Queen's Economics Department Working Paper No. 1062

# Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes

Katsumi Shimotsu
Queen's University

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

2-2006

# Gaussian semiparametric estimation of multivariate fractionally integrated processes

Katsumi Shimotsu*
*Department of Economics*
*Queen's University*

February 1, 2006

**Abstract**

This paper analyzes the semiparametric estimation of multivariate long-range dependent processes. The class of spectral densities considered is motivated by and includes those of multivariate fractionally integrated processes. The paper establishes the consistency of the multivariate Gaussian semiparametric estimator (GSE), which has not been shown in other work, and the asymptotic normality of the GSE estimator. The proposed GSE estimator is shown to have a smaller limiting variance than the two-step GSE estimator studied by Lobato (1999). Gaussianity is not assumed in the asymptotic theory. Some simulations confirm the relevance of the asymptotic results in samples of the size used in practical work.

*JEL Classification*: C22

*Key words and phrases*: Fractional integration, long memory, semiparametric estimation.

*Address: Department of Economics, Queen's University, Kingston, Ontario K7L 3N6, Canada. E-mail: shimotsu@qed.econ.queensu.ca. Phone: +1-613-533-6546. Fax: +1-613-533-6668.

# 1 Introduction

Consider a real-valued covariance stationary $q$-vector process $X_t$ that is generated by

$$\begin{pmatrix} (1-L)^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-L)^{d_q} \end{pmatrix} \begin{pmatrix} X_{1t}-EX_{1t} \\ \vdots \\ X_{qt}-EX_{qt} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{qt} \end{pmatrix}, \quad -\frac{1}{2} < d_1, \ldots, d_q < \frac{1}{2}, \tag{1}$$

where $u_t = (u_{1t}, \ldots u_{qt})'$ is a covariance stationary process whose spectral density $f_u(\lambda)$ is bounded and bounded away from zero (in the sense of positive definite matrices) at the zero frequency $\lambda = 0$. This is a multivariate extension of a scalar fractionally integrated process (the so-called $I(d)$ process), and the time series $X_{at}$ exhibits long-range dependence whenever $d_a > 0$. $X_t$ becomes a multivariate ARFIMA process when $u_t$ is a vector ARMA process, but the specification (1) does not require $u_t$ to be of this or any other parametric form.

Fractionally integrated processes have a time domain representation that naturally extends conventional ARMA models and are the most widely used long-range dependent time series in econometrics. The relationship between the value of the memory parameter and the persistence of a shock is easily understood in terms of the coefficient in the expansion (albeit this is only formal for $d > 0$)

$$(1-L)^{-d} = \sum_{k=0}^{\infty} \frac{\Gamma(d+k)}{\Gamma(d)k!} L^k,$$

where $\Gamma$ is the gamma function. Discussion and examples of recent empirical applications of fractional integration are found in, e.g., Bollerslev and Wright (2000), Brunetti and Gilbert (2000), and Henry and Zaffaroni (2003).

Let $f(\lambda)$ denote the spectral density of $X_t$, so that

$$E(X_t - EX_t)(X'_{t+k} - EX'_t) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda.$$

Define $\Phi(\lambda) = \text{diag}((1-e^{i\lambda})^{-d_1}, \cdots, (1-e^{i\lambda})^{-d_q})$, then $f(\lambda) = \Phi(\lambda) f_u(\lambda) \Phi^*(\lambda)$ (e.g., Hannan, 1970, p.61). The memory parameters, $d_a$, govern the long-run dynamics of the process and the behavior of $f(\lambda)$ around the origin. Therefore, if empirical interest lies in the long-run dynamics of the process, it is useful to specify the spectral density only locally in the vicinity of the origin and avoid specifying the short-run dynamics of $u_t$ explicitly. How this is done turns out to be a matter of some importance.

Assume $f_u(\lambda)$ satisfies the local condition

$$f_u(\lambda) \sim G, \quad \lambda \to 0,$$

where $G$ is a real, symmetric, finite, and positive definite matrix. This will be so for any $u_t$ having Wold representation $u_t = C(L)\varepsilon_t$ with $C(1)$ finite and of full rank. Since

$$(1-e^{i\lambda})^{\alpha} = \lambda^{\alpha} e^{-i\pi\alpha/2}(1 + O(\lambda)), \quad \lambda \to 0+, \tag{2}$$

(Lobato, 1997; Phillips and Shimotsu, 2004), it follows that

$$f(\lambda) \sim \operatorname{diag}(\lambda^{-d_a} e^{i\pi d_a/2}) G \operatorname{diag}(\lambda^{-d_a} e^{-i\pi d_a/2}), \quad \lambda \to 0+. \tag{3}$$

If $f_u(\lambda) = G(1 + O(\lambda^\beta))$ as $\lambda \to 0$, then a more refined local approximation to the spectral matrix at the origin is given by

$$f(\lambda) \sim \operatorname{diag}(\lambda^{-d_a} e^{i(\pi-\lambda)d_a/2}) G \operatorname{diag}(\lambda^{-d_a} e^{-i(\pi-\lambda)d_a/2}) \left[1 + O(\lambda^{\min\{\beta,2\}})\right], \quad \lambda \to 0+, \tag{4}$$

as shown in section 2.

When $f(\lambda)$ is specified locally as (3) or (4), estimation of $d_a$ is semiparametric and only uses information on the long-run dynamics of the process. Semiparametric estimators are robust to misspecification of the short-run dynamics, because they are agnostic to the behavior of the spectrum away from the origin.

In the univariate case where $f(\lambda) \sim G\lambda^{-2d}$ as $\lambda \to 0$, one attractive semiparametric estimator was proposed by Künsch (1987) and analyzed by Robinson (1995b). The estimator, a Gaussian semiparametric estimator (GSE), is based on the maximization of the frequency domain Gaussian likelihood function that is localized to the vicinity of the origin. The GSE generally has several advantages over other semiparametric estimators, including efficiency and weaker distributional assumptions. Lobato (1999) has already analyzed one version of the multivariate extension of GSE. His approach involves two-step estimation, which is based on a first-step univariate estimation of $d_1, \ldots, d_q$ and a Newton-type second step. Lobato shows asymptotic normality of this two-step estimator.

We consider semiparametric estimation of $d$ when the spectral density has the general local form given in (3) or (4). The specification (3) extends the local specification of the scalar spectrum $f(\lambda) \sim G\lambda^{-2d}$ to the multivariate case. It includes multivariate fractionally integrated processes and is general enough to accommodate the presence of poles and zeros at frequencies away from the origin (Phillips and Shimotsu, 2004, provide an example of the latter). In (3), the memory parameters appear in the two factors $\lambda^{-d_a}$ and $e^{i\pi d_a/2}$, and hence the estimation of the $d_a$ needs to take both $\lambda^{-d_a}$ and $e^{i\pi d_a/2}$ into account. The representation (4) involves both $\lambda^{-d_a}$ and $e^{i(\pi-\lambda)d_a/2}$, so that there is an additional linear factor in the complex exponential. These additional dependencies on the memory parameter in the multivariate spectrum make the analysis more difficult but utilize the correct specification of the spectral matrix around the origin. Lobato (1999) considered semiparametric estimation of $d$ from the following simpler alternate local form of spectral density[1]

$$\widetilde{f}(\lambda) \sim \operatorname{diag}(\lambda^{-d_a}) G \operatorname{diag}(\lambda^{-d_a}), \quad \lambda \to 0. \tag{5}$$

When $X_t$ is generated by a multivariate fractionally integrated process such as (1), however, estimation based on the specification (5) cannot provide efficient estimates

[1]The specification (5) is also used in Lobato and Robinson (1998) to construct a nonparametric test for weak dependence. Lobato and Velasco (2000) extend it to analyze the two-step Gaussian semiparametric estimation of multivariate nonstationary long-range dependent processes.

of $d_a$. This is because the off-diagonal elements of the spectral matrix of $X_t$ have a nonnegligible imaginary part in the neighborhood of the origin and thus a complex asymptote at $\lambda = 0$, as is clear from (4). So, $\widetilde{f}(\lambda)$ in (5) does not belong to the class of spectral densities specified in (3) or (4). Indeed, we are not aware of any physically realizable time domain model of multivariate time series whose spectral density follows (5), except those special cases where $G$ itself is diagonal (which implies there is no long run covariance between the elements of $X_t$), or where $d_a = d$ for all $a$ (in which case $\Phi(\lambda) = (1 - e^{i\lambda})^{-d} I_q$ and the long range dependence is identical across components). In general, when $G$ has nonzero off-diagonal elements, $f(\lambda)$ has complex off-diagonal elements involving $d_a$. In particular, the phase spectrum of $X_{at}$ and $X_{bt}$ is nonzero (and depends on $d_a$ and $d_b$) even at the zero frequency. This means that different memory patterns in $X_{at}$ and $X_{bt}$ induce phase shifts in the cross spectrum of these variables at the origin. Since there is information in the phase patterns of the data about memory, taking the correct local form (3) into account in GSE estimation should improve the efficiency of estimation. The results of this paper show this to be so and indicate that the impact on efficiency can be significant.

We also prove the consistency of our multivariate GSE. Two-step estimation is partly motivated by its computational ease, because a two-step estimation is faster in general than a high dimensional direct minimization. However, in view of modern computational resources, a direct minimization of the objective function with respect to the $q$ memory parameters is not likely to cause any practical difficulty. Indeed, the simulation in this study confirms it. Some direct minimization methods such as Nelder-Mead simplex algorithm dispense with numerical/analytical derivatives which are necessary for the evaluation of the score function and Hessian. Although the proof of the consistency of univariate GSE by Robinson (1995b) is not directly applicable to the multivariate case, a proper modification of this proof enables us to handle the nonuniform convergence of the objective function and establish the consistency of the multivariate GSE.

The GSE is shown to have a Gaussian limiting distribution. As anticipated, the limiting variance is different from, and smaller than, that of the GSE analyzed by Lobato (1999). As indicated above, the gain in efficiency arises because both real and imaginary parts of the spectral density and periodograms are utilized and the presence of $d_a$ in the factor $e^{i(\pi-\lambda)d_a/2}$ in (4) provides additional information on $d$. Simulations with multivariate fractionally integrated processes confirm this increase in efficiency in finite samples. In addition, we prove the consistency and asymptotic normality of the GSE of Lobato (1999) under (3) and show its limiting variance is different from the one derived in Lobato (1999).

The remainder of the paper is organized as follows. Section 2 describes the GSE. Consistency of the GSE is demonstrated in Section 3, and Section 4 derives its limiting distribution. Section 5 shows the consistency and asymptotic normality of the GSE of Lobato (1999) under (3) and compares it with our GSE. Section 6 reports some simulation results, and Section 7 concludes. Proofs are given in Appendix A in Section 8. Some technical results are collected in Appendix B in Section 9.

## 2 Multivariate semiparametric estimation

We consider semiparametric estimation of $d = (d_1, \cdots, d_q)'$, which uses only Fourier frequencies in the neighborhood of the origin and hence is nonparametric with respect to short-run dynamics of the data. Define the discrete Fourier transform (dft) and the periodogram of $X_t$ evaluated at frequency $\lambda$ as

$$w(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda}, \quad I(\lambda) = w(\lambda) w^*(\lambda),$$

where $x^*$ denotes the conjugate transpose of $x$. For the reason explained in Section 3, it is useful to consider a local approximation at the origin that is finer than that given in (3). Since $|1 - e^{i\lambda}| = |2\sin(\lambda/2)|$ and $arg(1 - e^{i\lambda}) = (\lambda - \pi)/2$ for $0 \le \lambda < \pi$, we have

$$\begin{aligned} (1 - e^{i\lambda})^\theta &= (|2\sin(\lambda/2)|)^\theta \exp[i(\lambda - \pi)\theta/2] \\ &= \lambda^\theta \exp[i(\lambda - \pi)\theta/2](1 + O(\lambda^2)). \end{aligned}$$

This is merely a refinement of (2), but the smaller error magnitude ($O(\lambda^2)$) will become essential in the analysis in Section 4. Since $f_u(\lambda) \sim G$ as $\lambda \to 0$, we have, for the Fourier frequencies $\lambda_j = 2\pi j/n$ with $j = 1, \ldots, m$ and $m = o(n)$,

$$f(\lambda_j) \sim \Lambda_j(d) G \Lambda_j^*(d), \quad \Lambda_j(d) = \mathrm{diag}\left(\Lambda_{ja}(d)\right); \quad \Lambda_{ja}(d) = \lambda_j^{-d_a} e^{i(\pi - \lambda_j)d_a/2}. \tag{6}$$

Therefore, the Gaussian log-likelihood function localized to the origin is

$$\begin{aligned} Q_m(G, d) &= \frac{1}{m} \sum_{j=1}^{m} \left\{ \log\det \Lambda_j(d) G \Lambda_j^*(d) + \mathrm{tr}\left[ \left( \Lambda_j(d) G \Lambda_j^*(d) \right)^{-1} I(\lambda_j) \right] \right\} \\ &= \frac{1}{m} \sum_{j=1}^{m} \left\{ \log\det \Lambda_j(d) G \Lambda_j^*(d) + \mathrm{tr}\left[ G^{-1} \mathrm{Re}\left[ \Lambda_j(d)^{-1} I(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right] \right\}, \end{aligned}$$

where the second line follows because both $Q_m(G, d)$ and $G$ are real. Using the fact that $\det AB = \det A \det B$ for any complex matrices $A$ and $B$, the first order condition with respect to $G$ gives

$$G = \frac{1}{m} \sum_{j=1}^{m} \mathrm{Re}[\Lambda_j(d)^{-1} I(\lambda_j) \Lambda_j^*(d)^{-1}].$$

Substituting this into $Q_m(G, d)$ in conjunction with the fact that

$$\begin{aligned} & \log\det \Lambda_j(d) + \log\det \Lambda_j^*(d) \\ = \; & \log\det \Lambda_j(d) \Lambda_j^*(d) = \log(\mathrm{diag}(\lambda_j^{-2d_a})) = -2 \sum_{a=1}^{q} d_a \log \lambda_j, \end{aligned}$$

we obtain the objective function

$$
\begin{aligned}
R(d) &= \log\det \widehat{G}(d) - 2\sum_{a=1}^{q} d_a \frac{1}{m}\sum_{j=1}^{m}\log\lambda_j, \\
\widehat{G}(d) &= \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[\Lambda_j(d)^{-1} I(\lambda_j)\Lambda_j^*(d)^{-1}\right].
\end{aligned}
$$

In the following, we denote the true parameter values by $G^0$ and $d^0$. The estimator is defined as

$$
\widehat{d} = \arg\min_{d\in\Theta} R(d),
$$

where the space of admissible estimates of $d^0$, $\Theta$, takes the form $\Theta = [\Delta_1, \Delta_2]^q$, with $-1/2 < \Delta_1 < \Delta_2 < 1/2$.

# 3 Consistency of the estimator

We now introduce the assumptions on $m$ and $f(\lambda)$ needed for the consistency of the estimator. Let $f_{ab}(\lambda)$ and $G^0_{ab}$ denote the $(a,b)$ th element of $f(\lambda)$ and $G^0$, respectively.

**Assumption 1** *As* $\lambda \to 0+$,

$$
f_{ab}(\lambda) = e^{i\pi(d_a^0 - d_b^0)/2} G^0_{ab} \lambda^{-d_a^0 - d_b^0} + o(\lambda^{-d_a^0 - d_b^0}), \quad a, b = 1, \ldots, q.
$$

**Assumption 2**

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

*where* $||\cdot||$ *denotes the supremum norm and* $E(\varepsilon_t|F_{t-1}) = 0$, $E(\varepsilon_t\varepsilon_t'|F_{t-1}) = I_q$ *a.s.,* $t = 0, \pm 1, \ldots$, *in which* $F_t$ *is the* $\sigma$*-field generated by* $\varepsilon_s$, $s \le t$, *and there exists a scalar random variable* $\varepsilon$ *such that* $E\varepsilon^2 < \infty$ *and for all* $\eta > 0$ *and some* $K > 0$, $\Pr(||\varepsilon_t||^2 > \eta) \le K \Pr(\varepsilon^2 > \eta)$.

**Assumption 3** *In a neighborhood* $(0, \delta)$ *of the origin,* $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ *is differentiable and*

$$
\frac{\partial}{\partial\lambda} A_a(\lambda) = O(\lambda^{-1}||A_a(\lambda)||) \quad \text{as } \lambda \to 0+,
$$

*where* $A_a(\lambda)$ *is the a'th row of* $A(\lambda)$.

