

Asymptotics of bivariate local Whittle estimators with applications to fractal connectivity^{*†}

Changryong Baek
Sungkyunkwan University

Stefanos Kechagias
SAS Institute

Vladas Pipiras
University of North Carolina

August 1, 2019

Abstract

Several methodological and numerical issues behind the local Whittle estimation of long and short memory in bivariate stationary time series with possible fractional cointegration are reexamined. These issues include the asymptotic normality for all model parameters, local Whittle plots for phase parameter and fractal connectivity, and others. For fractal connectivity, in particular, it is advocated to work with a model parametrization for which the model parameters associated with this phenomenon are identifiable and could be tested naturally within the local Whittle estimation framework. A simulation study and data applications are also considered.

1 Introduction

Our goal is to clarify a number of issues behind the local Whittle estimation in bivariate stationary systems proposed and studied by Robinson (2008), with particular attention to long memory (see also Nielsen (2007), Nielsen and Shimotsu (2007), Shimotsu (2007, 2012), Nielsen (2011)). More specifically, consider a bivariate (second-order) stationary time series $\{X_n\}_{n \in \mathbb{Z}} = \{(X_{1,n}, X_{2,n})'\}_{n \in \mathbb{Z}}$ with zero mean (for simplicity), the autocovariance matrix function $\gamma_X(h) = \mathbb{E}X_{n+h}X_n'$, $h \in \mathbb{Z}$, and the spectral density matrix $f_X(\lambda) = (f_{X,kl}(\lambda))_{k,l=1,2}$, $\lambda \in (-\pi, \pi)$, assumed to exist and, by convention, to satisfy $\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f_X(\lambda) d\lambda$. Suppose that as $\lambda \rightarrow 0^+$,

$$f_X(\lambda) = \begin{pmatrix} f_{X,11}(\lambda) & f_{X,12}(\lambda) \\ f_{X,21}(\lambda) & f_{X,22}(\lambda) \end{pmatrix} \sim \begin{pmatrix} g_{11}\lambda^{-2d_1} & g_{12}\lambda^{-d_1-d_2} \\ g_{21}\lambda^{-d_1-d_2} & g_{22}\lambda^{-2d_2} \end{pmatrix} = \lambda^{-D} G \lambda^{-D}, \quad (1.1)$$

where $D = \text{diag}(d_1, d_2)$, $-1/2 < d_1, d_2 < 1/2$, $\lambda^{-D} = \text{diag}(\lambda^{-d_1}, \lambda^{-d_2})$, and $G = (g_{kl})$ is Hermitian symmetric and positive definite (i.e. $g_{11} > 0, g_{22} > 0, \overline{g_{12}} = g_{21}$ and $g_{11}g_{22} - |g_{12}|^2 > 0$). In (1.1), the asymptotic equivalence $\sim$ is entry-wise and $f(\lambda) \sim C\lambda^{-p}$ for a constant C and a power p means that $f(\lambda)\lambda^p \rightarrow C$, including the case when $C = 0$. For univariate series, the case $d > 0$ is associated with long memory, $d \leq 0$ with short memory and $d < 0$ with anti-persistence (e.g. Robinson (2003), Giraitis, Koul and Surgailis (2012), Beran, Feng, Ghosh and Kulik (2013), Pipiras and Taqqu (2017)). Writing $g_{11} = \omega_{11} > 0$, $g_{22} = \omega_{22} > 0$, $g_{12} = \omega_{12}e^{-i\phi}$, $g_{21} = \omega_{12}e^{i\phi} =: \omega_{21}e^{i\phi}$ with $\omega_{12} \in \mathbb{R}$, $\phi \in (-\pi/2, \pi/2)$ and setting $\Omega = (\omega_{kl})_{k,l=1,2}$, the relation (1.1) can also be expressed as

$$f_X(\lambda) \sim \Phi_{D,\phi}(\lambda)^{-1} \Omega \overline{\Phi}_{D,\phi}(\lambda)^{-1}, \quad \text{as } \lambda \rightarrow 0^+, \quad (1.2)$$

^{*}AMS subject classification. Primary: 62M10, 62H12. Secondary: 62F12.

[†]Keywords and phrases: Long memory, multivariate time series, phase parameters, local Whittle estimation, fractional cointegration, fractal connectivity.

where $\Phi_{D,\phi}(\lambda) = \text{diag}(\lambda^{d_1}, \lambda^{d_2} e^{-i\phi})$ and Ω is a real-valued, symmetric, positive definite matrix. The bar above $\Phi_{D,\phi}$ in (1.2) indicates complex conjugation componentwise. The parameters d_1, d_2 are known as memory parameters (long-memory parameters when they are positive) and ϕ as a phase parameter (at the zero frequency).

We gather the model parameters above as

$$\text{Parametrization P: } \omega_{11}, \omega_{22}, \omega_{12}, \phi, d_1, d_2. \quad (1.3)$$

It will also be convenient to work in

$$\text{Parametrization C: } \omega_{11}, \omega_{22}, r_1, r_2, d_1, d_2, \quad (1.4)$$

where r_1, r_2 are the real and imaginary part of g_{12} , namely satisfying $r_1 - ir_2 = \omega_{12} e^{-i\phi} = g_{12}$. (The letters “P” in (1.3) and “C” in (1.4) stand for “Polar” and “Complex”, respectively.) One reason that we are interested in Parametrization C is related to the so-called *fractal connectivity*, which is associated with the case

$$\omega_{12} \neq 0 \quad (1.5)$$

and that of fractal non-connectivity with the case $\omega_{12} = 0$. See, for example, Achard, Bassett, Meyer-Lindenberg and Bullmore (2008), Wendt, Scherrer, Abry and Achard (2009), Kristoufek (2013), Wendt, Didier, Combrexelle and Abry (2017). But testing for fractal non-connectivity is not possible in Parametrization P since for $\omega_{12} = 0$, the model parameter ϕ is not identifiable. There is no such issue in Parameterization C, where fractal connectivity $\omega_{12} \neq 0$ is now associated with

$$r_1 \neq 0 \quad \text{or} \quad r_2 \neq 0, \quad (1.6)$$

and the case $\omega_{12} = 0$ with $r_1 = 0$ and $r_2 = 0$.

