
<td>(0.524, 0.760)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.052</td>
<td>1.107</td>
<td>1.045</td>
<td>1.032</td>
<td>1.039</td>
</tr>
<tr>
<td></td>
<td>(0.925, 1.180)</td>
<td>(1.031, 1.183)</td>
<td>(0.930, 1.160)</td>
<td>(0.851, 1.213)</td>
<td>(0.791, 1.287)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.609</td>
<td>0.594</td>
<td>0.621</td>
<td>0.554</td>
<td>0.487</td>
</tr>
<tr>
<td>$\beta_{NB}$</td>
<td>1.309</td>
<td>1.318</td>
<td>1.325</td>
<td>1.330</td>
<td>1.330</td>
</tr>
</tbody>
</table>

Note: Asymptotic 95% confidence intervals are in parentheses.