**Assumption 4** *As* $n \to \infty$,

$$\frac{1}{m} + \frac{m}{n} \to 0.$$

Assumptions 1-4 are multivariate extensions of Assumptions A1-A4 of Robinson (1995b) and analogous to the ones used in Robinson (1995a) and Lobato (1999). In Assumption 1, replacing $e^{i\pi(d_a^0-d_b^0)/2}$ with $e^{i(\pi-\lambda)(d_a^0-d_b^0)/2}$ does not make a difference because $e^{i\lambda}-1=o(1)$. Assumption 3 implies $\partial A_a(\lambda)/\partial\lambda = O(\lambda^{-d_a-1})$, because $||A_a(\lambda)|| \leq (A_a(\lambda)A_a^*(\lambda))^{1/2} = (2\pi f_{aa}(\lambda))^{1/2}$.

Under these conditions, we may now establish the consistency of $\widehat{d}$.

**Theorem 1** *Let Assumptions 1-4 hold. Then, for* $d_0 \in \Theta$, $\widehat{d} \to_p d^0$ *as* $n \to \infty$.

# 4 Asymptotic normality of the estimator

We introduce some further assumptions that are used in the results of this section. They are analogous to the assumptions in Lobato (1999).

**Assumption 1′** *For* $\beta \in (0,2]$ *and* $a,b = 1,\ldots,q$,

$$f_{ab}(\lambda) - e^{i(\pi-\lambda)(d_a^0-d_b^0)/2}\lambda^{-d_a^0-d_b^0}G_{ab}^0 = O(\lambda^{-d_a^0-d_b^0+\beta}) \quad \text{as } \lambda \to 0+.$$

**Assumption 2′** *Assumption 2 holds and also for* $a,b,c,d = 1,2$,

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}|F_{t-1}) = \mu_{abc} \quad a.s., \quad E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}|F_{t-1}) = \mu_{abcd}, \quad t = 0,\pm 1,\ldots,$$

*where* $|\mu_{abc}| < \infty$ *and* $|\mu_{abcd}| < \infty$.

**Assumption 3′** *Assumption 3 holds.*

**Assumption 4′** *As* $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{n^{2\beta}} + \frac{\log n}{m^\gamma} \to 0, \quad \text{for any } \gamma > 0.$$

**Assumption 5′** *There exists a finite real matrix* $H$ *such that*

$$\Lambda_j(d^0)^{-1}A(\lambda_j) = H + o(1), \quad \text{as } \lambda_j \to 0.$$

Assumption 1′ does not hold for $\beta > 1$ if we replace $e^{i(\pi-\lambda)(d_a^0-d_b^0)/2}$ with $e^{i\pi(d_a^0-d_b^0)/2}$, because $e^{i\lambda} = 1 + O(\lambda)$. Assumption 1′ is analogous to the ones used in Robinson (1995a) and Lobato (1999) and is satisfied by certain multivariate ARFIMA processes. See Robinson (1995a, p.1056) for further discussion. Assumption 4′ is slightly stronger than the assumptions in Robinson (1995b) and Lobato (1999), i.e., $m^{-1} + m^{1+2\beta}n^{-2\beta}(\log m)^2 \to 0$. It is satisfied if $m \sim Cn^\xi$ with a finite positive constant $C$ and $0 < \xi < 2\beta/(1+2\beta)$. The third term on the left hand side of Assumption

$4'$ is necessary in establishing the convergence of the Hessian. Assumption $5'$ complements Assumption $1'$ in that it controls the degree of approximation of the transfer function by $\Lambda_j(d^0)$. This assumption obviously implies $HH' = 2\pi G^0$ and is satisfied by multivariate ARFIMA models.

**Theorem 2** *Let Assumptions $1'$-$5'$ hold. Then, for $d^0 \in Int(\Theta)$, as $n \to \infty$,*

$$m^{1/2}\left(\widehat{d} - d_0\right) \to_d N\left(0, \Omega^{-1}\right), \quad \Omega = 2\left[G^0 \odot (G^0)^{-1} + I_q + \frac{\pi^2}{4}\left(G^0 \odot (G^0)^{-1} - I_q\right)\right],$$
$$\widehat{G}(\widehat{d}) \to_p G^0,$$

*where $\odot$ denotes the Hadamard product.*

## 5 Comparison with the estimator of Lobato (1999)

Lobato (1999) analyzes the two-step GSE that uses the objective function based on (5):

$$\widetilde{d} = \arg\min_{d\in\Theta} \widetilde{R}(d),$$

where

$$\widetilde{R}(d) = \log\det \widetilde{G}(d) - 2\sum_{a=1}^{q} d_a \frac{1}{m}\sum_{j=1}^{m}\log\lambda_j, \quad \widetilde{G}(d) = \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[\mathrm{diag}(\lambda_j^{d_a}) I(\lambda_j)\mathrm{diag}(\lambda_j^{d_a})\right],$$

and shows that, when the spectral density of $X_t$ follows (5), $m^{1/2}(\widetilde{d} - d_0) \to_d N(0, \Xi^{-1})$, where $\Xi = 2[G^0 \odot (G^0)^{-1} + I_q]$. Because $G^0 \odot (G^0)^{-1} - I_q$ is positive semidefinite (Horn and Johnson, 1985, p. 475), $\widehat{d}$ has a smaller (in a matrix sense) limiting variance matrix than $\widetilde{d}$ except when $G^0$ is diagonal.

The following Theorem establishes the asymptotic behavior of $\widetilde{d}$ under the Assumptions 1-4 and $1'$-$5'$. Intriguingly, $\widetilde{d}$ is consistent and asymptotically normal despite being based on a misspecified model (5). Define

$$\mathcal{E}_0 = \mathrm{diag}(e^{i\pi d_a^0/2}), \quad \widetilde{G}^0 = \mathrm{Re}\left[\mathcal{E}_0 G^0 \mathcal{E}_0^*\right], \quad \overline{G}^0 = \mathrm{Im}\left[\mathcal{E}_0 G^0 \mathcal{E}_0^*\right].$$

**Theorem 3** *(a) Let Assumptions 1-4 hold. Then, for $d_0 \in \Theta$, $\widetilde{d} \to_p d_0$ as $n \to \infty$. (b) Let Assumptions $1'$-$5'$ hold and assume $m^3 n^{-2}(\log m)^2 \to 0$. Then, for $d^0 \in Int(\Theta)$, as $n \to \infty$,*

$$m^{1/2}\left(\widetilde{d} - d_0\right) \to_d N(0, Q), \quad Q = \widetilde{\Omega}^{-1}\Upsilon\widetilde{\Omega}^{-1},$$
$$\widetilde{\Omega} = 2\left[\widetilde{G}^0 \odot (\widetilde{G}^0)^{-1} + I_q\right],$$
$$\Upsilon = 2\left[\widetilde{G}^0 \odot (\widetilde{G}^0)^{-1} + I_q\right] + 2\left((\widetilde{G}^0)^{-1}\,\overline{G}^0(\widetilde{G}^0)^{-1}\right) \odot \overline{G}^0$$
$$-2\left((\widetilde{G}^0)^{-1}\overline{G}^0\right) \odot \left((\widetilde{G}^0)^{-1}\overline{G}^0\right)',$$
$$\widetilde{G}(\widetilde{d}) \to_p \widetilde{G}^0.$$

The additional assumption $m^3 n^{-2}(\log m)^2 \to 0$ is necessary because the misspecification of the true spectral density by (5) involves the term $e^{i\lambda_j(d_a^0 - d_b^0)/2} = O(\lambda_j)$. Misspecificaton of the true spectral density does not affect the consistency of $\widetilde{d}$, but $\widetilde{G}(\widetilde{d})$ converges to $\widetilde{G}^0$ and hence is an inconsistent estimate of $G^0$. Since the $(a,b)$th element of $\widetilde{G}^0$ is $G_{ab}^0 \cos(\pi(d_a - d_b)/2)$, $\widetilde{G}(\widetilde{d})$ underestimates the off-diagonal elements $G^0$.

The asymptotic variance of $\widetilde{d}$ takes an involved form. An interesting special case is $d_1^0 = \cdots = d_q^0$, whence $\widetilde{G}^0 = G^0$, $\overline{G}^0 = 0$, and $Q$ reduces to $\Xi^{-1}$. In this case, (3) coincides with (5) under the true dgp. However, if not all $d_a$ are the same, the spectral density has a complex part, which provides an additional source of identification. $\widetilde{d}$ fails to take into account its presence and hence is less efficient than $\widehat{d}$. In more general cases where $d_a^0$ are not the same across all $a$, both $\widetilde{\Omega}$ and $\Upsilon$ depend on the value of $d^0$ and an explicit analytic comparison between $\Omega^{-1}$ and $Q$ is not available. A small numerical evaluation and the simulation evidence below indicate that $\widehat{d}$ is more efficient than $\widetilde{d}$, which comes as no surprise since $\widehat{d}$ is based on the correct specification.

We compare the diagonal elements of $\Omega^{-1}$ and $Q$ with the asymptotic variance of the univariate GSE ($= 0.25$) when $q = 2$. $G^0$ is chosen to be

$$G^0 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad \rho = 0.0, 0.2, 0.4, 0.6, 0.8.$$

$|d_1 - d_2|$ is set to 0.0, 0.2, and 0.4. The value of $Q$ depends on $d$ only via $|d_1 - d_2|$, and $\Omega$ does not depend on $d$. Table 1 reports $(\Omega^{-1})_{11}$ and $Q_{11}$ and their ratio to 0.25.

When $\rho \leq 0.2$, the variance of the three estimators is not substantially different. When $\rho \geq 0.4$, both $(\Omega^{-1})_{11}$ and $Q_{11}$ are noticeably smaller than 0.25, and they decrease as $\rho$ increases. $Q_{11}$ is always larger than $(\Omega^{-1})_{11}$ and approaches 0.25 as $|d_1 - d_2|$ increases, but it is always smaller than 0.25. Therefore, we may expect a nonnegligible gain in efficiency from estimating the elements of $d$ jointly, and the gain may be substantial, especially when both real and imaginary parts of the spectral density are taken into account.

## 6 Simulations

This section reports some simulations that were conducted to examine the finite sample performance of $\widehat{d}$ (hereafter GSE1) and $\widetilde{d}$ (hereafter GSE2). The sample size and band parameter $m$ were chosen to be $n = 128$, $512$ and $m = n^{0.65}$, and the statistics in the tables were computed using 10,000 replications. We generate $X_t$ by truncating the infinite order moving average representation of (1):

$$\begin{aligned} X_t &= \begin{pmatrix} (1-L)^{-d_1} & 0 \\ 0 & (1-L)^{-d_2} \end{pmatrix} \left[ \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} I\{t \geq 1\} \right], \\ \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} &\sim iidN\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \end{aligned}$$

$n$+2,000 observations of $X_t$ were generated, and the first 2,000 observations were discarded. Minimization of the objective function is carried out using the Nelder-Mead simplex method.

Tables 2-4 show the bias, standard deviation, and root mean squared error (RMSE) of both estimators with $n = 512$ and the ratio of their variance to the variance of the univariate GSE. The values of $d$ were chosen to be $(0.2, -0.2)$, $(0.2, 0.2)$, and $(0.2, 0.4)$. Three values of $\rho$ were used; $\rho = 0$, 0.4, 0.8. Table 2 shows the results for $\rho = 0$. Both GSE1 and GSE2 have little bias for all values of $d$. The standard deviation and RMSE of GSE1 are slightly higher than those of GSE2. The limiting variance of the two estimators is the same, and the simulation results appear to corroborate it. The bias, standard deviation, and RMSE do not appear to be affected by the value of $d$. Table 3 shows the results for $\rho = 0.4$. GSE1 has a smaller standard deviation and RMSE than GSE2. The variance of GSE1 is not affected by the value of $d$, while the variance of GSE2 increases as $|d_1 - d_2|$ increases. Table 4 shows the results for $\rho = 0.8$. Both GSE1 and GSE2 have smaller standard deviations than the case when $\rho = 0.4$. In Tables 2-4, the ratio of the variance of GSE1 and GSE2 to that of the univariate GSE is close to its theoretical value given in Table 1. A simulation for a single set of $(d_1, d_2, \rho)$ with 10,000 replications took around 30 minutes on a PC with a dual 2.0 Ghz CPU running the Linux operating system, so a direct minimization by the simplex algorithm did not cause computational problems.

Tables 5 and 6 compare the two-step version of the GSEs with those obtained by direct minimization by the Newton-Raphson method with line search for selected values of $d$ and $\rho$. The two-step estimator is computed by taking a Newton step from the first-stage estimates, which are obtained by the univariate GSE. Analytical derivatives are used for both two-step and NR GSEs. For GSE1 with $n = 128$ and $\rho = 0.8$, the estimator computed by direct minimization has a substantially smaller variance than its two-step counterpart. For all other cases, the two-step method and direct minimization give very similar performance. The results in Table 6 are also close to the corresponding results from the Nelder-Mead simplex method in Tables 3 and 4.

Table 7 compares the estimates of $2\pi G = [(1, \rho)' \ (\rho, 1)']$ by GSE1 and GSE2. From Theorem 3, $\widetilde{G}(\widetilde{d})_{12}$ will converge to $G_{12} \cos(\pi(d_1 - d_2)/2)$ instead of $G_{12}$, whereas the diagonal elements of $G$ are consistently estimated by GSE2. $\cos(\pi(d_1 - d_2)/2)$ takes the value of 0.81, 1.00, and 0.95 when $|d_1 - d_2|$ is 0.4, 0, and -0.2, respectively. The simulation results confirm the downward bias of GSE2 in the off-diagonal elements of $G$, although the bias is small except for $|d_1 - d_2| = 0.4$.

We examine the accuracy of asymptotic inference based on Theorem 2 by testing a hypothesis $H_0 : (d_1, d_2) = (d_1^0, d_2^0)$ by a Wald statistic:

$$W = m(\widehat{d} - d^0)'\widehat{\Omega}(\widehat{d} - d^0),$$

where $\widehat{\Omega}$ is obtained by replacing $G^0$ in the definition of $\Omega$ with $\widehat{G}(\widehat{d})$. In univariate GSE estimation, Hurvich and Chen (2000, p. 164) report that the finite sample variance of GSE estimators tends to exceed their asymptotic variance. Hurvich and Chen find replacing $m$ in the variance estimate by a number $c_m$ improves approximation, where $c_m$ is defined as[2]

$$c_m = \sum_{j=1}^{m} \nu_j^2, \quad \nu_j = \log \lambda_j - \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j.$$

Since $c_m/m \to 1$ as $m \to \infty$, this modification does not alter the asymptotic distribution of the test statistic. The modified Wald statistic takes the form

$$W_c = c_m(\widehat{d} - d^0)'\widehat{\Omega}(\widehat{d} - d^0).$$

Tables 8 and 9 report the rejection frequencies with 0.10, 0.05, and 0.01 asymptotic critical values for $n = 128$, 512 and various values of $(\rho, d_1, d_2)$. The unmodified Wald statistic $W$ always overrejects the null, and its size distortion is substantial, in particular when $n = 128$. The modified Wald statistic $W_c$ also overrejects, but its size distortion is much smaller than that of $W$, and it seems to have a reasonable size when $n = 512$. In view of the general overrejecting tendency of Wald tests, we may conclude that $W_c$ provides a good inferential tool when $n$ is not too small.

## 7 Concluding remarks

This paper analyzes the semiparametric estimation of multivariate long-range dependent processes. The class of spectral densities considered is motivated by and includes those of multivariate fractionally integrated processes.

This class of spectral densities has both real and complex parts even around the origin, and the memory parameter affects both the slope and phase of the spectral density around the origin. As a result, modeling this dependency correctly achieves the efficient estimation, while ignoring it results in misspecification.

A Gaussian semiparametric estimator (GSE) that takes this dependency into account is proposed. It is shown to be consistent and asymptotically normally distributed. Its limiting variance is independent of the memory parameter, with the potential for substantial efficiency gain over univariate estimation. The GSE that ignores the phase shift is still consistent and asymptotically normally distributed despite its misspecification. But it is less efficient than the GSE based on the correct specification, and its limiting variance depends on the memory parameter. Simulation results corroborate the asymptotic results, and a properly modified Wald statistic is shown to have a reasonable finite sample ($\simeq 500$) size.

This paper sheds light on the importance of and potential difficulty in extending the univariate semiparametric modeling and estimation of strongly dependent processes into a multivariate context.