Working with Parametrization P, Robinson (2008) introduced and studied the local Whittle estimators of Ω , ϕ and D defined as

$$(\hat{\Omega}, \hat{\phi}, \hat{D}) = \underset{(\Omega, \phi, D)}{\text{argmin}} Q(\Omega, \phi, D) \quad (1.7)$$

with

$$Q(\Omega, \phi, D) = \frac{1}{m} \sum_{j=1}^m (\log |\Phi_{D,\phi}(\lambda_j)^{-1} \Omega \bar{\Phi}_{D,\phi}(\lambda_j)^{-1}|) + \text{tr} (\Phi_{D,\phi}(\lambda_j) I_X(\lambda_j) \bar{\Phi}_{D,\phi}(\lambda_j) \Omega^{-1}), \quad (1.8)$$

where $|A|$ denotes the determinant of a matrix A , $\lambda_j = (2\pi j)/N$ are the Fourier frequencies for a sample size N ,

$$I_X(\lambda) = \frac{1}{2\pi N} \left(\sum_{n=1}^N X_n e^{-in\lambda} \right) \left(\sum_{n=1}^N X_n e^{in\lambda} \right)' \quad (1.9)$$

is the periodogram of the series X and m is the number of frequencies used in estimation. The optimization problem (1.7) is taken in theory essentially over $-1/2 < d_1, d_2 < 1/2$, $\phi \in (-\pi/2, \pi/2)$ and positive definite Ω ; in practice, as in their Whittle plots (e.g. Section 6 below), one often allows d_1, d_2 to exceed $1/2$. The asymptotic normality result for $\hat{\phi}, \hat{D}$ is provided in Robinson (2008) under suitable assumptions, in particular, on $m = m(N) \rightarrow \infty$.

Robinson (2008) also considers the case of possible fractional cointegration, where (1.1) is replaced by

$$B f_X(\lambda) B' \sim \lambda^{-D} G \lambda^{-D}, \quad \text{as } \lambda \rightarrow 0^+, \quad (1.10)$$

with

$$B = \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix}, \quad (1.11)$$

the case $\beta = 0$ corresponding to (1.1) with no fractional cointegration and the case $\beta \neq 0$ associated with fractional cointegration. Furthermore, it is assumed that $0 \leq d_1 < d_2 < 1/2$. In the local Whittle estimation (1.7)–(1.8), β is added as another parameter, with the negative log-likelihood having the same form as (1.8) but replacing $I_X(\lambda_j)$ by $BI_X(\lambda_j)B'$. Robinson (2008) also established the asymptotic normality of the local Whittle estimators $\hat{\beta}$, $\hat{\phi}$ and $\hat{D}$. We also note that the spectral density in the fractionally cointegrated case (1.10) with $\beta \neq 0$ can be expressed as (1.1) but with the parameters in (1.1) (not to be confused with those in (1.10)) satisfying

$$d_1 = d_2, \quad |\Omega| = \omega_{11}\omega_{22} - \omega_{12}^2 = 0 \quad (1.12)$$

(see also Remark 2.4 below). But the possibility of the second relation in (1.12) is excluded in the asymptotics of the local Whittle estimation (1.7)–(1.8).

Local Whittle estimation plays a fundamental role in the analysis of time series, especially when long memory is thought to be present. Our work contributes to the understanding of the estimation method in the following ways:

- We consider the asymptotic normality result for *all* model parameters in Parametrization P, and not just ϕ , d_1 , d_2 as in Robinson (2008). Through a series of remarks, we also make a number of additional points related to the established asymptotic normality result: we correct the asymptotic covariance matrix appearing in Robinson (2008), we correct the earlier form of normalization for some local Whittle and related estimators, and others.
- We also establish an asymptotic normality result for all model parameters in Parametrization C. As a consequence, in connection to fractal connectivity, we suggest a test statistic that can be used naturally to test for fractal non-connectivity $r_1 = 0$ and $r_2 = 0$.
- We shall also present asymptotic normality results in Parametrizations P and C with the additional parameter β , for the fractionally cointegrated model (1.10). As for the non-cointegrated model, we again pay particular attention to fractal (non-)connectivity. For example, an analogous test statistic is suggested for testing fractal non-connectivity in the case of fractional cointegration.
- We show that a number of considered optimization problems, such as that in (1.7), can be reduced to problems involving fewer variables.
- In addition to the asymptotic normality results, we consider a number of issues related to local Whittle plots. These are plots of parameter estimates and quantities of interest as functions of the tuning parameter m in (1.8) or (3.2). In this work, we discuss what local Whittle plots, in our view, should be considered in connection to long memory and fractal connectivity. We also slightly modify the local Whittle plot for the phase parameter ϕ .

In all our asymptotic results for both parametrizations, we provide closed form expressions for the limiting covariance matrices of the local Whittle estimators.

The rest of the paper is organized as follows. Our asymptotic normality results are presented in Section 2, and optimization issues are discussed in Section 3. We present a small numerical study to assess the validity and relevance of the derived asymptotic results and suggested test statistics in Section 4. We discuss local Whittle plots in Section 5, and consider real data applications in Section 6. Section 7 contains conclusions, including some open questions. All technical proofs are moved to Appendices A, B and C.

2 Asymptotic normality results

As noted in Section 1, Robinson (2008) derived the asymptotic normality of the local Whittle estimators $\hat{\phi}, \hat{d}_1, \hat{d}_2$. Here, we are interested in all the model parameters in either Parametrization P or Parametrization C, and the asymptotic normality of their local Whittle estimators.

We also consider separately the case of fractional cointegration. We should also mention that though all our results are stated in the form of asymptotic normality, their proofs focus only on the calculation of the respective information matrices. The asymptotic normality itself is somewhat “standard” and could be established as in Robinson (2008). To distinguish between the information and limiting covariance matrices of the two Parametrizations we use the subscripts p (Theorems 2.1 and 2.2) and c (Theorems 2.3 and 2.4) .

2.1 Case of Parametrization P without fractional cointegration

As in Robinson (2008), we shall make the following assumptions. For a matrix function $A(\lambda)$, the notation $A(\lambda) = O(\lambda^p)$ is interpreted as the standard $O(\lambda^p)$ notation but entry-wise.