## Acknowledgements

The author thanks the Editor, an Associate Editor and a referee for helpful and constructive comments, including the suggestion to conduct the asymptotic analysis


[2] Hurvich and Chen use $2\sin(\lambda_j/2)$ instead of $\lambda_j$, but their difference is small for $\lambda_j \sim 0$. For more details, see Hurvich and Chen (2000).

in Section 5. The author thanks Peter C. B. Phillips for a discussion that led to this paper and for valuable comments. The financial support from ESRC under Grant R000223986 to University of Essex (his former affiliation) is gratefully acknowledged.

# 8 Appendix A: Proofs

In this and the following sections, $C$ denotes a generic constant such that $C \in (1, \infty)$ unless specified otherwise, and it may take different values in different places.

## 8.1 Proof of Theorem 1

Define $\theta = (\theta_1, \cdots, \theta_q)' = d - d^0$ and $S(d) = R(d) - R(d^0)$. Fix $1/2 > \delta > 0$, and define $\overline{N}_\delta = \{d : ||d - d^0|| \geq \delta\}$, where $||\cdot||$ denotes the supremum norm. For arbitrary small $\Delta > 0$, define $\Theta_1 = \{\theta : \theta \in [-1/2 + \Delta, 1/2]^q\}$ and $\Theta_2 = \Theta \backslash \Theta_1$, possibly empty. Without loss of generality, assume $\Delta < 1/4$. Then we have (c.f. Robinson, 1995b, p. 1634)

$$\begin{aligned} \Pr\left(||\widehat{d} - d^0|| > \delta\right) &\leq \Pr\left(\inf_{\overline{N}_\delta \cap \Theta} S(d) \leq 0\right) \\ &\leq \Pr\left(\inf_{\overline{N}_\delta \cap \Theta_1} S(d) \leq 0\right) + \Pr\left(\inf_{\Theta_2} S(d) \leq 0\right). \end{aligned} \quad (7)$$

For the first probability on the right of (7), rewrite $S(d)$ as

$$\begin{aligned} S(d) &= \log \det \widehat{G}(d) - \log \det \widehat{G}(d^0) - 2\sum_{a=1}^{q} \theta_a \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j \\ &= \log \det \widehat{G}(d) + \log \left(\frac{2\pi m}{n}\right)^{-2(\theta_1 + \cdots + \theta_q)} - \log \det \widehat{G}(d^0) \\ &\quad -2\sum_{a=1}^{q} \theta_a \left(\frac{1}{m}\sum_{j=1}^{m} \log j - \log m\right) \\ &= \log A(d) - \log B(d) - \log A(d^0) + \log B(d^0) + S_2(d), \end{aligned}$$

where

$$\begin{aligned} A(d) &= \left(\frac{2\pi m}{n}\right)^{-2(\theta_1 + \cdots + \theta_q)} \det \widehat{G}(d), \quad B(d) = \prod_{a=1}^{q} (2\theta_a + 1)^{-1} \det G^0, \\ S_2(d) &= -2\sum_{a=1}^{q} \theta_a \left(\frac{1}{m}\sum_{j=1}^{m} \log j - \log m\right) - \sum_{a=1}^{q} \log(2\theta_a + 1). \end{aligned} \quad (8)$$

Since $m^{-1} \sum_1^m \log j - \log m + 1 = O(m^{-1} \log m)$ (see, e.g. Robinson, 1995b, Lemma 2), we have

$$S_2(d) = \sum_{a=1}^{q} [2\theta_a - \log(2\theta_a + 1)] + O(m^{-1} \log m).$$

Because $x - \log(x+1)$ achieves a unique global minimum on $(-1, \infty)$ at $x = 0$ and $x - \log(x+1) \geq x^2/6$ for $0 \leq |x| < 1$, for all sufficiently large $n$

$$\inf_{\overline{N}_\delta \cap \Theta_1} S_2(d) \geq \delta^2/8.$$

For $A(d)$ and $B(d)$, if there exists nonrandom $\Xi(d)$ such that

$$\text{(i)} \quad \sup_{\Theta_1} |A(d) - \Xi(d)| = o_p(1), \quad \text{(ii)} \quad \Xi(d) \geq B(d), \quad \text{(iii)} \quad \Xi(d^0) = B(d^0), \quad (9)$$

as $n \to \infty$, then, since $\inf_{\Theta_1} \Xi(d) \geq \inf_{\Theta_1} B(d) > 0$, we have, uniformly in $\Theta_1$,

$$\begin{aligned}
\log A(d) - \log B(d) &\geq \log A(d) - \log \Xi(d) \\
&= \log\left([\Xi(d) + o_p(1)]/\Xi(d)\right) = o_p(1), \\
\log A(d^0) - \log B(d^0) &= \log\left([\Xi(d^0) + o_p(1)]/\Xi(d^0)\right) = o_p(1),
\end{aligned}$$

and $\Pr(\inf_{\overline{N}_\delta \cap \Theta_1} S(d) \leq 0) \to 0$ follows.

We proceed to show (9). For (i), recall that $\Lambda_j(d)^{-1} = \text{diag}(\lambda_j^{d_a} e^{i(\lambda_j - \pi)d_a/2})$ and

$$\Lambda_j(d)^{-1} = \Lambda_j(d - d^0)^{-1}\Lambda_j(d^0)^{-1} = \Lambda_j(\theta)^{-1}\Lambda_j(d^0)^{-1}.$$

It follows that

$$\begin{aligned}
A(d) &= \left(\frac{2\pi m}{n}\right)^{-2(\theta_1 + \cdots + \theta_q)} \\
&\quad \times \det\left\{\frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[\Lambda_j(\theta)^{-1}\Lambda_j(d^0)^{-1} I(\lambda_j)\Lambda_j^*(d^0)^{-1}\Lambda_j^*(\theta)^{-1}\right]\right\} \\
&= \det\left\{\frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[M_j(\theta)\Lambda_j(d^0)^{-1} I(\lambda_j)\Lambda_j^*(d^0)^{-1} M_j^*(\theta)\right]\right\}, \qquad (10)
\end{aligned}$$

where $M_j(\theta) = \text{diag}(e^{i(\lambda_j - \pi)\theta_a/2}(j/m)^{\theta_a})$. Hereafter let $I_j$ denote $I(\lambda_j)$ and $w_{aj}$ denote $w_a(\lambda_j)$, the $a$th element of $w(\lambda_j)$. Observe that the $(a, b)$th element of the inside of $\det\{\cdot\}$ in (10) is (recall that $\Lambda_j(d) = \text{diag}(\Lambda_{ja}(d))$ with $\Lambda_{ja}(d) = \lambda_j^{-d_a} e^{i(\pi - \lambda_j)d_a/2}$, as defined in (6))

$$\frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[e^{i(\lambda_j - \pi)(\theta_a - \theta_b)/2}\left(\frac{j}{m}\right)^{\theta_a + \theta_b} \frac{w_{aj} w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)}\right].$$

Summation by parts (c.f., Robinson, 1995b, p. 1636) and Lemma 1 (a) give, uniformly in $(a,b)$,

$$\begin{aligned}
&\sup_{\Theta_1}\left|\frac{1}{m}\sum_{j=1}^{m} e^{i(\lambda_j-\pi)(\theta_a-\theta_b)/2}\left(\frac{j}{m}\right)^{\theta_a+\theta_b}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)}-G_{ab}^0\right)\right| \\
\leq\ & \frac{1}{m}\sum_{r=1}^{m-1}\sup_{\Theta_1}\left|\left(\frac{r}{m}\right)^{\theta_a+\theta_b} e^{i(\lambda_r-\pi)(\theta_a-\theta_b)/2}-\left(\frac{r+1}{m}\right)^{\theta_a+\theta_b} e^{i(\lambda_{r+1}-\pi)(\theta_a-\theta_b)/2}\right| \\
&\times\left|\sum_{j=1}^{r}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)}-G_{ab}^0\right)\right|+\left|\frac{1}{m}\sum_{j=1}^{m}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)}-G_{ab}^0\right)\right| \\
\leq\ & C\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2\Delta}\frac{1}{r^2}\left|\sum_{j=1}^{r}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)}-G_{ab}^0\right)\right| \\
&+\frac{1}{m}\left|\sum_{j=1}^{m}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)}-G_{ab}^0\right)\right| = o_p(1). \qquad (11)
\end{aligned}$$

It follows that, uniformly in $\Theta_1$,

$$\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[M_j(\theta)\Lambda_j(d^0)^{-1}I_j\Lambda_j^*(d^0)^{-1}M_j^*(\theta)\right]=\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[M_j(\theta)G^0M_j^*(\theta)\right]+o_p(1).$$

We proceed to derive an approximation of the right hand side. From Lemma 2 of Robinson (1995b), we have $\sup_{C\geq\gamma\geq\varepsilon}\left|\gamma m^{-1}\sum_{j=1}^{m}(j/m)^{\gamma-1}-1\right| = O\left(m^{-\varepsilon}\right)$ for $0<\varepsilon<C<\infty$. Also $e^{i(\lambda-\pi)(\theta_a-\theta_b)/2}=e^{-i\pi(\theta_a-\theta_b)/2}+O(\lambda)$. Define $\mathcal{E}\left(\theta\right)$ and $M_\infty(\theta)$ to be matrices whose $(a,b)$ elements are $e^{-i\pi(\theta_a-\theta_b)/2}$ and $(1+\theta_a+\theta_b)^{-1}=\int_0^1 x^{\theta_a+\theta_b}dx$, respectively. Then it follows that

$$\frac{1}{m}\sum_{j=1}^{m}\left[M_j(\theta)G^0M_j^*(\theta)\right]=\mathcal{E}\left(\theta\right)\odot M_\infty(\theta)\odot G^0+O(mn^{-1})+O(m^{-2\Delta}), \qquad (12)$$

where $\odot$ denotes a Hadamard product. Because the determinant is a continuous function of each element and the matrices $\mathcal{E}\left(\theta\right)$, $M_\infty(\theta)$, and $G^0$ are finite for $\theta\in\Theta_1$, (i) of (9) follows with

$$\Xi(d)=\det(\mathrm{Re}\left[\mathcal{E}\left(\theta\right)\right]\odot M_\infty(\theta)\odot G^0).$$

For (ii) and (iii) of (9), rewrite $\mathcal{E}\left(\theta\right)=\xi\xi^*$ with $\xi=(e^{-i\pi\theta_1/2},\cdots,e^{-i\pi\theta_q/2})'$. Then

$$\mathrm{Re}\left[\mathcal{E}\left(\theta\right)\right]=\mathrm{Re}\left(\xi\xi^*\right)=\mathrm{Re}\left[\xi\right]\left(\mathrm{Re}\left[\xi\right]\right)'+\mathrm{Im}\left[\xi\right]\left(\mathrm{Im}\left[\xi\right]\right)', \qquad (13)$$

and it follows that $\mathrm{Re}\left[\mathcal{E}\left(\theta\right)\right]$ is positive semidefinite. Since $M_\infty(\theta)$ and $G^0$ are positive semidefinite, $\mathrm{Re}\left[\mathcal{E}\left(\theta\right)\right]\odot M_\infty(\theta)$ is also positive semidefinite (Lütkepohl, 1996, p.152). Therefore, it follows from Oppenheim's inequality

If $A,B$ are $m\times m$ and positive semidefinite, then $\det(A\odot B)\geq\det A\prod_{i=1}^{m}b_{ii}$.

(Lütkepohl, 1996, p.56) that

$$\Xi(d) \geq \prod_{a=1}^{q} \left(\mathrm{Re}\left[\mathcal{E}\left(\theta\right)\right] \odot M_{\infty}(\theta)\right)_{aa} \det(G^0) = \prod_{a=1}^{q} \left[M_{\infty}(\theta)\right]_{aa} (\det G^0) = B(d),$$

giving (ii) of (9). (iii) follows because $\Xi(d^0) = \det(M_{\infty}(0) \odot G^0) = B(d^0)$, since all elements of $\mathcal{E}\left(0\right)$ are one.

We move to bound the second probability in (7). Observe that

$$\begin{aligned} S(d) &= \log\det \widehat{G}(d) - \log\det \widehat{G}(d^0) - 2\sum_{a=1}^{q} \theta_a \frac{1}{m}\sum_{j=1}^{m} \log \lambda_j \\ &= \log\det \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[\Lambda_j(\theta)^{-1}\Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1}\Lambda_j^*(\theta)^{-1}\right] \\ &\quad -2\sum_{a=1}^{q} \theta_a \frac{1}{m}\sum_{j=1}^{m} \log\lambda_j - \log\det \widehat{G}(d^0) \\ &= \log\det \widehat{D}(d) - \log\det \widehat{D}(d^0), \end{aligned} \tag{14}$$

where

$$\begin{aligned} \widehat{D}(d) &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[P_j(\theta)\Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1} P_j^*(\theta)\right], \\ P_j(\theta) &= \mathrm{diag}(e^{i(\lambda_j-\pi)\theta_a/2}(j/p)^{\theta_a}), \quad p = \exp(m^{-1}\textstyle\sum_{j=1}^{m} \log j) \sim m/e. \end{aligned}$$

Since $\log x$ is a monotone increasing function of $x$, $\Pr(\inf_{\Theta_2} S(d) \leq 0) \to 0$ follows if

$$\Pr(\inf\nolimits_{\Theta_2} \det \widehat{D}(d) - \det \widehat{D}(d^0) \leq 0) \to 0 \quad \text{as } n \to \infty. \tag{15}$$

For a $q$-vector $W_j$, we can write down each summand of $\widehat{D}(d)$ as

$$\begin{aligned} & \mathrm{Re}[P_j(\theta)\Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1} P_j^*(\theta)] \\ = \; & \mathrm{Re}[W_j W_j^*] = \mathrm{Re}[W_j](\mathrm{Re}[W_j])' + \mathrm{Im}[W_j](\mathrm{Im}[W_j])', \end{aligned}$$

which is positive semidefinite. Thus $\widehat{D}(d)$ is a sum of $m$ positive semidefinite matrices. For a fixed $\kappa \in (0,1)$, define

$$\begin{aligned} \widehat{D}_{\kappa}(d) &= \frac{1}{m}\sum_{j=[\kappa m]}^{m} \mathrm{Re}\left[P_j(\theta)\Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1} P_j^*(\theta)\right], \\ K_{\kappa}(d) &= \frac{1}{m}\sum_{j=[\kappa m]}^{m} \mathrm{Re}\left[P_j(\theta) G^0 \; P_j^*(\theta)\right]. \end{aligned}$$

Then, it follows from Lütkepohl (1996, p. 55) that

$$\det \widehat{D}(d) \geq \det \widehat{D}_{\kappa}(d). \tag{16}$$

$\widehat{D}_\kappa(d)$ is uniformly approximated by $K_\kappa(d)$. The $(a,b)$th element of $\widehat{D}_\kappa(d) - K_\kappa(d)$ is

$$\begin{aligned} & \frac{1}{m}\sum_{j=[\kappa m]}^{m} \mathrm{Re}\left[ e^{i(\lambda_j-\pi)(\theta_a-\theta_b)/2}\left(\frac{j}{p}\right)^{\theta_a+\theta_b}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)} - G_{ab}^0\right)\right] \\ = & \left(\frac{m}{p}\right)^{\theta_a+\theta_b} \mathrm{Re}\left[\frac{1}{m}\sum_{j=[\kappa m]}^{m} e^{i(\lambda_j-\pi)(\theta_a-\theta_b)/2}\left(\frac{j}{m}\right)^{\theta_a+\theta_b}\left(\frac{w_{aj}w_{bj}^*}{\Lambda_{ja}(d^0)\Lambda_{jb}^*(d^0)} - G_{ab}^0\right)\right] \\ = & o_p(1) \quad \text{uniformly in } \theta \in \Theta_2, \end{aligned}$$

where the third line is derived similarly to (11) from summation by parts, Lemma 1 (a), and Lemma 5.4 of Shimotsu and Phillips (2005). It follows that, for any $\kappa \in (0,1)$,

$$\sup_{\Theta_2}\left|\det \widehat{D}_\kappa(d) - \det K_\kappa(d)\right| = o_p(1), \quad \text{as } n \to \infty.$$

The proof is completed by deriving the lower bound of $K_\kappa(d)$ for $d \in \Theta_2$. Rewrite $K_\kappa(d)$ as

$$K_\kappa(d) = M_m^\kappa(\theta) \odot G^0,$$

where a positive semidefinite matrix $M_m^\kappa(\theta)$ is defined as

$$M_m^\kappa(\theta) = \frac{1}{m}\sum_{j=[\kappa m]}^{m} \mathrm{Re}\left[Z_j Z_j^*\right], \quad Z_j = \left(e^{i(\lambda_j-\pi)\theta_1/2}(j/p)^{\theta_1}, \cdots, e^{i(\lambda_j-\pi)\theta_q/2}(j/p)^{\theta_q}\right)'.$$

In view of Oppenheim's inequality, Lemma 5.5 of Shimotsu and Phillips (2005), and Lemma 2, there exist $\varepsilon \in (0, 0.1)$ and $\bar{\kappa} \in (0, 1/4)$ such that, for sufficiently large $m$ and all $\kappa \in (0, \bar{\kappa})$,

$$\begin{aligned} \inf_{\Theta_2} \det K_\kappa(d) & \geq \det G^0 \inf_{\Theta_2} \prod_{a=1}^{q} \frac{1}{m}\sum_{j=[\kappa m]}^{m}\left(\frac{j}{p}\right)^{2\theta_a} \\ & \geq \det G^0 (1+2\varepsilon)(1-\kappa^{2\Delta})^{q-1} + o(1). \end{aligned}$$

Choose $\kappa$ sufficiently small so that $(1+2\varepsilon)(1-\kappa^{2\Delta})^{q-1} \geq 1+\varepsilon$. Then, it follows that

$$\inf_{\Theta_2} \det \widehat{D}_\kappa(d) = \inf_{\Theta_2} \det K_\kappa(d) + o_p(1) \geq \det G^0(1+\varepsilon) + o_p(1).$$

From the results for $d \in \Theta_1$, we have $\det \widehat{D}(d^0) = \det \widehat{G}(d^0) \to_p \det G^0$ as $n \to \infty$. Therefore,

$$\Pr(\inf_{\Theta_2} \det \widehat{D}_\kappa(d) - \det \widehat{D}(d^0) \leq 0) \to 0 \quad \text{as } n \to \infty,$$

and (15) follows in view of (16), completing the proof. □

## 8.2 Proof of Theorem 2

We follow the approach developed by Lobato (1999). Theorem 1 holds under the current conditions and implies that with probability approaching one, as $n \to \infty$, $\widehat{d}$ satisfies

$$0 = \left.\frac{\mathrm{d}R(d)}{\mathrm{d}d}\right|_{\widehat{d}} = \left.\frac{\mathrm{d}R(d)}{\mathrm{d}d}\right|_{d^0} + \left(\left.\frac{\mathrm{d}^2R(d)}{\mathrm{d}d\mathrm{d}d'}\right|_{\bar{d}}\right)(\widehat{d} - d^0).$$

where $||\overline{d}-d^0|| \leq ||\widehat{d}-d^0||$. $\widehat{d}$ has the stated limiting distribution if, for any $q \times 1$ vector $\eta$, as $n \to \infty$,

$$\eta' \sqrt{m} \left.\frac{\mathrm{d}R(d)}{\mathrm{d}d}\right|_{d^0} = \sum_{a=1}^{q} \eta_a \sqrt{m} \left.\frac{\partial R(d)}{\partial d_a}\right|_{d^0} \to_d N(0, \eta'\Omega\eta), \tag{17}$$

$$\left.\frac{\mathrm{d}^2 R(d)}{\mathrm{d}d\mathrm{d}d'}\right|_{\overline{d}} \to_p \Omega, \quad \Omega = 2\left[G^0 \odot (G^0)^{-1} + I_q + \frac{\pi^2}{4}\left(G^0 \odot (G^0)^{-1} - I_q\right)\right] \tag{18}$$

#### 8.2.1 Score vector approximation

First we show (17). The proof is similar to that of Lobato (1999). Observe that

$$\sqrt{m}\frac{\partial R(d)}{\partial d_a} = -\frac{2}{\sqrt{m}}\sum_{j=1}^{m} \log \lambda_j + \mathrm{tr}\left[\widehat{G}(d)^{-1}\sqrt{m}\frac{\partial \widehat{G}(d)}{\partial d_a}\right]. \tag{19}$$

Let $i_a$ be a $q \times q$ matrix whose $a$th diagonal element is one and all other elements are zero, and let $\Lambda_j^0$ denote $\Lambda_j(d^0)$ in the following. From $\Lambda_j(d)^{-1} = \mathrm{diag}(\lambda_j^{d_a} e^{i(\lambda_j - \pi)d_a/2})$ and $\mathrm{Re}[(a+bi)(c+di)] = ac - bd$, we obtain

$$\begin{aligned}
\sqrt{m} \left.\frac{\partial \widehat{G}(d)}{\partial d_a}\right|_{d^0} &= \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \mathrm{Re}\left[\left(\log \lambda_j + \frac{\lambda_j - \pi}{2} i\right)(\Lambda_j^0)^{-1} i_a I_j (\Lambda_j^{0*})^{-1}\right] \\
&\quad + \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \mathrm{Re}\left[\left(\log \lambda_j - \frac{\lambda_j - \pi}{2} i\right)(\Lambda_j^0)^{-1} I_j i_a (\Lambda_j^{0*})^{-1}\right] \\
&= \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \log \lambda_j \mathrm{Re}\left[(\Lambda_j^0)^{-1}\left(i_a I_j + I_j i_a\right)(\Lambda_j^{0*})^{-1}\right] \\
&\quad + \frac{1}{\sqrt{m}}\sum_{j=1}^{m} \frac{\lambda_j - \pi}{2} \mathrm{Im}\left[(\Lambda_j^0)^{-1}\left(-i_a I_j + I_j i_a\right)(\Lambda_j^{0*})^{-1}\right], \\
&= H_{1a} + H_{2a}.
\end{aligned}$$

Therefore, $\sum_{a=1}^{q} \eta_a \sqrt{m}(\partial R(d))/(\partial d_a)|_{d^0}$ is equal to

$$\sum_{a=1}^{q} \eta_a \left\{-\frac{2}{\sqrt{m}}\sum_{j=1}^{m} \log \lambda_j + \mathrm{tr}\left[\widehat{G}\left(d^0\right)^{-1} H_{1a}\right]\right\} + \sum_{a=1}^{q} \eta_a \mathrm{tr}\left[\widehat{G}\left(d^0\right)^{-1} H_{2a}\right] = R_1 + R_2.$$

We proceed to find an approximation of $R_1$ and $R_2$. First, we obtain, with $\nu_j =$

$\log\lambda_j - m^{-1}\sum_1^m \log\lambda_j = \log j - m^{-1}\sum_1^m \log j = O(\log m)$,

$$
\begin{aligned}
& -\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j + \mathrm{tr}\left[\widehat{G}\left(d^0\right)^{-1}H_{1a}\right] \\
= \;& \mathrm{tr}\left[\widehat{G}(d^0)^{-1}\left(H_{1a} - \frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j\widehat{G}(d^0)i_a\right)\right] \\
= \;& \mathrm{tr}\left[\widehat{G}(d^0)^{-1}\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\nu_j\mathrm{Re}\left[(\Lambda_j^0)^{-1}I_j(\Lambda_j^{0*})^{-1}\right]i_a\right] \\
= \;& \left(g^a + o_p(1)\right)\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\nu_j\left\{\mathrm{Re}\left[(\Lambda_j^0)^{-1}I_j(\Lambda_j^{0*})^{-1}\right]\right\}_a, \qquad (20)
\end{aligned}
$$

where $g^a$ is the $a$th row of $(G^0)^{-1}$ and $\{A\}_a$ denotes the $a$th column of matrix $A$. For the moment, we proceed ignoring the $o_p(1)$ term in (20), but later it becomes clear that doing so does not affect the result. If follows from summation by parts, Lemma 1 (b1), and $\sum_1^m \nu_j = 0$ that

$$
\begin{aligned}
& \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\nu_j(\Lambda_j^0)^{-1}I_j(\Lambda_j^{0*})^{-1} \\
= \;& \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\nu_j\left[(\Lambda_j^0)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)(\Lambda_j^{0*})^{-1} - G^0\right] + o_p(1). \qquad (21)
\end{aligned}
$$

Thus

$$
R_1 = \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left(g^a\left\{\mathrm{Re}\left[(\Lambda_j^0)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)(\Lambda_j^{0*})^{-1}\right]\right\}_a - 1\right) + o_p(1). \qquad (22)
$$

The first term on the right is equal to

$$
\begin{aligned}
& \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left(g^a\left\{\mathrm{Re}\left[(\Lambda_j^0)^{-1}A(\lambda_j)\left(\frac{1}{2\pi n}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t'\right)A^*(\lambda_j)(\Lambda_j^{0*})^{-1}\right]\right\}_a - 1\right) \\
& + \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left(g^a\left\{\mathrm{Re}\left[(\Lambda_j^0)^{-1}A(\lambda_j)\left(\frac{1}{2\pi n}\sum\sum_{t\neq s}\varepsilon_t\varepsilon_s'e^{i(t-s)\lambda_j}\right)A^*(\lambda_j)(\Lambda_j^{0*})^{-1}\right]\right\}_a\right).
\end{aligned}
$$

The first part is $o_p(1)$ from applying the proof of Lemma 1 in Appendix D of Lobato (1999). It follows that

$$
R_1 = \sum_{t=2}^{n}\varepsilon_t'\sum_{s=1}^{t-1}\Phi_{t-s}\varepsilon_s + o_p(1), \quad \Phi_s = \frac{1}{\pi\sqrt{m}n}\sum_{j=1}^{m}\nu_j\mathrm{Re}\left[\Omega_j e^{-is\lambda_j} + \Omega_j' e^{is\lambda_j}\right], \qquad (23)
$$

where $\Omega_j$ is defined as

$$\Omega_j = \sum_{a=1}^{q} \eta_a \left\{ A^*(\lambda_j)(\Lambda_j^{0*})^{-1} \right\}_a g^a (\Lambda_j^0)^{-1} A(\lambda_j). \tag{24}$$

We can simplify $R_1$ further as

$$R_1 = \sum_{t=2}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \Theta_{t-s} \varepsilon_s + o_p(1), \qquad \Theta_s = \frac{1}{\pi\sqrt{mn}} \sum_{j=1}^{m} \nu_j \mathrm{Re}\left[\Omega_j + \Omega_j'\right] \cos(s\lambda_j). \tag{25}$$

This is because we can rewrite $\sum_{t=1}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \Phi_{t-s}\varepsilon_s - \sum_{t=1}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \Theta_{t-s}\varepsilon_s$ as

$$\sum_{t=2}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \left\{ \frac{1}{\pi\sqrt{mn}} \sum_{j=1}^{m} \nu_j \mathrm{Im}\left[\Omega_j - \Omega_j'\right] \sin(t-s)\lambda_j \right\} \varepsilon_s. \tag{26}$$

This is $o_p(1)$ because its second moment is equal to

$$\begin{aligned} &\frac{1}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j=1}^{m} \nu_j^2 \mathrm{tr}\left\{ \left(\mathrm{Im}\left[\Omega_j - \Omega_j'\right]\right)' \mathrm{Im}\left[\Omega_j - \Omega_j'\right] \right\} \sin^2(s\lambda_j) \\ &+ \frac{1}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum\sum_{j \neq k}^{m} \nu_j \nu_k \mathrm{tr}\left\{ \left(\mathrm{Im}\left[\Omega_j - \Omega_j'\right]\right)' \mathrm{Im}\left[\Omega_k - \Omega_k'\right] \right\} \sin(s\lambda_j)\sin(s\lambda_k), \end{aligned}$$

which is $o(1)$ from $\mathrm{Im}[\Omega_j] \to 0$ by Assumption $5'$, $\sum_{j=1}^{m} \nu_j^2 = O(m)$, and Lemma 3 (c) and (d).

We move to $R_2$. Proceeding similarly to (20) - (21), we obtain

$$\mathrm{tr}\left[\widehat{G}\left(d^0\right)^{-1} H_{2a}\right] = (g^a + o_p(1)) \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\lambda_j - \pi) \left\{ \mathrm{Im}\left[(\Lambda_j^0)^{-1} I_j (\Lambda_j^{0*})^{-1}\right] \right\}_a. \tag{27}$$

It follows from Lemma 1 (b2), Assumption $5'$, and the uncorrelatedness of $I_{\varepsilon j}$ and $I_{\varepsilon k}$ for $j \neq k$ that

$$\begin{aligned} &\frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\lambda_j - \pi) \mathrm{Im}\left[(\Lambda_j^0)^{-1} I_j (\Lambda_j^{0*})^{-1}\right] \\ =\ & -\frac{\pi}{\sqrt{m}} \sum_{j=1}^{m} \mathrm{Im}\left[(\Lambda_j^0)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) (\Lambda_j^{0*})^{-1}\right] + o_p(1). \end{aligned}$$

Therefore, ignoring the $o_p(1)$ term in (27) and proceeding as between (21) and (23), we find that

$$R_2 = -\frac{\pi}{2} \sum_{t=2}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \widetilde{\Phi}_{t-s} \varepsilon_s + o_p(1), \quad \widetilde{\Phi}_s = \frac{1}{\pi\sqrt{mn}} \sum_{j=1}^{m} \mathrm{Im}\left[\Omega_j e^{-is\lambda_j} + \Omega_j' e^{is\lambda_j}\right],$$

where $\Omega_j$ is defined in (24). Using a decomposition similar to (26) in conjunction with $\mathrm{Im}[\Omega_j] \to 0$ and Lemma 3 (a) and (b), we can simplify $R_2$ as

$$R_2 = \sum_{t=1}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \widetilde{\Theta}_{t-s} \varepsilon_s + o_p(1), \qquad \widetilde{\Theta}_s = \frac{\pi}{2} \frac{1}{\pi\sqrt{m}n} \sum_{j=1}^{m} \mathrm{Re}\left[\Omega_j - \Omega_j'\right] \sin(s\lambda_j). \quad (28)$$

It follows from (25) and (28) that, with $z_1 = 0$,

$$\sum_{a=1}^{q} \eta_a \sqrt{m} \frac{\partial R(d)}{\partial d_a}\bigg|_{d^0} = \sum_{t=1}^{n} z_t + o_p(1), \quad z_t = \varepsilon_t' \sum_{s=1}^{t-1} \left[\Theta_{t-s} + \widetilde{\Theta}_{t-s}\right] \varepsilon_s.$$

By a standard martingale CLT, (17) follows if

$$\sum_{t=1}^{n} E(z_t^2 | F_{t-1}) - \sum_{a=1}^{q} \sum_{b=1}^{q} \eta_a \eta_b \Omega_{ab} \quad \to \quad {}_p 0, \quad (29)$$

$$\sum_{t=1}^{n} E(z_t^2 I(|z_t| \quad > \quad \delta)) \to 0 \quad \text{for all } \delta > 0. \quad (30)$$

Applying the argument in Lobato (1999, pp. 149-51) to our $\Theta_s$ and $\widetilde{\Theta}_s$ , we obtain $||\Theta_s||, ||\widetilde{\Theta}_s|| = O(n^{-1} m^{1/2} \log m)$ for $1 \le s \le n/2$ and $||\Theta_s||, ||\widetilde{\Theta}_s|| = O(m^{-1/2} s^{-1} \log m)$, and Assumption 1 implies that $\Omega_j = O(1)$. Therefore, Lemmas 2 and 3 in Lobato (1999) hold for our $\Theta_s$ and $\widetilde{\Theta}_s$. Hence, we can apply the arguments in Lobato (1999, Proof of (C2), pp. 142-43) to show that (30) holds. For (29), from the results in Lobato (1999, page 142 line 1 and Lemmas 2 and 3), we have

$$\sum_{t=1}^{n} E(z_t^2 | F_{t-1}) = \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathrm{tr}\left[\left(\Theta_{t-s} + \widetilde{\Theta}_{t-s}\right)' \left(\Theta_{t-s} + \widetilde{\Theta}_{t-s}\right)\right] + o_p(1). \quad (31)$$

Now

$$\begin{aligned}
& \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathrm{tr}\left[\Theta_{t-s}' \Theta_{t-s} + \widetilde{\Theta}_{t-s}' \widetilde{\Theta}_{t-s}\right] \\
= \ & \frac{1}{\pi^2 m n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j=1}^{m} \nu_j^2 \mathrm{tr}\left\{\mathrm{Re}\left[\Omega_j' + \Omega_j\right] \mathrm{Re}\left[\Omega_j + \Omega_j'\right]\right\} \cos^2(s\lambda_j) \\
& + \frac{1}{\pi^2 m n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j \neq k}^{m} \sum \nu_j \nu_k \mathrm{tr}\left\{\mathrm{Re}\left[\Omega_j' + \Omega_j\right] \mathrm{Re}\left[\Omega_k + \Omega_k'\right]\right\} \cos(s\lambda_j) \cos(s\lambda_k) \\
& + \frac{\pi^2}{4} \frac{1}{\pi^2 m n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j=1}^{m} \mathrm{tr}\left\{\mathrm{Re}\left[\Omega_j' - \Omega_j\right] \mathrm{Re}\left[\Omega_j - \Omega_j'\right]\right\} \sin^2(s\lambda_j) \\
& + \frac{\pi^2}{4} \frac{1}{\pi^2 m n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j \neq k}^{m} \sum \mathrm{tr}\left\{\mathrm{Re}\left[\Omega_j' - \Omega_j\right] \mathrm{Re}\left[\Omega_k - \Omega_k'\right]\right\} \sin(s\lambda_j) \sin(s\lambda_k).
\end{aligned}$$