- (C1) The spectral density $f_X(\lambda)$ of a bivariate stationary series $\{X_n\}$ satisfies (1.2). Furthermore, for some $b \in (0, 2]$, $\Phi_{D,\phi}(\lambda)C(\lambda) - P = O(\lambda^b)$, as $\lambda \rightarrow 0^+$, where P is a real 2×2 matrix satisfying $PP' = \Omega$, and $C(\lambda)$ is a 2×2 matrix-valued function differentiable in a neighborhood of $\lambda = 0$ such that $f_X(\lambda) = C(\lambda)\overline{C}(\lambda)'$ and $\Phi_{D,\phi}(\lambda)dC(\lambda)/d\lambda = O(\lambda^{-1})$, as $\lambda \rightarrow 0^+$.
- (C2) The series $\{X_n\}_{n \in \mathbb{Z}}$ admits a representation $X_n = \mathbb{E}X_n + \sum_{j \in \mathbb{Z}} C_j \varepsilon_{n-j}$, $\sum_{j \in \mathbb{Z}} \|C_j\|^2 < \infty$ with the Euclidean norm $\|\cdot\|$, where $C_j = (2\pi)^{-1} \int_{-\pi}^{\pi} C(\lambda) e^{-ij\lambda} d\lambda$ and $\{\varepsilon_n\}_{n \in \mathbb{Z}}$ is a $p \times 1$ vector series such that ε_n has almost sure constant first, second, third and fourth moments and respective cross-moments conditionally on $\mathcal{F}_{n-1} = \sigma\{\varepsilon_m, m \leq n-1\}$ and such that $\mathbb{E}\varepsilon_n = 0$, $\mathbb{E}\varepsilon_n \varepsilon_n' = I_p$, $\mathbb{E}\varepsilon_n \varepsilon_m = 0$, $n \neq m$. Moreover, $\mathbb{P}(\varepsilon_n' \varepsilon_n > \eta) \leq K\mathbb{P}(X > \eta)$ for all $\eta > 0$ and some scalar non-negative random variable X such that $\mathbb{E}X < \infty$.
- (C3) Let $\theta = (\omega_{11}, \omega_{22}, \omega_{12}, \phi, d_1, d_2)'$ be the parameter vector of interest and such that $\theta \in \Theta_\omega \times \Theta_\phi \times \Theta_D$, where $\Theta_\omega = \{(x, y, z) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} : xy - z^2 > 0\}$, $\Theta_\phi = (-\pi/2, \pi/2)$, $\Theta_D = (-1/2, 1/2)^2$.
- (C4) The number of frequencies m satisfies $(\log m)^2 m^{1+2b}/N^{2b} \rightarrow 0$ and $(\log N)^C/m \rightarrow 0$ as $N \rightarrow \infty$ for any $C < \infty$, where $b \in (0, 2]$ appears in (C1).
- (C5) $0 < |\omega_{12}|/(\omega_{11}\omega_{22})^{1/2} < 1$.

The assumptions (C1)–(C5) are, respectively, the assumptions (B1), (B2), (B4), (B5) and (A6) of Robinson (2008) with (C3) also restricting Ω to be positive definite. Note that Assumption (C5) excludes the fractal non-connectivity case $\omega_{12} = 0$ (cf. (1.5)) and the fractionally cointegrated case $\omega_{11}\omega_{22} - \omega_{12}^2 = 0$ (cf. (1.12)). The next theorem is the asymptotic normality result for all bivariate local Whittle estimators. It is followed by a discussion and a number of remarks that cast further light on the form of the asymptotic covariance matrix. The proof can be found in Appendix A. (As throughout the paper, the notation $\log$ below refers to the natural logarithm.)

Theorem 2.1 Suppose that the assumptions (C1)–(C5) hold. Then, as $N \rightarrow \infty$,

$$\sqrt{m}\hat{\eta}_p := \sqrt{m} \begin{pmatrix} \frac{1}{\log(N/m)}(\hat{\omega}_{11} - \omega_{11}) \\ \frac{1}{\log(N/m)}(\hat{\omega}_{22} - \omega_{22}) \\ \frac{1}{\log(N/m)}(\hat{\omega}_{12} - \omega_{12}) \\ \hat{\phi} - \phi \\ \hat{d}_1 - d_1 \\ \hat{d}_2 - d_2 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Gamma_p), \quad (2.1)$$

where the symmetric limiting covariance matrix Γ_p (presented in compact form) is

$$\Gamma_p = \begin{pmatrix} \frac{\omega_{11}(\omega_{11}\omega_{22}+|\Omega|)}{2\omega_{22}} & \frac{\omega_{12}^2}{2} & \frac{\omega_{11}\omega_{12}}{2} & 0 & -\frac{\omega_{11}\omega_{22}+|\Omega|}{4\omega_{22}} & -\frac{\omega_{12}^2}{4\omega_{22}} \\ & \frac{\omega_{22}(\omega_{11}\omega_{22}+|\Omega|)}{2\omega_{11}} & \frac{\omega_{12}\omega_{22}}{2} & 0 & -\frac{\omega_{12}^2}{4\omega_{11}} & -\frac{\omega_{11}\omega_{22}+|\Omega|}{4\omega_{11}} \\ & & \frac{\omega_{12}^2}{2} & 0 & -\frac{\omega_{12}}{4} & -\frac{\omega_{12}}{4} \\ & & & \frac{|\Omega|}{2\omega_{12}^2} & 0 & 0 \\ & & & & \frac{\omega_{11}\omega_{22}+|\Omega|}{8\omega_{11}\omega_{22}} & \frac{\omega_{12}^2}{8\omega_{11}\omega_{22}} \\ & & & & & \frac{\omega_{11}\omega_{22}+|\Omega|}{8\omega_{11}\omega_{22}} \end{pmatrix}, \quad (2.2)$$

with $|\Omega| = \omega_{11}\omega_{22} - \omega_{12}^2$.

It is instructive to comment first on how the different convergence rates $\sqrt{m}$ and $\sqrt{m}/\log(N/m)$ in (2.1) actually arise, as some of these arguments will take a more central place in the following sections. The information matrix associated with the estimated model parameters (at the rate $\sqrt{m}$) is given in (A.5) of Appendix A in its asymptotic form as:

$$\begin{pmatrix} \frac{\omega_{22}^2}{|\Omega|^2} & \frac{\omega_{12}^2}{|\Omega|^2} & -\frac{2\omega_{12}\omega_{22}}{|\Omega|^2} & 0 & -\frac{2\omega_{22}}{|\Omega|}L & 0 \\ & \frac{\omega_{11}^2}{|\Omega|^2} & -\frac{2\omega_{11}\omega_{12}}{|\Omega|^2} & 0 & 0 & -\frac{2\omega_{11}}{|\Omega|}L \\ & & \frac{2(\omega_{11}\omega_{22}+\omega_{12}^2)}{|\Omega|^2} & 0 & \frac{2\omega_{12}}{|\Omega|}L & \frac{2\omega_{12}}{|\Omega|}L \\ & & & \frac{2\omega_{12}^2}{|\Omega|} & 0 & 0 \\ & & & & \frac{2(2\omega_{11}\omega_{22}-\omega_{12}^2)}{|\Omega|}M & -\frac{2\omega_{12}^2}{|\Omega|}M \\ & & & & & \frac{2(2\omega_{11}\omega_{22}-\omega_{12}^2)}{|\Omega|}M \end{pmatrix}. \quad (2.3)$$

where L and M are given by

$$L = \frac{1}{m} \sum_{j=1}^m \log \lambda_j, \quad M = \frac{1}{m} \sum_{j=1}^m (\log \lambda_j)^2 \quad (2.4)$$

and $\lambda_j = 2\pi j/N$ are the Fourier frequencies. It is known (see Appendix A) that

$$L \sim \log(m/N), \quad M \sim (\log(m/N))^2. \quad (2.5)$$

Somewhat surprisingly perhaps, one can check that substituting the latter asymptotic expressions for L and M in (2.3) (or substituting M by L^2 in (2.3)) actually yields a singular information

matrix. A way to overcome this is to compute the inverse of (2.3) by treating L and M as distinct. More specifically, by using a symbolic matrix inversion in Mathematica of Wolfram Research (2017), the inverse of (2.3) is

$$\begin{pmatrix} \frac{\omega_{11}(2\omega_{11}\omega_{22}M - \omega_{12}^2L^2)}{2\omega_{22}(M-L^2)} & \frac{\omega_{12}^2(2M-L^2)}{2(M-L^2)} & \frac{\omega_{11}\omega_{12}(2M-L^2)}{2(M-L^2)} & 0 & \frac{(\omega_{11}\omega_{22}+|\Omega|)L}{4\omega_{22}(M-L^2)} & \frac{\omega_{12}^2L}{4\omega_{22}(M-L^2)} \\ & \frac{\omega_{22}(2\omega_{11}\omega_{22}M - \omega_{12}^2L^2)}{2\omega_{11}(M-L^2)} & \frac{\omega_{12}\omega_{22}(2M-L^2)}{2(M-L^2)} & 0 & \frac{\omega_{12}^2L}{4\omega_{11}(M-L^2)} & \frac{(\omega_{11}\omega_{22}+|\Omega|)L}{4\omega_{11}(M-L^2)} \\ & & \frac{\omega_{22}}{2} + \frac{\omega_{12}^2M}{2\omega_{11}(M-L^2)} & 0 & \frac{\omega_{12}L}{4(M-L^2)} & \frac{\omega_{12}L}{4(M-L^2)} \\ & & & \frac{|\Omega|}{2\omega_{12}^2} & 0 & 0 \\ & & & & \frac{\omega_{11}\omega_{22}+|\Omega|}{8\omega_{11}\omega_{22}(M-L^2)} & \frac{\omega_{12}^2}{8\omega_{11}\omega_{22}(M-L^2)} \\ & & & & & \frac{\omega_{11}\omega_{22}+|\Omega|}{8\omega_{11}\omega_{22}(M-L^2)} \end{pmatrix}. \quad (2.6)$$

One can show (Appendix A) that

$$M - L^2 = 1 + o(1). \quad (2.7)$$

The relations (2.5), (2.6) and (2.7) then imply the different convergence rates in (2.1) and the form of the limiting covariance matrix in (2.2). For example, the asymptotic covariance between $\hat{\omega}_{11}$ and $\hat{d}_1$ is calculated from the negative (1,5)th entry of (2.6). It is given by $-(\omega_{11}\omega_{22} + |\Omega|)/4\omega_{22}$ with additional factor $1/\log(N/m)$ due to L , since $M - L^2 \sim 1$ by (2.7).

But we also note that (2.6) is more informative than (2.2). For example, one could check from (2.2) by using the delta method that the asymptotic variance of $\hat{\omega}_{12}^2/(\hat{\omega}_{11}\hat{\omega}_{22})$ is zero at the rate $\sqrt{m}/\log(N/m)$. But using (2.6), the asymptotic variance would be nonzero at the faster rate $\sqrt{m}$.

Remark 2.1 Note that the limiting covariance matrix Γ_p in (2.2) depends only on $\omega_{11}, \omega_{12}, \omega_{22}$. Moreover, the phase parameter estimator $\hat{\phi}$ is asymptotically uncorrelated with respect to other estimators. The asymptotic variance of the (normalized) memory parameter estimators $\hat{d}_1, \hat{d}_2$ is

$$\frac{1}{4} - \frac{\omega_{12}^2}{8\omega_{11}\omega_{22}}. \quad (2.8)$$

When $\omega_{12} = 0$ (or close to 0, since strictly speaking $\omega_{12} = 0$ is not permitted under Assumption (C5)), the quantity (2.8) becomes $1/4$, that is, the usual asymptotic variance of the (normalized) memory estimator in the univariate case. By this analogy, the asymptotic variance of the estimator $\hat{\omega}_{11}$ in the univariate case is ω_{11}^2 , after substituting $\omega_{12} = 0$ in the (1,1) entry of Γ_p . By (2.8), the largest reduction in the variance $1/4$ that can be achieved from another dependent long-memory series is arbitrarily close to $1/8$.