The second and fourth terms are $o(1)$ from $\Omega_j = O(1)$ and Lemma 3 (b) and (d). For the first and third terms, observe that

$$\begin{cases} \operatorname{tr}\left\{(4\pi^2)^{-1}\mathrm{Re}\left[\Omega_j'\right]\mathrm{Re}\left[\Omega_j\right]\right\} \to \sum_{a=1}^q \sum_{b=1}^q \eta_a\eta_b G_{ab}^0 (G^0)_{ab}^{-1}, \\ \operatorname{tr}\left\{(4\pi^2)^{-1}\mathrm{Re}\left[\Omega_j\right]\mathrm{Re}\left[\Omega_j\right]\right\} \to \sum_{a=1}^q \eta_a^2, \end{cases} \tag{32}$$

as $\lambda_j \to 0$. Therefore, in view of Lemma 3 (a) and (c), we have $\sum_{t=2}^n \sum_{s=1}^{t-1} \operatorname{tr}[\Theta_{t-s}'\Theta_{t-s} + \widetilde{\Theta}_{t-s}'\widetilde{\Theta}_{t-s}] \to \sum_{a=1}^q \sum_{b=1}^q \eta_a\eta_b\Omega_{ab}$. Finally,

$$\begin{aligned} & \sum_{t=2}^{n}\sum_{s=1}^{t-1} \Theta_{t-s}'\widetilde{\Theta}_{t-s} \\ = \ & \frac{1}{2\pi mn^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\sum_{j=1}^{m}\sum_{k=1}^{m} \nu_j\nu_k \operatorname{tr}\left\{\mathrm{Re}\left[\Omega_j' + \Omega_j\right]\mathrm{Re}\left[\Omega_k - \Omega_k'\right]\right\}\cos(s\lambda_j)\sin(s\lambda_k) = 0, \end{aligned}$$

because $\operatorname{tr}((A'+A)(B-B')) = 0$ for any real matrices $A, B$. Therefore, $\sum_{t=2}^n \sum_{s=1}^{t-1} \operatorname{tr}[(\Theta_{t-s} + \widetilde{\Theta}_{t-s})'(\Theta_{t-s} + \widetilde{\Theta}_{t-s})] \to \sum_{a=1}^q \sum_{b=1}^q \eta_a\eta_b\Omega_{ab}$, establishing (29). Since $m^{-1/2}\sum_{j=1}^m \nu_j \mathrm{Re}[(\Lambda_j^0)^{-1} I_j (\Lambda_j^{0*})^{-1}]$ and $m^{-1/2}\sum_{j=1}^m \mathrm{Im}[(\Lambda_j^0)^{-1} I_j (\Lambda_j^{0*})^{-1}]$ are $O_p(1)$ from the above argument, the $o_p(1)$ terms in (20) and (27) do not affect the result.

### 8.2.2 Hessian approximation

The proof is similar to that of Lobato (1999). Fix $\varepsilon > 0$ and let $M = \{d : (\log n)^4 ||d - d^0|| < \varepsilon\} = \{\theta : (\log n)^4 ||\theta|| < \varepsilon\}$. First, we show $\Pr(\overline{d} \notin M) \to 0$ as $n \to \infty$. Using the notations in the proof of Theorem 1, $\inf_{\Theta_1 \setminus M} S_2(d)$ is bounded as

$$\inf_{\Theta_1 \setminus M} S_2(d) \geq \varepsilon^2 (\log n)^8 / 8.$$

By applying Lemma 1 (b2) to (11), we strengthen (i) of (9) to

$$\sup_{\Theta_1} |A(d) - \Xi(d)| = O_p(m^\beta n^{-\beta} + m^{-2\Delta}\log m + mn^{-1}).$$

It follows that, uniformly in $\Theta_1$,

$$\begin{aligned} \log A(d) - \log B(d) &\geq \log\left([\Xi(d) + o_p((\log n)^{-8})]/\Xi(d)\right) = o_p((\log n)^{-8}), \\ \log A(d^0) - \log B(d^0) &= \log\left([\Xi(d^0) + o_p((\log n)^{-8})]/\Xi(d^0)\right) = o_p((\log n)^{-8}). \end{aligned}$$

Therefore, as $n \to \infty$, $\Pr(\inf_{\Theta_1 \setminus M} S(d) \leq 0) \to 0$ and $\Pr(\overline{d} \notin M) \to 0$ follow.

Observe that

$$\frac{\partial^2 R(d)}{\partial d_a \partial d_b} = \operatorname{tr}\left[-\widehat{G}^{-1}(d)\frac{\partial \widehat{G}(d)}{\partial d_a}\widehat{G}^{-1}(d)\frac{\partial \widehat{G}(d)}{\partial d_b} + \widehat{G}^{-1}(d)\frac{\partial^2 \widehat{G}(d)}{\partial d_a \partial d_b}\right]. \tag{33}$$

The derivatives of $\widehat{G}(d)$ are given by

$$\begin{aligned} \frac{\partial \widehat{G}(d)}{\partial d_a} &= \frac{1}{m}\sum_{j=1}^m \mathrm{Re}\left[\left(\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right) i_a \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1}\right] \\ &\quad + \frac{1}{m}\sum_{j=1}^m \mathrm{Re}\left[\left(\log\lambda_j - \frac{\lambda_j - \pi}{2}i\right) \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1} i_a\right], \end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 \widehat{G}(d)}{\partial d_a \partial d_b} &= \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[\left(\log\lambda_j + \frac{\lambda_j-\pi}{2}i\right)^2 i_a i_b \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1}\right] \\ &+\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[\left|\log\lambda_j - \frac{\lambda_j-\pi}{2}i\right|^2 i_a \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1} i_b\right] \\ &+\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[\left|\log\lambda_j - \frac{\lambda_j-\pi}{2}i\right|^2 i_b \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1} i_a\right] \\ &+\frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left[\left(\log\lambda_j - \frac{\lambda_j-\pi}{2}i\right)^2 \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1} i_a i_b\right].\end{aligned}$$

Define, for $k=0,1,2$,

$$\begin{aligned}\widehat{G}_k(d) &= m^{-1}\textstyle\sum_{j=1}^{m}(\log\lambda_j)^k \mathrm{Re}\left[\Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1}\right], \\ \overline{G}_k(d) &= m^{-1}\textstyle\sum_{j=1}^{m}(\log\lambda_j)^k \mathrm{Im}\left[\Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1}\right],\end{aligned}$$

Then it follows that

$$\begin{aligned}\frac{\partial \widehat{G}(d)}{\partial d_a} &= i_a\widehat{G}_1(d) + \widehat{G}_1(d) i_a + (\pi/2) i_a \overline{G}_0(d) - (\pi/2)\overline{G}_0(d) i_a + o_p((\log n)^{-1}), \\ \frac{\partial^2 \widehat{G}(d)}{\partial d_a \partial d_b} &= i_a i_b \widehat{G}_2(d) + i_a \widehat{G}_2(d) i_b + i_b \widehat{G}_2(d) i_a + \widehat{G}_2(d) i_a i_b \\ &+(\pi^2/4)\left[-i_a i_b \widehat{G}_0(d) + i_a\widehat{G}_0(d) i_b + i_b \widehat{G}_0(d) i_a - \widehat{G}_0(d) i_a i_b\right] \\ &+\pi i_a i_b \overline{G}_1(d) - \pi \overline{G}_1(d) i_a i_b + o_p(1),\end{aligned}$$

where the order of the reminder terms follows from summation by parts, $\sum_{j=1}^{r}\Lambda_j(d)^{-1}I_j\Lambda_j^*(d)^{-1} = O_p(r)$, and Assumption 4′. We state the following properties, to be established later. Uniformly in $d\in M$,

$$\widehat{G}_k(d) = G^0 m^{-1}\textstyle\sum_{j=1}^{m}(\log\lambda_j)^k + o_p((\log n)^{k-2}), \quad \overline{G}_k(d) = o_p((\log n)^{k-2}). \tag{34}$$

The assumption $m^{-\gamma}\log n \to 0$ is necessary in showing (34), because the terms with $\overline{G}_1(d)$ do not cancel out even if we take the trace of $\widehat{G}^{-1}(d)(\partial^2\widehat{G}(d))/(\partial d_a \partial d_b)$. Define $G_{1a}^0 = i_a G^0 + G^0 i_a$, $G_{2ab}^0 = i_a i_b G^0 + i_a G^0 i_b + i_b G^0 i_a + G^0 i_a i_b$, and $G_{3ab}^0 = -i_a i_b G^0 + i_a G^0 i_b + i_b G^0 i_a - G^0 i_a i_b$. It follows from (34) that

$$\begin{aligned}\widehat{G}^{-1}(\overline{d}) &= (G^0)^{-1} + o_p((\log n)^{-2}), \\ \partial\widehat{G}(\overline{d})/\partial d_a &= m^{-1}\textstyle\sum_{j=1}^{m}(\log\lambda_j) G_{1a}^0 + o_p((\log n)^{-1}), \\ \partial^2\widehat{G}(\overline{d})/\partial d_a\partial d_b &= m^{-1}\textstyle\sum_{j=1}^{m}(\log\lambda_j)^2 G_{2ab}^0 + (\pi^2/4) G_{3ab}^0 + o_p(1).\end{aligned}$$

Since $\mathrm{tr}[(G^0)^{-1}G_{1a}^0(G^0)^{-1}G_{1b}^0] = \mathrm{tr}[(G^0)^{-1}G_{2ab}^0]$ and $m^{-1}\sum_{j=1}^{m}(\log\lambda_j)^2 - [m^{-1}\sum_{j=1}^{m}(\log\lambda_j)]^2 \to 1$, we obtain

$$\frac{\partial^2 R(\widetilde{d})}{\partial d_a \partial d_b} = \mathrm{tr}\left[(G^0)^{-1}G_{2ab}^0 + (\pi^2/4)(G^0)^{-1}G_{3ab}^0\right] + o_p(1),$$

and (18) follows. $\widehat{G}(\bar{d}) \to_p G^0$ follows from (34).

It remains to show (34). Define

$$F_k(d) = m^{-1} \sum_{j=1}^m (\log \lambda_j)^k \Lambda_j(\theta)^{-1} G^0 \Lambda_j^*(\theta)^{-1},$$

then (34) follows if

$$\sup_{d\in M} \left\| m^{-1} \sum_{j=1}^m (\log \lambda_j)^k \Lambda_j(d)^{-1} I_j \Lambda_j^*(d)^{-1} - F_k(d) \right\| = o_p((\log n)^{k-2}), \quad (35)$$

$$\sup_{d\in M} \left\| F_k(d) - G^0 m^{-1} \sum_{j=1}^m (\log \lambda_j)^k \right\| = o((\log n)^{k-2}). \quad (36)$$

For (35), rewrite it as

$$\sup_{d\in M} \left\| m^{-1} \sum_{j=1}^m (\log \lambda_j)^k \Lambda_j(\theta)^{-1} \left[ \Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1} - G^0 \right] \Lambda_j^*(\theta)^{-1} \right\|. \quad (37)$$

Define $b_{nj}(\theta) = (\log \lambda_j)^k e^{i(\lambda_j - \pi)(\theta_a - \theta_b)/2} \lambda_j^{\theta_a + \theta_b}$, then the $(a, b)$th element of the inside of $\sup_{d\in M}$ in (37 ) is equal to

$$m^{-1} \sum_{j=1}^m b_{nj}(\theta) \left[ e^{i(\lambda_j - \pi)(d_a^0 - d_b^0)/2} \lambda_j^{d_a^0 + d_b^0} w_{aj} w_{bj}^* - G_{ab}^0 \right].$$

It is easily seen that $b_{nj}(\theta) - b_{n,j+1}(\theta) = O((\log n)^k j^{-1})$ and $b_{nm} = O((\log n)^k)$ uniformly in $\theta \in M$. Therefore, it follows from summation by parts and Lemma 1 (b2) that $(37) = O_p((\log n)^k m^{-1} \sum_{r=1}^m (r^\beta n^{-\beta} + r^{-1/2} \log r)) = o_p((\log n)^{k-2})$.

We move to the proof of (36). The $(a, b)$th element of the inside of $\sup_{d\in M}$ in (36) is equal to

$$m^{-1} \sum_{j=1}^m (\log \lambda_j)^k \left[ e^{i(\lambda_j - \pi)(\theta_a - \theta_b)/2} \lambda_j^{\theta_a + \theta_b} - 1 \right] G_{ab}^0. \quad (38)$$

Since, for $\theta \in M$ and $0 < \lambda_j \le 1$, $|\lambda_j^{\theta_a + \theta_b} - 1| / |\theta_a + \theta_b| \le |\log \lambda_j| n^{|\theta_a| + |\theta_b|} \le C \log n$ and $|e^{i(\lambda_j - \pi)(\theta_a - \theta_b)/2} - 1| \le C(|\theta_a| + |\theta_b|)$, we have

$$\sup_{d\in M} |e^{i(\lambda_j - \pi)(\theta_a - \theta_b)/2} \lambda_j^{\theta_a + \theta_b} - 1| \le C \sup_{d\in M} (|\theta_a| + |\theta_b|) \log n = O((\log n)^{-3}),$$

and hence (38) is $o((\log n)^{k-2})$. Therefore, we show (34) and complete the proof. □

## 8.3 Proof of Theorem 3 (a)

The proof follows the logic of the proof of Theorem 1, with corresponding modifications. Define

$$\begin{aligned} \widetilde{S}(d) &= \widetilde{R}(d) - \widetilde{R}(d^0) \\ &= \log \det \widetilde{G}(d) - \log \det \widetilde{G}(d^0) - 2 \sum_{a=1}^q \theta_a \frac{1}{m} \sum_{j=1}^m \log \lambda_j \\ &= \log \widetilde{A}(d) - \log \widetilde{B}(d) - \log \widetilde{A}(d^0) + \log \widetilde{B}(d^0) + S_2(d), \end{aligned}$$

where $S_2(d)$ is defined in (8) and

$$\widetilde{A}(d) = \left(\frac{2\pi m}{n}\right)^{-2(\theta_1+\cdots+\theta_q)} \det \widetilde{G}(d), \quad \widetilde{B}(d) = \prod_{a=1}^{q} (2\theta_a + 1)^{-1} \det \widetilde{G}^0.$$

From the proof of Theorem 1, $\Pr(\inf_{\overline{N}_\delta \cap \Theta_1} \widetilde{S}(d) \leq 0) \to 0$ follows if we find a non-random $\widetilde{\Xi}(d)$ such that, as $n \to \infty$,

$$\text{(i)} \quad \sup_{\Theta_1} \left|\widetilde{A}(d) - \widetilde{\Xi}(d)\right| = o_p(1), \quad \text{(ii)} \quad \widetilde{\Xi}(d) \geq \widetilde{B}(d), \quad \text{(iii)} \quad \widetilde{\Xi}(d^0) = \widetilde{B}(d^0). \tag{39}$$

Define $\widetilde{M_j}(\theta)$ =diag$(e^{-i(\lambda_j - \pi)d_a^0/2}(j/m)^{\theta_a})$. Then, the expression of $\widetilde{A}(d)$ corresponding to (10) is

$$\widetilde{A}(d) = \det\left\{\frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[\widetilde{M_j}(\theta)\Lambda_j(d^0)^{-1} I(\lambda_j)\Lambda_j^*(d^0)^{-1}\widetilde{M_j^*}(\theta)\right]\right\}. \tag{40}$$

Applying summation by parts, Lemma 1 (a), and the bound provided by (11) gives, uniformly in $\Theta_1$,

$$\widetilde{A}(d) = \det\left\{\frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[\widetilde{M_j}(\theta) G^0 \widetilde{M_j^*}(\theta)\right] + o_p(1)\right\}.$$