Remark 2.2 The presence of the multiplicative factor $\log(N/m)$ in the estimation of $\omega_{11}, \omega_{22}, \omega_{12}$ is expected in view of the analogous results when using the so-called GPH estimation. For example, a similar factor $\log N$ appears in the GPH estimation of ω_{11} and ω_{12} in Robinson (1995), Theorem 3. In fact, it seems that $\log N$ should actually be $\log(N/m)$ in Robinson (1995) as well. A mistake in Robinson (1995) appears, for example, in relation (5.2) which for $J = 1$ yields:

$$\sum_{k=l+1}^m (-2 \log \lambda_k) = 2m(\log N + 1) + O(l \log N), \quad (2.9)$$

where, in particular, $m/N \rightarrow 0$ and $l/m \rightarrow 0$. Indeed, by Stirling's formula, $\sum_{k=1}^m \log k = n \log n - n + (1/2) \log n + O(1)$, as $n \rightarrow \infty$. Hence,

$$\begin{aligned} \sum_{k=l+1}^m (-2 \log \lambda_k) &= -2 \sum_{k=1}^m \log k + 2 \sum_{k=1}^l \log k + 2(m-l) \log \frac{N}{2\pi} \\ &= 2m \log \frac{N}{2\pi m} + 2m - \frac{1}{2} \log m + O(1) - 2l \log \frac{N}{2\pi l} + 2l - \frac{1}{2} \log l + O(1) \\ &= 2m \left(\log \frac{N}{2\pi m} + 1 \right) + O \left(l \log \frac{N}{l} \right), \end{aligned} \quad (2.10)$$

where we also used the assumption $(\log m)/l \rightarrow 0$ (which follows from Assumption 6 in Robinson (1995)). Note that (2.10) does not imply (2.9): while $O(l \log N/l)$ is obviously also $O(l \log N)$, the same is not expected, for example, for $m \log m$ (arising from the difference of (2.10) and (2.9)), since $(m \log m)/(l \log N) \rightarrow \infty$ for $m = N^a$ and arbitrary power $a > 0$.

Remark 2.3 The asymptotic covariance matrix of $(\hat{\phi}, \hat{d}_1, \hat{d}_2)'$ appearing in Theorem 4 of Robinson (2008) (see also Remark 5) is denoted as Σ^{-1} with $\Sigma = (\sigma_{jk})_{j,k=2,3,4}$ and σ_{jk} specified before the theorem. In particular, it is stated that

$$\sigma_{33} = \sigma_{44} = 4 - 2\mu\rho^2 = 4 - \frac{2\rho}{1 - \rho^2} = \frac{4\omega_{11}\omega_{22} - 6\omega_{12}^2}{\omega_{11}\omega_{22} - \omega_{12}^2}, \quad (2.11)$$

where $\mu = (1 - \rho^2)^{-1}$ and $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2}$. In fact, these entries of Σ should be¹

$$\sigma_{33} = \sigma_{44} = \frac{4\omega_{11}\omega_{22} - 2\omega_{12}^2}{\omega_{11}\omega_{22} - \omega_{12}^2} = 2(\mu + 1) \quad (2.12)$$

and (correcting the entries σ_{33} and σ_{44}) the matrices Σ and Σ^{-1} should read

$$\Sigma^{-1} = \begin{pmatrix} 2(\mu - 1) & 0 & 0 \\ 0 & 2(\mu + 1) & 2(\mu - 1) \\ 0 & 2(\mu - 1) & 2(\mu + 1) \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2(\mu - 1)} & 0 & 0 \\ 0 & \frac{(\mu + 1)}{8\mu} & \frac{(\mu - 1)}{8\mu} \\ 0 & \frac{(\mu - 1)}{8\mu} & \frac{(\mu + 1)}{8\mu} \end{pmatrix}. \quad (2.13)$$

One can check that (2.13) is the same as the matrix consisting of the last three columns and the last three rows of Γ_p in (2.2).

2.2 Case of Parametrization P with fractional cointegration

We now turn to the model (1.10) allowing for fractional cointegration when $\beta \neq 0$. The model parameters $\omega_{11}, \omega_{22}, \omega_{12}, \phi, d_1, d_2$ and their local Whittle estimators below refer to (1.10), and not (1.1) as in the previous section (see also Remark 2.4 below).

Theorem 2.2 *Suppose that the assumptions (C1)–(C5) hold. In addition, suppose that $0 \leq d_1 < d_2 < 1/2$. Then, as $N \rightarrow \infty$,*

$$\sqrt{m} \begin{pmatrix} \lambda_m^{-(d_2 - d_1)} (\hat{\beta} - \beta) \\ \hat{\eta}_p \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Upsilon_p), \quad (2.14)$$

¹As defined in Robinson (2008), the matrix Σ is generally not positively definite. Non-positive definiteness happens when $(4 - 2\rho/(1 - \rho^2))^2 - (2\rho^2/(1 - \rho^2))^2 = 4(3\rho - 2)(\rho + 1)(\rho + 2)(\rho - 1)/(1 - \rho^2)^2 < 0$ or $3\rho - 2 > 0$ or $\rho > 2/3$.

where the vector $\hat{\eta}_p$ is defined in (2.1), and by using the corresponding parameters to subscript the entries of Υ_p , the diagonal entries of Υ_p are given by

$$\begin{aligned}\Upsilon_\beta &= -\frac{\omega_{11}|\Omega|(\nu_0 - 1)^4(2\nu_0 - 1)}{2\omega_{22}\nu_0^2 K_1}, \quad \frac{\Upsilon_{\omega_{11}}}{\omega_{11}^2} = \frac{\Upsilon_{\omega_{22}}}{\omega_{22}^2} = \frac{K_1 + |\Omega|(\nu_0 - 1)^2}{2K_1}, \\ \Upsilon_{\omega_{12}} &= \frac{\omega_{12}^2}{2}, \quad \Upsilon_\phi = \frac{|\Omega|}{2\omega_{12}^2}, \quad \Upsilon_{d_1} = \Upsilon_{d_2} = \frac{K_1 + |\Omega|(\nu_0 - 1)^2}{8K_1},\end{aligned}$$

and the off-diagonal entries of Υ_p are given by

$$\begin{aligned}& \frac{2\nu_0 K_1}{|\Omega|(\nu_0 - 1)^2(2\nu_0 - 1)\cos\phi} (\Upsilon_{\beta\omega_{11}}, \Upsilon_{\beta\omega_{22}}, \Upsilon_{\beta\omega_{12}}, \Upsilon_{\beta\phi}, \Upsilon_{\beta d_1}, \Upsilon_{\beta d_2}) \\ &= \left(\frac{\omega_{11}\omega_{12}}{\omega_{22}}, -\omega_{12}, -\frac{\omega_{11}(\nu_0 - 1)}{\nu_0}, \frac{\omega_{11}(\nu_0 - 1)}{\omega_{12}}, \frac{\omega_{12}}{2\omega_{22}}, -\frac{\omega_{12}}{2\omega_{22}} \right), \\ & \frac{4K_1}{|\Omega|(\nu_0 - 1)(2\nu_0 - 1)\sin^2\phi} (\Upsilon_{\phi\omega_{11}}, \Upsilon_{\phi\omega_{22}}, \Upsilon_{\phi\omega_{12}}, \Upsilon_{\phi d_1}, \Upsilon_{\phi d_2}) \\ &= \left(-\omega_{11}, \omega_{22}, \frac{\omega_{11}\omega_{22}(\nu_0 - 1)}{\omega_{12}\nu_0^2}, -\frac{1}{2\nu_0}, \frac{1}{2\nu_0} \right), \\ & (\Upsilon_{\omega_{12}\omega_{11}}, \Upsilon_{\omega_{12}\omega_{22}}, \Upsilon_{\omega_{12}d_1}, \Upsilon_{\omega_{12}d_2}) = \left(\frac{\omega_{11}\omega_{12}}{2}, \frac{\omega_{12}\omega_{22}}{2}, \frac{\omega_{12}}{4}, \frac{\omega_{12}}{4} \right), \\ & (\Upsilon_{\omega_{11}\omega_{22}}, \Upsilon_{\omega_{11}d_1}, \Upsilon_{\omega_{11}d_2}) = \left(\frac{\omega_{11}\omega_{22}K_2}{2K_1}, \frac{\omega_{11}(K_1 + |\Omega|(\nu_0 - 1)^2)}{4K_1}, \frac{\omega_{11}K_2}{4K_1 - 1} \right), \\ & (\Upsilon_{\omega_{22}d_1}, \Upsilon_{\omega_{22}d_2}) = \left(\frac{\omega_{22}K_2}{4K_1}, \frac{\omega_{22}(K_1 + |\Omega|(\nu_0 - 1)^2)}{4K_1} \right), \quad \Upsilon_{d_1 d_2} = \frac{K_2}{4K_1}\end{aligned}$$

with $\nu_0 = d_2 - d_1$ and

$$K_1 = \omega_{11}\omega_{22}(\nu_0 - 1)^2 + (2\nu_0 - 1)\omega_{12}^2 \sin^2\phi, \quad K_2 = \omega_{11}^2\nu_0^2 - (2\nu_0 - 1)\omega_{11}^2 \sin^2\phi.$$

Remark 2.4 Though the same parameters $\omega_{11}, \omega_{22}, \omega_{12}, \phi, d_1, d_2$ are used in the case of fractional cointegration, we stress again that these parameter appear in the model (1.10) with transformed spectral density $f_X(\lambda)$. When $\beta \neq 0$ (and $d_1 < d_2$), the spectral density $f_X(\lambda)$ itself satisfies the asymptotic relation

$$f_X(\lambda) \sim \omega_{22}\lambda^{-2d_2} \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}, \quad \text{as } \lambda \rightarrow 0^+. \quad (2.15)$$

Thus, for the original series X , the long-memory and other model parameters satisfy (1.12). We also note for later reference that the phase parameter for the model (2.15) is always $\phi = 0$.

Remark 2.5 As in the discussion following Theorem 2.1, we note that the information matrix for the fractionally cointegrated model is computed in Appendix A and has the same entries as the information matrix for the non-cointegrated model for all the parameters except β (see (A.9) for the entries related to β). The inverse of the information matrix in the form (2.6) could also be computed but it will not be included here for the shortness sake.

2.3 Case of Parametrization C without fractional cointegration

The next result presents the asymptotic normality of the parameter estimators in Parametrization C in (1.4). Note that the different parameters (and their estimators) in Parametrizations P and C are related through $r_1 - ir_2 = \omega_{12}e^{-i\phi}$ and, for $r_1 \neq 0$,

$$\phi = \arctan\left(\frac{r_2}{r_1}\right), \quad \omega_{12} = \text{sign}(r_1)\sqrt{r_1^2 + r_2^2}. \quad (2.16)$$

In principle, the asymptotic normality result in Parametrization C follows from Theorem 2.1 through the delta method, at least when $r_1 \neq 0$. But it will be easier to derive the result directly, and in contrast to Theorem 2.1, allowing $r_1 = 0$ or/and $r_2 = 0$. We will also state the result using the asymptotic covariance matrix in the form (2.6) given in the discussion following Theorem 2.1 for Parametrization P that will prove useful in connection to fractal connectivity. As already noted at the end of that discussion, the asymptotic covariance form (2.6) carries more information than the form (2.2), and this will also be the case with the form (2.19) given below. To use the latter form, we shall need the notation

$$\zeta_N \stackrel{d}{\sim} \mathcal{N}(0, \Psi_N), \quad \text{as } N \rightarrow \infty, \quad (2.17)$$

with a random vector ζ_N and a covariance matrix Ψ_N , to mean that $\text{diag}(\Psi_N)^{-1/2}\zeta_N \stackrel{d}{\rightarrow} \mathcal{N}(0, \Psi)$ for some covariance matrix Ψ . The relation (2.17) thus expresses the idea that ζ_N is asymptotically normal with mean 0 and covariance Ψ_N . One thinks here of Ψ_N having entries that diverge to infinity. The appearance of $\text{diag}(\Psi_N)^{1/2}$ in the above interpretation should then not be surprising since this normalization is also expected to be that for the cross covariances in Ψ_N .