Proceeding as in the proof of Theorem 1 with $e^{i\lambda} = 1 + O(\lambda)$, we obtain the bound corresponding to (12):

$$\frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[\widetilde{M_j}(\theta) G^0 \widetilde{M_j^*}(\theta)\right] = M_\infty(\theta) \odot \widetilde{G}^0 + O(mn^{-1}) + O(m^{-2\Delta}). \tag{41}$$

Therefore, (i) and (iii) of (39) follow with $\widetilde{\Xi}(d) = \det(M_\infty(\theta) \odot \widetilde{G}^0)$. Define $\mathcal{E}^0$ to be a matrix whose $(a, b)$ elements are $\exp[i\pi(d_a^0 - d_b^0)/2]$. Since we can rewrite $\widetilde{G}^0 = \text{Re}[\mathcal{E}^0] \odot G^0$ and $\text{Re}[\mathcal{E}^0]$ is positive semidefinite (c.f. (13)), $\widetilde{G}^0$ is positive semidefinite. Therefore, we may apply Oppenheim's inequality to $\widetilde{\Xi}(d)$, and (ii) of (39) follows because

$$\widetilde{\Xi}(d) \geq \prod_{a=1}^{q} [M_\infty(\theta)]_{aa} \det \widetilde{G}^0 = \widetilde{B}(d).$$

The proof completes by showing $\Pr(\inf_{\Theta_2} \widetilde{S}(d) \leq 0) \to 0$. In place of (14), we obtain

$$\widetilde{S}(d) = \log\det \widetilde{D}(d) - \log\det \widetilde{D}(d^0),$$

where

$$\begin{aligned}
\widetilde{D}(d) &= \frac{1}{m}\sum_{j=1}^{m} \text{Re}\left[\widetilde{P}_j(\theta)\Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1}\widetilde{P}_j^*(\theta)\right], \\
\widetilde{P}_j(\theta) &= \text{diag}(e^{-i(\lambda_j - \pi)d_a^0/2}(j/p)^{\theta_a}).
\end{aligned}$$

$\Pr(\inf_{\Theta_2} \widetilde{S}(d) \leq 0) \to 0$ follows if

$$\Pr(\inf_{\Theta_2} \det \widetilde{D}(d) - \det \widetilde{D}(d^0) \leq 0) \to 0 \quad \text{as } n \to \infty. \tag{42}$$

For a fixed $\kappa \in (0, 1/4)$, define

$$\widetilde{D}_\kappa(d) = \frac{1}{m} \sum_{j=[\kappa m]}^{m} \mathrm{Re}\left[\widetilde{P}_j(\theta)\Lambda_j(d^0)^{-1} I_j \Lambda_j^*(d^0)^{-1} \widetilde{P}_j^*(\theta)\right],$$

and $\widetilde{K}_\kappa(d) = m^{-1} \sum_{j=[\kappa m]}^{m} \mathrm{Re}[\widetilde{P}_j(\theta) G^0 \ \widetilde{P}_j^*(\theta)]$. Then we obtain $\det \widetilde{D}(d) \geq \det \widetilde{D}_\kappa(d)$ and $\sup_{\Theta_2} |\det \widetilde{D}_\kappa(d) - \det \widetilde{K}_\kappa(d)| = o_p(1)$ by applying the same argument as the proof of Theorem 1. We move to derive the lower bound of $\widetilde{K}_\kappa(d)$. Rewrite $\widetilde{K}_\kappa(d)$ as

$$\widetilde{K}_\kappa(d) = \widetilde{M}_m^\kappa(\theta) \odot \mathcal{E}_m^\kappa \odot G^0,$$

where a positive semidefinite matrices $\widetilde{M}_m^\kappa(\theta)$ and $\mathcal{E}_m^\kappa$ are defined as

$$\begin{aligned}
\widetilde{M}_m^\kappa(\theta) &= \frac{1}{m} \sum_{j=[\kappa m]}^{m} \widetilde{Z}_j \widetilde{Z}_j', \quad \widetilde{Z}_j = \left((j/p)^{\theta_1}, \cdots, (j/p)^{\theta_q}\right)', \\
\mathcal{E}_m^\kappa &= \frac{1}{m} \sum_{j=[\kappa m]}^{m} \mathrm{Re}\left[\xi_j \xi_j^*\right], \quad \xi_j = \left(e^{-i(\lambda_j - \pi)d_1^0/2}, \cdots, e^{-i(\lambda_j - \pi)d_q^0/2}\right)'.
\end{aligned}$$

From Oppenheim's inequality, Lemma 5.5 of Shimotsu and Phillips (2005), and Lemma 2, it follows that there exist $\varepsilon \in (0, 0.1)$ and $\bar{\kappa} \in (0, 1/4)$ such that, for sufficiently large $m$ and all $\kappa \in (0, \bar{\kappa})$,

$$\begin{aligned}
\inf_{\Theta_2} \det \widetilde{K}_\kappa(d) &\geq \det\left\{\mathcal{E}_m^\kappa \odot G^0\right\} \inf_{\Theta_2} \prod_{a=1}^{q} \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left(\frac{j}{p}\right)^{2\theta_a} \\
&\geq \det\left\{(1-\kappa)\widetilde{G}^0 + o(1)\right\}(1+2\varepsilon)(1-\kappa^{2\Delta})^{q-1} + o(1).
\end{aligned}$$

Choosing $\kappa$ sufficiently small gives $\inf_{\Theta_2} \det \widetilde{D}_\kappa(d) = \inf_{\Theta_2} \det K_\kappa(d) + o_p(1) \geq (1 + \varepsilon) \det \widetilde{G}^0 + o_p(1)$. From the results for $d \in \Theta_1$, $\det \widetilde{D}(d^0) = \det \widetilde{G}(d^0) \to_p \det \widetilde{G}^0$, and (42) follows and we complete the proof. □

## 8.4 Proof of Theorem 3 (b)

The proof follows the logic of the proof of Theorem 2, with corresponding modifications. See the proof of Theorem 2 for the relevant definitions if not stated herein. The stated result follows if, with $\ddot{d} \in [d^0, \widetilde{d}]$,

$$\eta' \sqrt{m} \left.\frac{\mathrm{d}\widetilde{R}(d)}{\mathrm{d}d}\right|_{d^0} = \sum_{a=1}^{q} \eta_a \sqrt{m} \left.\frac{\partial \widetilde{R}(d)}{\partial d_a}\right|_{d^0} \to_d N(0, \eta' \Upsilon \eta), \tag{43}$$

$$\left.\frac{\mathrm{d}^2 \widetilde{R}(d)}{\mathrm{d}d\mathrm{d}d'}\right|_{\ddot{d}} \to_p \widetilde{\Omega}. \tag{44}$$

#### 8.4.1 Score vector approximation

First we show (43). Define $\mathcal{E}_j$ =diag(exp$[i(\pi-\lambda_j)d_a^0/2]$) for $j=0,\ldots,m$, then

$$\widetilde{G}\left(d^0\right)=\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left[\mathcal{E}_j(\Lambda_j^0)^{-1}I_j(\Lambda_j^{0*})^{-1}\mathcal{E}_j^*\right],$$

and

$$\sqrt{m}\left.\frac{\partial\widetilde{G}\left(d\right)}{\partial d_a}\right|_{d^0}=\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j\mathrm{Re}\left[\mathcal{E}_j(\Lambda_j^0)^{-1}\left(i_aI_j+I_ji_a\right)(\Lambda_j^{0*})^{-1}\mathcal{E}_j^*\right].$$

Using an argument similar to (19) and (20), we obtain

$$\begin{aligned}
&\sum_{a=1}^{q}\eta_a\sqrt{m}\left.\frac{\partial\widetilde{R}(d)}{\partial d_a}\right|_{d^0}\\
=\;&\sum_{a=1}^{q}\eta_a\left\{-\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j+\mathrm{tr}\left[\widetilde{G}\left(d^0\right)^{-1}\sqrt{m}\frac{\partial\widetilde{G}\left(d^0\right)}{\partial d_a}\right]\right\}\\
=\;&\sum_{a=1}^{q}\eta_a\left(\widetilde{g}^a+o_p(1)\right)\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\nu_j\left\{\mathrm{Re}\left[\mathcal{E}_j(\Lambda_j^0)^{-1}I_j(\Lambda_j^{0*})^{-1}\mathcal{E}_j^*\right]\right\}_a=\widetilde{R}_1,
\end{aligned}$$

where $\widetilde{g}^a$ is the $a$th row of $(\widetilde{G}^0)^{-1}$ and $\{A\}_a$ denotes the $a$th column of matrix $A$. Using summation by parts, Lemma 1 (b1), $\sum_1^m\nu_j=0$, and $m^3n^{-2}(\log m)^2\to 0$, in place of (21) we have

$$\begin{aligned}
&\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\nu_j\mathcal{E}_j(\Lambda_j^0)^{-1}I_j(\Lambda_j^{0*})^{-1}\mathcal{E}_j^*\\
=\;&\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\nu_j\left[\mathcal{E}_0(\Lambda_j^0)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)(\Lambda_j^{0*})^{-1}\mathcal{E}_0^*-\mathcal{E}_0G^0\mathcal{E}_0^*\right]+o_p(1).
\end{aligned}\tag{45}$$

Then, proceeding as in the proof of Theorem 2, in place of (22) and (23) we obtain

$$\begin{aligned}
\widetilde{R}_1&=\frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left(\widetilde{g}^a\left\{\mathrm{Re}\left[\mathcal{E}_0(\Lambda_j^0)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)(\Lambda_j^{0*})^{-1}\mathcal{E}_0^*\right]\right\}_a-1\right)+o_p(1)\\
&=\sum_{t=2}^{n}\varepsilon_t'\sum_{s=1}^{t-1}\Psi_{t-s}\varepsilon_s+o_p(1),
\end{aligned}$$

where

$$\Psi_s = \frac{1}{\pi\sqrt{mn}} \sum_{j=1}^{m} \nu_j \mathrm{Re}\left[\ddot{\Omega}_j e^{-s\lambda_j} + \ddot{\Omega}_j' e^{s\lambda_j}\right]$$
$$= \frac{1}{\pi\sqrt{mn}} \sum_{j=1}^{m} \nu_j \mathrm{Re}\left[\ddot{\Omega}_j + \ddot{\Omega}_j'\right] \cos(s\lambda_j) + \frac{1}{\pi\sqrt{mn}} \sum_{j=1}^{m} \nu_j \mathrm{Im}\left[\ddot{\Omega}_j - \ddot{\Omega}_j'\right] \sin(s\lambda_j),$$
$$\ddot{\Omega}_j = \sum_{a=1}^{q} \eta_a \left\{A^*(\lambda_j)(\Lambda_j^{0*})^{-1}\mathcal{E}_0^*\right\}_a \widetilde{g}^a \mathcal{E}_0 (\Lambda_j^0)^{-1} A(\lambda_j).$$

Since $||\Psi_s||$ is bounded by the bound of $||\Theta_s||$, (43) follows if (c.f. (29)-(31))

$$\sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathrm{tr}\left[\Psi_{t-s}' \Psi_{t-s}\right] - \sum_{a=1}^{q} \sum_{b=1}^{q} \eta_a \eta_b \Upsilon_{ab} \to_p 0. \tag{46}$$

Now

$$\sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathrm{tr}\left[\Psi_{t-s}' \Psi_{t-s}\right]$$
$$= \frac{1}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j=1}^{m} \nu_j^2 \mathrm{tr}\left\{\mathrm{Re}\left[\ddot{\Omega}_j' + \ddot{\Omega}_j\right] \mathrm{Re}\left[\ddot{\Omega}_j + \ddot{\Omega}_j'\right]\right\} \cos^2(s\lambda_j)$$
$$+ \frac{1}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum\sum_{j\neq k}^{m} \nu_j \nu_k \mathrm{tr}\left\{\mathrm{Re}\left[\ddot{\Omega}_j' + \ddot{\Omega}_j\right] \mathrm{Re}\left[\ddot{\Omega}_k + \ddot{\Omega}_k'\right]\right\} \cos(s\lambda_j)\cos(s\lambda_k)$$
$$+ \frac{1}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j=1}^{m} \nu_j^2 \mathrm{tr}\left\{\mathrm{Im}\left[\ddot{\Omega}_j' - \ddot{\Omega}_j\right] \mathrm{Im}\left[\ddot{\Omega}_j - \ddot{\Omega}_j'\right]\right\} \sin^2(s\lambda_j)$$
$$+ \frac{1}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum\sum_{j\neq k}^{m} \nu_j \nu_k \mathrm{tr}\left\{\mathrm{Im}\left[\ddot{\Omega}_j' - \ddot{\Omega}_j\right] \mathrm{Im}\left[\ddot{\Omega}_k - \ddot{\Omega}_k'\right]\right\} \sin(s\lambda_j)\sin(s\lambda_k)$$
$$+ \frac{2}{\pi^2 mn^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} \sum_{j=1}^{m} \sum_{k=1}^{m} \nu_j \nu_k \mathrm{tr}\left\{\mathrm{Re}\left[\ddot{\Omega}_j' + \ddot{\Omega}_j\right] \mathrm{Im}\left[\ddot{\Omega}_k - \ddot{\Omega}_k'\right]\right\} \cos(s\lambda_j)\sin(s\lambda_k).$$

The second and fourth terms are $o(1)$ from $\ddot{\Omega}_j = O(1)$ and Lemma 3 (b) and (d). The fifth term is 0 because $\mathrm{tr}((A' + A)(B - B')) = 0$ for any real matrices $A, B$. Define $\widetilde{H} = \mathcal{E}_0 H$ so that $\mathcal{E}_0(\Lambda_j^0)^{-1}A(\lambda_j) \to \widetilde{H}$ and $\ddot{\Omega}_j \to \ddot{\Omega} = \sum_{a=1}^{q} \eta_a \{\widetilde{H}^*\}_a \widetilde{g}^a \widetilde{H}$ as $\lambda_j \to 0$. Then, from Lemma 3 (a) and (c), the sum of the first and third terms converges to

$$(4\pi^2)^{-1} \mathrm{tr}\left\{\mathrm{Re}\left[\ddot{\Omega}' + \ddot{\Omega}\right] \mathrm{Re}\left[\ddot{\Omega} + \ddot{\Omega}'\right]\right\} + (4\pi^2)^{-1} \mathrm{tr}\left\{\mathrm{Im}\left[\ddot{\Omega}' - \ddot{\Omega}\right] \mathrm{Im}\left[\ddot{\Omega} - \ddot{\Omega}'\right]\right\}.$$

Since $\mathrm{Re}[AA^*] = \mathrm{Re}[A]\mathrm{Re}[A'] + \mathrm{Im}[A]\mathrm{Im}[A']$ and $\mathrm{Re}[AA] = \mathrm{Re}[A]\ \mathrm{Re}[A] - \mathrm{Im}[A]\mathrm{Im}[A]$ for any square matrix $A$, this reduces to

$$2(4\pi^2)^{-1} \mathrm{tr}\left\{\mathrm{Re}\left[\ddot{\Omega}\ddot{\Omega}^*\right] + \mathrm{Re}\left[\ddot{\Omega}\ddot{\Omega}\right]\right\}.$$

Since $\widetilde{G}^0 = \text{Re}[\mathcal{E}_0 G^0 \mathcal{E}_0^*] = \text{Re}[\widetilde{H}\widetilde{H}^*]/2\pi$ and $\overline{G}^0 = \text{Im}[\mathcal{E}_0 G^0 \mathcal{E}_0^*] = \text{Im}[\widetilde{H}\widetilde{H}^*]/2\pi$,

$$\begin{aligned}
(4\pi^2)^{-1}\text{tr}\left\{\text{Re}\left[\ddot{\Omega}\ddot{\Omega}^*\right]\right\} &= (4\pi^2)^{-1}\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\text{Re}\left[\text{tr}\left[\left\{\widetilde{H}^*\right\}_a \widetilde{g}^a\widetilde{H}\widetilde{H}^*(\widetilde{g}^b)'\left\{\widetilde{H}^*\right\}_b^*\right]\right] \\
&= (4\pi^2)^{-1}\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\text{Re}\left[\text{tr}\left[\widetilde{g}^a\widetilde{H}\widetilde{H}^*(\widetilde{g}^b)'\{b\text{th row of } \widetilde{H}\}\left\{\widetilde{H}^*\right\}_a\right]\right] \\
&= \sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\left\{\text{tr}\left[\widetilde{g}^a\widetilde{G}^0(\widetilde{g}^b)'\widetilde{G}^0_{ba}\right] - \text{tr}\left[\widetilde{g}^a\overline{G}^0(\widetilde{g}^b)'\overline{G}^0_{ba}\right]\right\} \\
&= \sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\left[\widetilde{G}^0_{ab}\ (\widetilde{G}^0)^{-1}_{ab}\ + \widetilde{g}^a\overline{G}^0(\widetilde{g}^b)'\overline{G}^0_{ab}\right],
\end{aligned}$$

and

$$\begin{aligned}
(4\pi^2)^{-1}\text{tr}\left\{\text{Re}\left[\ddot{\Omega}\ddot{\Omega}\right]\right\} &= (4\pi^2)^{-1}\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\text{Re}\left[\text{tr}\left[\left\{\widetilde{H}^*\right\}_a\widetilde{g}^a\widetilde{H}\left\{\widetilde{H}^*\right\}_b\widetilde{g}^b\widetilde{H}\right]\right] \\
&= (4\pi^2)^{-1}\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\text{Re}\left[\text{tr}\left[\widetilde{H}\left\{\widetilde{H}^*\right\}_a\widetilde{g}^a\widetilde{H}\left\{\widetilde{H}^*\right\}_b\widetilde{g}^b\right]\right] \\
&= \sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\left\{\text{tr}\left[\{\widetilde{G}^0\}_a\widetilde{g}^a\{\widetilde{G}^0\}_b\widetilde{g}^b\right] - \text{tr}\left[\{\overline{G}^0\}_a\widetilde{g}^a\{\overline{G}^0\}_b\widetilde{g}^b\right]\right\} \\
&= \sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b\left[\mathbf{1}\{a=b\} - \left\{(\widetilde{G}^0)^{-1}\overline{G}^0\right\}_{ab}\left\{(\widetilde{G}^0)^{-1}\overline{G}^0\right\}_{ba}\right].
\end{aligned}$$

Therefore, (46) follows and the proof completes.