We need to modify some of the assumptions stated in Section 2 as well:

- (C1)' The spectral density $f_X(\lambda)$ of a bivariate stationary series $\{X_n\}$ satisfies (1.1). Furthermore, for some $b \in (0, 2]$, $\lambda^D C(\lambda) - Q = O(\lambda^b)$, as $\lambda \rightarrow 0^+$, where Q is a complex-valued 2×2 matrix satisfying $Q\overline{Q}' = G$, and $C(\lambda)$ is a 2×2 matrix-valued function differentiable in a neighborhood of $\lambda = 0$ such that $f_X(\lambda) = C(\lambda)\overline{C}(\lambda)'$ and $\Phi_{D,\phi}(\lambda)dC(\lambda)/d\lambda = O(\lambda^{-1})$, as $\lambda \rightarrow 0^+$.
- (C3)' Let $\theta = (\omega_{11}, \omega_{22}, r_1, r_2, d_1, d_2)'$ be the parameter vector of interest and such that $\theta \in \Theta_{\omega_r} \times \Theta_D$, where $\Theta_{\omega_r} = \{(x, y, z, w) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R} : xy - (z^2 + w^2) > 0\}$, $\Theta_D = (-1/2, 1/2)^2$.

Theorem 2.3 Suppose that the assumptions (C1)', (C2), (C3)' and (C4) hold. Then, as $N \rightarrow \infty$,

$$\sqrt{m}\widehat{\eta}_c := \sqrt{m} \begin{pmatrix} \widehat{\omega}_{11} - \omega_{11} \\ \widehat{\omega}_{22} - \omega_{22} \\ \widehat{r}_1 - r_1 \\ \widehat{r}_2 - r_2 \\ \widehat{d}_1 - d_1 \\ \widehat{d}_2 - d_2 \end{pmatrix} \stackrel{d}{\sim} \mathcal{N}(0, \Psi_{c,N}), \quad (2.18)$$

where $\stackrel{d}{\sim}$ is defined following (2.17) and $\Psi_{c,N}$ is

$$\begin{pmatrix} \frac{\omega_{11}(2\omega_{11}\omega_{22}M - (r_1^2 + r_2^2)L^2)}{2\omega_{22}(M-L^2)} & \frac{(r_1^2 + r_2^2)(2M-L^2)}{2(M-L^2)} & \frac{\omega_{11}r_1(2M-L^2)}{2(M-L^2)} & \frac{\omega_{11}r_2(2M-L^2)}{2(M-L^2)} & \frac{(\omega_{11}\omega_{22} + |G|)L}{4\omega_{22}(M-L^2)} & \frac{(r_1^2 + r_2^2)L}{4\omega_{22}(M-L^2)} \\ \frac{\omega_{22}(2\omega_{11}\omega_{22}M - (r_1^2 + r_2^2)L^2)}{2\omega_{11}(M-L^2)} & \frac{\omega_{22}r_1(2M-L^2)}{2(M-L^2)} & \frac{\omega_{22}r_2(2M-L^2)}{2(M-L^2)} & \frac{(r_1^2 + r_2^2)L}{4\omega_{11}(M-L^2)} & \frac{(\omega_{11}\omega_{22} + |G|)L}{4\omega_{11}(M-L^2)} & \frac{(\omega_{11}\omega_{22} + |G|)L}{4\omega_{11}(M-L^2)} \\ \frac{|G|}{2} + \frac{r_1^2(2M-L^2)}{2(M-L^2)} & \frac{r_1r_2(2M-L^2)}{2(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} \\ \frac{|G|}{2} + \frac{r_2^2(2M-L^2)}{2(M-L^2)} & \frac{r_1r_2(2M-L^2)}{2(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} \\ \frac{\omega_{11}\omega_{22} + |G|}{8\omega_{11}\omega_{22}(M-L^2)} & \frac{r_1^2 + r_2^2}{8\omega_{11}\omega_{22}(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} \\ \frac{\omega_{11}\omega_{22} + |G|}{8\omega_{11}\omega_{22}(M-L^2)} & \frac{r_1^2 + r_2^2}{8\omega_{11}\omega_{22}(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} & \frac{r_1L}{4(M-L^2)} & \frac{r_2L}{4(M-L^2)} \end{pmatrix} \quad (2.19)$$

with L and M defined in (2.4), and satisfying (2.5) and (2.7).

Focus now on the estimators $\hat{r}_1$ and $\hat{r}_2$ that are of particular interest in connection to fractal (non-)connectivity (cf. (1.6)). Note that by (2.18) and (2.19),

$$\sqrt{m} \begin{pmatrix} \hat{r}_1 - r_1 \\ \hat{r}_2 - r_2 \end{pmatrix} \stackrel{d}{\sim} \mathcal{N}(0, \Psi_{r,N}) \quad \text{with} \quad \Psi_{r,N} = \begin{pmatrix} \frac{|G|}{2} + \frac{r_1^2(2M-L^2)}{2(M-L^2)} & \frac{r_1r_2(2M-L^2)}{2(M-L^2)} \\ \frac{r_1r_2(2M-L^2)}{2(M-L^2)} & \frac{|G|}{2} + \frac{r_2^2(2M-L^2)}{2(M-L^2)} \end{pmatrix}. \quad (2.20)$$

That is, by using the relations (2.5) and (2.7) for M, L ,

$$\sqrt{m} \begin{pmatrix} \frac{\hat{r}_1 - r_1}{\frac{1_{\{r_1=0\}} + (\log N/m)1_{\{r_1 \neq 0\}}}{\hat{r}_2 - r_2}} \\ \frac{\hat{r}_2 - r_2}{\frac{1_{\{r_2=0\}} + (\log N/m)1_{\{r_2 \neq 0\}}}{\hat{r}_1 - r_1}} \end{pmatrix} \stackrel{d}{\sim} \mathcal{N}(0, \Psi_{r_1, r_2}) \quad (2.21)$$

with

$$\Psi_{r_1, r_2} = \begin{pmatrix} \frac{|G|}{2}1_{\{r_1=0\}} + \frac{r_1^2}{2}1_{\{r_1 \neq 0\}} & \frac{r_1r_2}{2} \\ \frac{r_1r_2}{2} & \frac{|G|}{2}1_{\{r_2=0\}} + \frac{r_2^2}{2}1_{\{r_2 \neq 0\}} \end{pmatrix}.$$