### 8.4.2 Hessian approximation

$\Pr(\ddot{d} \notin M) \to 0$ follows from replacing $A(d)$ and $\Xi(d)$ in the proof of Theorem 2 with $\widetilde{A}(d)$ and $\widetilde{\Xi}(d)$. Now, in place of (33), we have

$$\frac{\partial^2\widetilde{R}(d)}{\partial d_a\partial d_b} = \text{tr}\left[-\widetilde{G}^{-1}(d)\frac{\partial\widetilde{G}(d)}{\partial d_a}\widetilde{G}^{-1}(d)\frac{\partial\widetilde{G}(d)}{\partial d_b} + \widetilde{G}^{-1}(d)\frac{\partial^2\widetilde{G}(d)}{\partial d_a\partial d_b}\right]. \tag{47}$$

Define $\widetilde{G}_k(d) = m^{-1}\sum_{j=1}^m(\log\lambda_j)^k\text{Re}[\text{diag}(\lambda_j^{da})I_j\text{diag}(\lambda_j^{da})]$ for $k = 0,1,2$, then $\widetilde{G}_0(d) = \widetilde{G}(d)$ and

$$\begin{aligned}
\partial\widetilde{G}(d)/\partial d_a &= i_a\widetilde{G}_1(d) + \widetilde{G}_1(d)i_a, \\
\partial^2\widehat{G}(d)/\partial d_a\partial d_b &= i_ai_b\widetilde{G}_2(d) + i_a\widetilde{G}_2(d)i_b + i_b\widetilde{G}_2(d)i_a + \widetilde{G}_2(d)i_ai_b.
\end{aligned}$$

Using an argument similar to (35) and (36), we obtain, in place of (34),

$$\begin{aligned}
\widetilde{G}_k(d) &= \text{Re}\left[\text{diag}\left(e^{i\pi d_a^0/2}\right)G^0\text{diag}\left(e^{-i\pi d_a^0/2}\right)\right]m^{-1}\textstyle\sum_{j=1}^m(\log\lambda_j)^k + o_p\left((\log n)^{k-2}\right) \\
&= \widetilde{G}^0 m^{-1}\textstyle\sum_{j=1}^m(\log\lambda_j)^k + o_p\left((\log n)^{k-2}\right),
\end{aligned}$$

uniformly in $d \in M$. Define $\widetilde{G}^0_{1a} = i_a \widetilde{G}^0 + \widetilde{G}^0 i_a$ and $\widetilde{G}^0_{2ab} = i_a i_b \widetilde{G}^0 + i_a \widetilde{G}^0 i_b + i_b \widetilde{G}^0 i_a + \widetilde{G}^0 i_a i_b$ and repeat the arguments following (34), then it follows that $\partial^2 \widetilde{R}(\ddot{d})/\partial d_a \partial d_b = \mathrm{tr}[(\widetilde{G}^0)^{-1} \widetilde{G}^0_{2ab}] + o_p(1)$, giving (44) and completing the proof. □

# 9 Appendix B: technical lemmas

Lemmas 5.4 and 5.5 of Shimotsu and Phillips (2005) are given for the convenience of readers and are to be removed from the final version.

**Shimotsu and Phillips (2005), Lemma 5.4**

*For* $\kappa \in (0,1)$, *as* $m \to \infty$,

$$(a) \qquad \sup_{-C \le \gamma \le C} \left| \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left( \frac{j}{m} \right)^{\gamma} - \int_{\kappa}^{1} x^{\gamma} dx \right| = O\left(m^{-1}\right),$$

$$(b) \qquad \begin{array}{c} \sup_{-C \le \gamma \le C} |m^{-1} \sum_{j=[\kappa m]}^{m} (j/m)^{\gamma}| = O(1), \\ \liminf_{m \to \infty} \inf_{-C \le \gamma \le C} |m^{-1} \sum_{j=[\kappa m]}^{m} (j/m)^{\gamma}| > \varepsilon > 0. \end{array}$$

**Shimotsu and Phillips (2005), Lemma 5.5**

*For* $p \sim m/e$ *as* $m \to \infty$ *and* $\Delta \in (0, 1/(2e))$, *there exist* $\varepsilon \in (0, 0.1)$ *and* $\bar{\kappa} \in (0, 1/4)$ *such that, for sufficiently large* $m$ *and all fixed* $\kappa \in (0, \bar{\kappa})$,

$$(a) \inf_{-C \le \gamma \le -1+2\Delta} \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left( \frac{j}{p} \right)^{\gamma} \ge 1 + 2\varepsilon, \quad (b) \inf_{1 \le \gamma \le C} \frac{1}{m} \sum_{j=[\kappa m]}^{m} \left( \frac{j}{p} \right)^{\gamma} \ge 1 + 2\varepsilon.$$

**Lemma 1** *Let* $A_a(\lambda_j)$ *be the* $a$*th row of* $A(\lambda_j) = \sum_{k=0}^{\infty} A_k e^{ik\lambda_j}$ *and* $A_b^*(\lambda_j)$ *be the* $b$*th column of* $A^*(\lambda_j)$.
*(a) Under the assumptions of Theorem 1, as* $n \to \infty$, *for* $1 \le v < r \le m$,

$$\max_{a,b} \sum_{j=v}^{r} \left( e^{i(\lambda_j - \pi)(d_a^0 - d_b^0)/2} \lambda_j^{d_a^0 + d_b^0} w_{aj} w_{bj}^* - G_{ab}^0 \right) = A_{vr} + B_{vr},$$

*where* $\max_{1 \le v < r \le m} |r^{-1} A_{vr}| = o_p(1)$ *and* $E|B_{vr}| = O(r^{1/2} \log r)$.
*(b) Under the assumptions of Theorem 2, as* $n \to \infty$, *for* $1 \le v < r \le m$,

$$\begin{aligned} (b1) \qquad & \max_{a,b} \sum_{j=v}^{r} e^{i(\lambda_j - \pi)(d_a^0 - d_b^0)/2} \lambda_j^{d_a^0 + d_b^0} \left( w_{aj} w_{bj}^* - A_a(\lambda_j) I_{\varepsilon j} A_b^*(\lambda_j) \right) \\ & = O_p(r^{1/3} (\log r)^{2/3} + \log r + r^{1/2} n^{-1/4}). \\ (b2) \qquad & \max_{a,b} \sum_{j=v}^{r} \left( e^{i(\lambda_j - \pi)(d_a^0 - d_b^0)/2} \lambda_j^{d_a^0 + d_b^0} w_{aj} w_{bj}^* - G_{ab}^0 \right) = O_p(r^{\beta+1} n^{-\beta} + r^{1/2} \log r). \end{aligned}$$

**Proof** Decompose the term inside the summation as $H_{1j} + H_{2j} + H_{3j}$, where

$$\begin{aligned} H_{1j} &= e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\lambda_j^{d_a^0+d_b^0}\left[w_{aj}w_{bj}^* - A_a(\lambda_j)I_{\varepsilon j}A_b^*(\lambda_j)\right] \\ H_{2j} &= e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\lambda_j^{d_a^0+d_b^0}\left[A_a(\lambda_j)I_{\varepsilon j}A_b^*(\lambda_j) - f_{ab}(\lambda_j)\right] \\ H_{3j} &= e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\lambda_j^{d_a^0+d_b^0}f_{ab}(\lambda_j) - G_{ab}^0. \end{aligned}$$

We prove part (a) first. Assumption 1 implies that, for any $\eta > 0$, $n$ can be chosen such that $\max_{a,b}|H_{3j}| \le \eta$ uniformly in $j = 1, \ldots, m$, and $\max_{v,r}\max_{a,b}|r^{-1}\sum_{j=v}^r H_{3j}| = o(1)$ follows. For the contribution from $H_{1j}$, from the proof of Theorem 2 of Robinson (1995a) (also see Robinson (1995b) p. 1673) we have

$$\begin{cases} EI_j = f_j\{1 + O(j^{-1}\log(j+1))\}, \\ Ew_{aj}w_{\varepsilon j}^* = A_a(\lambda_j)/2\pi + O(j^{-1}\log(j+1)\lambda_j^{-d_a}), \\ EI_{\varepsilon j} = I_n/2\pi + O(j^{-1}\log(j+1)), \end{cases} \quad j = 1, \ldots, m. \tag{48}$$

Rewrite $H_{1j}$ as

$$e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\lambda_j^{d_a^0+d_b^0}\left\{\left[w_{aj} - A_a(\lambda_j)w_{\varepsilon j}\right]w_{bj}^* + A_a(\lambda_j)w_{\varepsilon j}\left[w_{bj}^* - w_{\varepsilon j}^*A_b^*(\lambda_j)\right]\right\}. \tag{49}$$

From (48), $A_a(\lambda_j)A_a^*(\lambda_j)/2\pi = f_{aa}(\lambda_j)$, and $\lambda_j^{2d_a^0}f_{aa}(\lambda_j) \sim G_{aa}^0$, we obtain

$$E\left|w_{aj} - A_a(\lambda_j)w_{\varepsilon j}\right|^2 = O(\lambda_j^{-2d_a^0}j^{-1}\log(j+1)), \quad Ew_{bj}w_{bj}^* = O(\lambda_j^{-2d_b^0}),$$

and similarly for $E|w_{bj}^* - w_{\varepsilon j}^*A_b^*(\lambda_j)|^2$. Therefore, applying the Cauchy-Schwartz inequality to the terms in the brace in (49) gives $E|(49)| = O(j^{-1/2}\log(j+1))$ and $\max_{a,b}E|\sum_{j=v}^r H_{1j}| = O(r^{1/2}\log m)$ follows.

For the contribution from $H_{2j}$, as in Lobato (1999, p.148) use $I_{\varepsilon j} = (2\pi n)^{-1}(\sum_{t=1}^n \varepsilon_t\varepsilon_t' + \sum\sum_{s\neq t}\varepsilon_s\varepsilon_t' e^{i(s-t)\lambda_j})$ to rewrite $\sum_{j=v}^r H_{2j}$ as

$$e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\frac{1}{2\pi}\sum_{j=v}^r \lambda_j^{d_a^0+d_b^0}A_a(\lambda_j)\frac{1}{n}\sum_{t=1}^n\left(\varepsilon_t\varepsilon_t' - I_q\right)A_b^*(\lambda_j) \tag{50}$$

$$+e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\frac{1}{2\pi}\sum_{j=v}^r \lambda_j^{d_a^0+d_b^0}A_a(\lambda_j)\left(\frac{1}{n}\sum\sum_{s\neq t}\varepsilon_s\varepsilon_t' e^{i(s-t)\lambda_j}\right)A_b^*(\lambda_j). \tag{51}$$

(50) is $o_p(r)$ uniformly in $1 \le v < r \le m$ because $n^{-1}\sum_1^n(\varepsilon_t\varepsilon_t' - I_q) \to_p 0$ from Theorem 1 of Heyde and Senata (1972) and $||A_b^*(\lambda_j)A_a(\lambda_j)|| = O(\lambda_j^{-d_a^0-d_b^0})$. Rewrite (51) as $e^{i(\lambda_j-\pi)(d_a^0-d_b^0)/2}\sum\sum_{s\neq t}\varepsilon_t'\Xi_{t-s}\varepsilon_s$, where

$$\Xi_{t-s} = \frac{1}{2\pi n}\sum_{j=v}^r \lambda_j^{d_a^0+d_b^0}A_b^*(\lambda_j)A_a(\lambda_j)e^{i(s-t)\lambda_j}.$$

$\sum\sum_{s\neq t}\varepsilon_t'\Xi_{t-s}\varepsilon_s$ has mean zero and variance $\sum\sum_{s\neq t}(\text{vec}'\Xi_{t-s}\text{vec}\Xi_{t-s}+\text{vec}'\Xi_{t-s}'\text{vec}\Xi_{s-t})$. Trivially, we have $||\Xi_s|| = ||\Xi_{-s}|| = ||\Xi_{n-s}||$ and $\Xi_s = O(rn^{-1})$, and summation by parts gives $\Xi_s = O(s^{-1}\log r)$ for $1 \leq s \leq n/2$. Therefore, the variance of $\sum\sum_{s\neq t}\varepsilon_t'\Xi_{t-s}\varepsilon_s$ is $O(n\sum_{h=1}^{[n/r]}||\Xi_h||^2 + n\sum_{h=[n/r]}^{[n/2]}||\Xi_h||^2) = O(r(\log r)^2)$, and part (a) follows.

For part (b), (b1) holds because $\max_{a,b}|\sum_{j=v}^{r}H_{1j}| = O_p(r^{1/3}(\log r)^{2/3} + \log r + r^{1/2}n^{-1/4})$, which follows from applying the proof of (C.2) in Lobato (1999). For (b2), in addition to the bound on $\max_{a,b}|\sum_{j=v}^{r}H_{1j}|$, we have $\max_{a,b}|\sum_{j=v}^{r}H_{2j}| = O_p(r^{1/2}\log r)$ because (50)$= O_p(r^{1/2})$ since $n^{-1}\sum_{t=1}^{n}(\varepsilon_t\varepsilon_t' - I_q) = O_p(n^{-1/2})$ from Assumption 2′ and (51)$= O_p(r^{1/2}\log r)$. Assumption 1′ implies $\max_{a,b}|\sum_{j=v}^{r}H_{3j}| = O(r^{\beta+1}n^{-\beta})$, giving (b2). □

**Lemma 2** *For $p \sim m/e$ as $m \to \infty$, $\Delta \in (0,1)$, and $\kappa \in (0,1)$, we have, for sufficiently large $m$,*

$$\inf_{-1+2\Delta\leq\gamma\leq 1}\frac{1}{m}\sum_{j=[\kappa m]}^{m}\left(\frac{j}{p}\right)^{\gamma} \geq 1 - \kappa^{2\Delta} + o(1).$$

**Proof** It follows from Lemma 5.4 of Shimotsu and Phillips (2005) that

$$\frac{1}{m}\sum_{j=[\kappa m]}^{m}\left(\frac{j}{p}\right)^{\gamma} = \left(\frac{m}{p}\right)^{\gamma}\frac{1}{m}\sum_{j=[\kappa m]}^{m}\left(\frac{j}{m}\right)^{\gamma} = e^{\gamma}\int_{\kappa}^{1}x^{\gamma}dx + o(1) = \frac{e^{\gamma}(1-\kappa^{\gamma+1})}{\gamma+1} + o(1).$$

The stated result follows because $e^{\gamma}/(\gamma+1) \geq 1$ for $\gamma \in [-1+2\Delta, 1]$. □

**Lemma 3** *For $j,k = 1,\ldots,m$ with $m = O(n)$, as $n \to \infty$,*

(*a*) $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos^2(s\lambda_j) = (1/4)n^2 + o(n^2)$,
(*b*) $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos(s\lambda_j)\cos(s\lambda_k) = O(n)$, $\quad j \neq k$,
(*c*) $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\sin^2(s\lambda_j) = (1/4)n^2 + o(n^2)$,
(*d*) $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\sin(s\lambda_j)\sin(s\lambda_k) = O(n)$, $\quad j \neq k$,

**Proof** Robinson (1995b, p. 1645) shows that $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos^2(s\lambda_j) = (n-1)^2/4$, $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos(s\lambda_j) = -n/2$, and $\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\cos(s\lambda_j)\cos(s\lambda_k) = -n/2$ for $j,k = 1,\ldots,m < \frac{1}{2}n$, $j \neq k$, giving parts (a) and (b). Part (c) follows from

$$\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\sin^2(s\lambda_j) = \sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\left\{1-\cos^2(s\lambda_j)\right\} = \frac{n(n-1)}{2} - \frac{(n-1)^2}{4} = \frac{n^2(1+o(1))}{4}.$$

Part (d) follows from $2\sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\sin(s\lambda_j)\sin(s\lambda_k) = \sum_{t=1}^{n-1}\sum_{s=1}^{n-t}\{\cos(s\lambda_{j-k}) - \cos(s\lambda_{j+k})\} = O(n)$. □

## References


Bollerslev, T., Wright, J.H., 2000. Semiparametric Estimation of Long-Memory Volatility Dependencies: The Role of High-Frequency Data. Journal of Econometrics 98, 81-106.

Brunetti, C., Gilbert, C.L., 2000. Bivariate FIGARCH and Fractional Cointegration. Journal of Empirical Finance 7, 509-530.

Hannan, E.J., 1970. Multiple Time Series. Wiley.

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

Heyde, C.C., Senata, E., 1972. Estimation Theory for Growth and Immigration Rates in a Multiplicative Process. Journal of Applied Probability 9, 235-256.

Horn, R.A., Johnson, C.R., 1985. Matrix Analysis, Cambridge University Press.

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

Künsch, H., 1987. Statistical aspects of self-similar processes. In: Prokhorov, Yu., Sazanov, V.V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society (VNU Science Press, Utrecht) 67-74.

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

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

Lobato, I.N., Robinson, P.M., 1998. A nonparametric test for I(0). Review of Economic Studies 65, 475-495.

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

Lütkepohl, H., 1996. Handbook of Matrices. Wiley.

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

Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048-1072.

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

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

**Table 1. Comparison of asymptotic variance**

| $\rho$ | | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 |
|---|---|---|---|---|---|---|
| univariate | | 0.250 | 0.250 | 0.250 | 0.250 | 0.250 |
| $(\Omega^{-1})_{11}$ | | 0.250 | 0.234 | 0.200 | 0.167 | 0.142 |
| $Q_{11}$ | $\|d_1 - d_2\| = 0.0$ | 0.250 | 0.245 | 0.230 | 0.205 | 0.170 |
| $Q_{11}$ | $\|d_1 - d_2\| = 0.2$ | 0.250 | 0.245 | 0.232 | 0.211 | 0.188 |
| $Q_{11}$ | $\|d_1 - d_2\| = 0.4$ | 0.250 | 0.247 | 0.238 | 0.225 | 0.218 |
| $(\Omega^{-1})_{11}/0.25$ | | 1.000 | 0.937 | 0.801 | 0.669 | 0.570 |
| $Q_{11}/0.25$ | $\|d_1 - d_2\| = 0.0$ | 1.000 | 0.980 | 0.920 | 0.820 | 0.680 |
| $Q_{11}/0.25$ | $\|d_1 - d_2\| = 0.2$ | 1.000 | 0.982 | 0.929 | 0.845 | 0.753 |
| $Q_{11}/0.25$ | $\|d_1 - d_2\| = 0.4$ | 1.000 | 0.987 | 0.951 | 0.901 | 0.870 |

**Table 2. Simulation results:** $n = 512$, $m = n^{0.65} = 57$

| | GSE1 | | | | GSE2 | | | |
|---|---|---|---|---|---|---|---|---|
| | bias | s.d. | RMSE | $\frac{\text{var(GSE1)}}{\text{var(uni)}}$ | bias | s.d. | RMSE | $\frac{\text{var(GSE2)}}{\text{var(uni)}}$ |
| | | | | $\rho = 0.0$ | | | | |
| | | | | $(d_1, d_2) = (0.2, -0.2)$ | | | | |
| $d_1$ | -0.0069 | 0.0777 | 0.0780 | 1.0173 | -0.0071 | 0.0773 | 0.0776 | 1.0069 |
| $d_2$ | -0.0035 | 0.0780 | 0.0780 | 1.0159 | -0.0037 | 0.0775 | 0.0776 | 1.0049 |
| | | | | $(d_1, d_2) = (0.2, 0.2)$ | | | | |
| $d_1$ | -0.0065 | 0.0772 | 0.0775 | 1.0083 | -0.0066 | 0.0770 | 0.0772 | 1.0018 |
| $d_2$ | -0.0051 | 0.0777 | 0.0779 | 1.0182 | -0.0052 | 0.0772 | 0.0784 | 1.0053 |
| | | | | $(d_1, d_2) = (0.2, 0.4)$ | | | | |
| $d_1$ | -0.0057 | 0.0775 | 0.0777 | 1.0168 | -0.0058 | 0.0770 | 0.0772 | 1.0034 |
| $d_2$ | -0.0020 | 0.0779 | 0.0780 | 1.0135 | -0.0020 | 0.0776 | 0.0776 | 1.0054 |

**Table 3. Simulation results:** $n = 512$, $m = n^{0.65} = 57$

| | GSE1 | | | | GSE2 | | | |
|---|---|---|---|---|---|---|---|---|
| | bias | s.d. | RMSE | $\frac{\text{var(GSE1)}}{\text{var(uni)}}$ | bias | s.d. | RMSE | $\frac{\text{var(GSE2)}}{\text{var(uni)}}$ |
| | | | | $\rho = 0.4$ | | | | |
| | | | | $(d_1, d_2) = (0.2, -0.2)$ | | | | |
| $d_1$ | -0.0053 | 0.0694 | 0.0696 | 0.8101 | -0.0049 | 0.0751 | 0.0753 | 0.9509 |
| $d_2$ | -0.0021 | 0.0704 | 0.0704 | 0.8135 | -0.0019 | 0.0760 | 0.0761 | 0.9500 |
| | | | | $(d_1, d_2) = (0.2, 0.2)$ | | | | |
| $d_1$ | -0.0056 | 0.0691 | 0.0694 | 0.8082 | -0.0067 | 0.0738 | 0.0741 | 0.9197 |
| $d_2$ | -0.0044 | 0.0688 | 0.0689 | 0.7969 | -0.0051 | 0.0742 | 0.0744 | 0.9285 |
| | | | | $(d_1, d_2) = (0.2, 0.4)$ | | | | |
| $d_1$ | -0.0038 | 0.0692 | 0.0693 | 0.8105 | -0.0052 | 0.0741 | 0.0743 | 0.9289 |
| $d_2$ | -0.0017 | 0.0690 | 0.0690 | 0.7982 | -0.0013 | 0.0748 | 0.0748 | 0.9370 |

**Table 4. Simulation results:** $n = 512$, $m = n^{0.65} = 57$

| | GSE1 | | | | GSE2 | | | |
|---|---|---|---|---|---|---|---|---|
| | bias | s.d. | RMSE | $\frac{\text{var(GSE1)}}{\text{var(uni)}}$ | bias | s.d. | RMSE | $\frac{\text{var(GSE2)}}{\text{var(uni)}}$ |
| | | | | $\rho = 0.8$ | | | | |
| | | | | $(d_1, d_2) = (0.2, -0.2)$ | | | | |
| $d_1$ | 0.0017 | 0.0594 | 0.0594 | 0.5936 | 0.0059 | 0.0711 | 0.0714 | 0.8523 |
| $d_2$ | 0.0046 | 0.0604 | 0.0606 | 0.6026 | 0.0090 | 0.0721 | 0.0727 | 0.8578 |
| | | | | $(d_1, d_2) = (0.2, 0.2)$ | | | | |
| $d_1$ | -0.0043 | 0.0581 | 0.0583 | 0.5708 | -0.0066 | 0.0635 | 0.0638 | 0.6820 |
| $d_2$ | -0.0038 | 0.0576 | 0.0577 | 0.5607 | -0.0054 | 0.0637 | 0.0640 | 0.6864 |
| | | | | $(d_1, d_2) = (0.2, 0.4)$ | | | | |
| $d_1$ | 0.0003 | 0.0585 | 0.0585 | 0.5795 | -0.0009 | 0.0662 | 0.0662 | 0.7412 |
| $d_2$ | 0.0004 | 0.0582 | 0.0582 | 0.5653 | 0.0033 | 0.0671 | 0.0672 | 0.7528 |

**Table 5. Simulation results:** $n = 128$, $m = n^{0.65} = 23$

| | **GSE1 (2-step)** | | | **GSE1 (NR)** | | |
|---|---|---|---|---|---|---|
| | bias | s.d. | RMSE | bias | s.d. | RMSE |
| | | | $\rho = 0.4$, $(d_1, d_2) = (0.2, -0.2)$ | | | |
| $d_1$ | -0.0173 | 0.1265 | 0.1277 | -0.0139 | 0.1239 | 0.1247 |
| $d_2$ | -0.0095 | 0.1268 | 0.1272 | -0.0054 | 0.1237 | 0.1239 |
| | | | $\rho = 0.8$, $(d_1, d_2) = (0.2, 0.4)$ | | | |
| $d_1$ | -0.0175 | 0.1304 | 0.1316 | -0.0070 | 0.1042 | 0.1044 |
| $d_2$ | -0.0190 | 0.1257 | 0.1271 | -0.0078 | 0.1041 | 0.1044 |
| | **GSE2 (2-step)** | | | **GSE2 (NR)** | | |
| | bias | s.d. | RMSE | bias | s.d. | RMSE |
| | | | $\rho = 0.4$, $(d_1, d_2) = (0.2, -0.2)$ | | | |
| $d_1$ | -0.0142 | 0.1341 | 0.1349 | -0.0133 | 0.1341 | 0.1347 |
| $d_2$ | -0.0061 | 0.1341 | 0.1343 | -0.0051 | 0.1337 | 0.1338 |
| | | | $\rho = 0.8$, $(d_1, d_2) = (0.2, 0.4)$ | | | |
| $d_1$ | -0.0122 | 0.1200 | 0.1206 | -0.0085 | 0.1198 | 0.1201 |
| $d_2$ | -0.0078 | 0.1194 | 0.1197 | -0.0041 | 0.1193 | 0.1194 |

**Table 6. Simulation results:** $n = 512$, $m = n^{0.65} = 57$

| | **GSE1 (2-step)** | | | **GSE1 (NR)** | | |
|---|---|---|---|---|---|---|
| | bias | s.d. | RMSE | bias | s.d. | RMSE |
| | | | $\rho = 0.4$, $(d_1, d_2) = (0.2, -0.2)$ | | | |
| $d_1$ | -0.0068 | 0.0695 | 0.0698 | -0.0053 | 0.0693 | 0.0695 |
| $d_2$ | -0.0037 | 0.0705 | 0.0706 | -0.0021 | 0.0704 | 0.0704 |
| | | | $\rho = 0.8$, $(d_1, d_2) = (0.2, 0.4)$ | | | |
| $d_1$ | -0.0052 | 0.0589 | 0.0592 | -0.0010 | 0.0585 | 0.0585 |
| $d_2$ | -0.0055 | 0.0595 | 0.0597 | -0.0007 | 0.0592 | 0.0592 |
| | **GSE2 (2-step)** | | | **GSE2 (NR)** | | |
| | bias | s.d. | RMSE | bias | s.d. | RMSE |
| | | | $\rho = 0.4$, $(d_1, d_2) = (0.2, -0.2)$ | | | |
| $d_1$ | -0.0053 | 0.0751 | 0.0753 | -0.0049 | 0.0751 | 0.0753 |
| $d_2$ | -0.0022 | 0.0760 | 0.0761 | -0.0019 | 0.0760 | 0.0761 |
| | | | $\rho = 0.8$, $(d_1, d_2) = (0.2, 0.4)$ | | | |
| $d_1$ | -0.0038 | 0.0667 | 0.0668 | -0.0020 | 0.0667 | 0.0667 |
| $d_2$ | 0.0001 | 0.0682 | 0.0682 | 0.0019 | 0.0682 | 0.0682 |

**Table 7. Simulation results:** $n = 512$, $m = n^{0.65} = 57$

| $2\pi G_{11} = 2\pi G_{12} = 1.0, \quad 2\pi G_{12} = 2\pi G_{21} = 0.4$ | | | | | | |
|---|---|---|---|---|---|---|
| | **mean(GSE1)** | | | **mean(GSE2)** | | |
| $(d_1, d_2)$ | $2\pi\widehat{G}_{11}$ | $2\pi\widehat{G}_{12}$ | $2\pi\widehat{G}_{22}$ | $2\pi\widehat{G}_{11}$ | $2\pi\widehat{G}_{12}$ | $2\pi\widehat{G}_{22}$ |
| $(0.2, -0.2)$ | 1.0274 | 0.4082 | 1.0148 | 1.0280 | 0.3447 | 1.0157 |
| $(0.2, 0.2)$ | 1.0300 | 0.4121 | 1.0243 | 1.0340 | 0.4105 | 1.0279 |
| $(0.2, 0.4)$ | 1.0244 | 0.4110 | 1.0318 | 1.0296 | 0.3937 | 1.0325 |

**Table 8. Simulation results:** $n = 128$, $m = n^{0.65} = 23$

| | | | rejection frequencies ($W$) | | | rejection frequencies ($W_c$) | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | $d_1$ | $d_2$ | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 |
| 0.0 | 0.2 | -0.2 | 0.2802 | 0.1924 | 0.0866 | 0.1471 | 0.0874 | 0.0280 |
| 0.0 | 0.2 | 0.2 | 0.2801 | 0.1961 | 0.0921 | 0.1510 | 0.0927 | 0.0328 |
| 0.0 | 0.2 | 0.4 | 0.2899 | 0.2020 | 0.0933 | 0.1562 | 0.0947 | 0.0326 |
| 0.4 | 0.2 | -0.2 | 0.2706 | 0.1847 | 0.0819 | 0.1375 | 0.0825 | 0.0248 |
| 0.4 | 0.2 | 0.2 | 0.2673 | 0.1854 | 0.0881 | 0.1451 | 0.0884 | 0.0289 |
| 0.4 | 0.2 | 0.4 | 0.2814 | 0.1975 | 0.0858 | 0.1504 | 0.0871 | 0.0260 |
| 0.8 | 0.2 | -0.2 | 0.2688 | 0.1843 | 0.0769 | 0.1380 | 0.0775 | 0.0227 |
| 0.8 | 0.2 | 0.2 | 0.2580 | 0.1798 | 0.0797 | 0.1395 | 0.0802 | 0.0251 |
| 0.8 | 0.2 | 0.4 | 0.2736 | 0.1884 | 0.0819 | 0.1438 | 0.0829 | 0.0252 |

**Table 9. Simulation results:** $n = 512$, $m = n^{0.65} = 57$

| | | | rejection frequencies ($W$) | | | rejection frequencies ($W_c$) | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | $d_1$ | $d_2$ | 0.10 | 0.05 | 0.01 | 0.10 | 0.05 | 0.01 |
| 0.0 | 0.2 | -0.2 | 0.1929 | 0.1188 | 0.0422 | 0.1272 | 0.0695 | 0.0198 |
| 0.0 | 0.2 | 0.2 | 0.1891 | 0.1191 | 0.0403 | 0.1277 | 0.0705 | 0.0185 |
| 0.0 | 0.2 | 0.4 | 0.1936 | 0.1211 | 0.0412 | 0.1284 | 0.0726 | 0.0188 |
| 0.4 | 0.2 | -0.2 | 0.1942 | 0.1217 | 0.0397 | 0.1294 | 0.0723 | 0.0162 |
| 0.4 | 0.2 | 0.2 | 0.1832 | 0.1132 | 0.0370 | 0.1205 | 0.0652 | 0.0173 |
| 0.4 | 0.2 | 0.4 | 0.1845 | 0.1142 | 0.0379 | 0.1211 | 0.0678 | 0.0166 |
| 0.8 | 0.2 | -0.2 | 0.2005 | 0.1226 | 0.0427 | 0.1304 | 0.0720 | 0.0203 |
| 0.8 | 0.2 | 0.2 | 0.1773 | 0.1048 | 0.0342 | 0.1127 | 0.0586 | 0.0163 |
| 0.8 | 0.2 | 0.4 | 0.1819 | 0.1106 | 0.0364 | 0.1182 | 0.0644 | 0.0134 |