Indeed, for example, when $r_1 \neq 0$ and $r_2 \neq 0$, note from (2.5) and (2.7) that $\Psi_{r,N}$ behaves as

$$\Psi_{r,N} \sim \begin{pmatrix} \frac{r_1^2}{2}(\log N/m)^2 & \frac{r_1r_2}{2}(\log N/m)^2 \\ \frac{r_1r_2}{2}(\log N/m)^2 & \frac{r_2^2}{2}(\log N/m)^2 \end{pmatrix},$$

which yields (2.21). Note that Ψ_{r_1, r_2} is singular when $r_1 \neq 0, r_2 \neq 0$. This is expected in view of the asymptotic normality result for the estimators $\hat{\omega}_{12}, \hat{\phi}$ in Parametrization P, where the convergence rate for $\hat{\omega}_{12}$ is $\sqrt{m}/\log(N/m)$ but that for $\hat{\phi}$ is the faster rate $\sqrt{m}$.

The asymptotic relation (2.21) provides a means for testing the fractal non-connectivity hypothesis

$$H_0 : r_1 = r_2 = 0 \quad \text{against} \quad H_1 : \text{not } H_0, \quad (2.22)$$

(cf. (1.6)). Setting a test statistic as

$$\hat{\xi}_N = \frac{2m}{|\hat{G}|}(\hat{r}_1^2 + \hat{r}_2^2) = \frac{2m}{\hat{\omega}_{11}\hat{\omega}_{22}}(\hat{r}_1^2 + \hat{r}_2^2), \quad (2.23)$$

we have by (2.21) that

$$\hat{\xi}_N \stackrel{d}{\rightarrow} \chi^2(2) \quad (2.24)$$

under H_0 , where $\chi^2(k)$ denotes the chi-square distribution with k degrees of freedom, and also $\hat{\xi}_N \xrightarrow{P} +\infty$ under the alternative H_1 . Testing for H_0 can then be carried out in a standard way by comparing $\hat{\xi}_N$ to an appropriate critical value associated with the $\chi^2(2)$ distribution. Some related local Whittle plots are discussed in Section 5.

Remark 2.6 We note that under the hypothesis H_0 in (2.22),

$$\hat{\phi} = \arctan\left(\frac{\hat{r}_2}{\hat{r}_1}\right) = \arctan\left(\frac{\sqrt{m}\hat{r}_2}{\sqrt{m}\hat{r}_1}\right) \xrightarrow{d} \arctan\left(\frac{X_2}{X_1}\right) =: U_0 \stackrel{d}{=} U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \quad (2.25)$$

where X_1, X_2 are independent $\mathcal{N}(0, 1)$ random variables and $U(-\pi/2, \pi/2)$ refers to a uniform distribution on $(-\pi/2, \pi/2)$. The relations (2.23) and (2.25) clarify what happens with the local Whittle estimators $\hat{\omega}_{12}^2 = \hat{r}_1^2 + \hat{r}_2^2$ and $\hat{\phi}$ in the case $r_1 = r_2 = 0$.

Remark 2.7 We also note here that it seems to be impossible to derive any inference procedures that would accommodate both the hypothesis H_0 in (2.22) and its alternative H_1 . More specifically, note from Theorem 2.1 that a confidence interval for $\omega_{12}^2 = r_1^2 + r_2^2$ (or ω_{12} could also be considered similarly) is

$$\hat{\omega}_{12}^2 \pm z_{1-\alpha} \frac{\sqrt{2} \log(N/m) \hat{\omega}_{12}^2}{\sqrt{m}}, \quad (2.26)$$

where $z_{1-\alpha}$ is an appropriate critical value for the standard normal distribution. Since $z_{1-\alpha} \sqrt{2} \log(N/m) / \sqrt{m}$, is expected to be smaller than 1 for most m 's of interest, the confidence interval (2.26) will *never* contain $\omega_{12}^2 = r_1^2 + r_2^2 = 0$. That is, one could not expect a confidence interval for ω_{12}^2 assuming $\omega_{12} \neq 0$ to be valid for $r_1^2 + r_2^2$ assuming $r_1 = r_2 = 0$.

2.4 Case of Parametrization C and fractional cointegration

In the following result, we also provide the asymptotic normality result in the case of fractional cointegration but assuming that the right-hand side of (1.10) is parametrized by (1.4).

Theorem 2.4 *Suppose that the assumptions $(C1)'$, $(C2)$ and $(C3)'$ hold. In addition, suppose that $0 \leq d_1 < d_2 < 1/2$. Then, as $N \rightarrow \infty$,*

$$\sqrt{m} \begin{pmatrix} \lambda_m^{-(d_2-d_1)} (\hat{\beta} - \beta) \\ \hat{\eta}_c \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Upsilon_c), \quad (2.27)$$

where the vector $\hat{\eta}_c$ is defined in (2.18) and Υ_c coincides with Υ_p in Theorem 2.2 after the transformations in (2.16), except the following entries:

$$\begin{aligned} \Upsilon_{\beta r_1} &= -\frac{\omega_{11}|G|(\nu_0 - 1)^4(2\nu_0 - 1)}{2\omega_{22}\nu_0^2 K_1}, & \Upsilon_{\beta r_2} &= 0, \\ \Upsilon_{\omega_{11}r_1} &= -\frac{\omega_{11}r_1}{2}, & \Upsilon_{\omega_{11}r_2} &= -\frac{\omega_{11}r_2}{2}, & \Upsilon_{\omega_{22}r_1} &= \frac{\omega_{22}r_1}{2}, & \Upsilon_{\omega_{22}r_2} &= \frac{\omega_{22}r_2}{2}, \\ \Upsilon_{r_1} &= \frac{r_1^2}{2}, & \Upsilon_{r_1r_2} &= \frac{r_1r_2}{2}, & \Upsilon_{r_1,d_1} &= \Upsilon_{r_1,d_2} = \frac{r_1}{4}, & \Upsilon_{r_2} &= \frac{r_2^2}{2}, & \Upsilon_{r_2,d_1} &= \Upsilon_{r_2,d_2} = \frac{r_2}{4}. \end{aligned}$$

As stated, Theorem 2.4 is not suitable for making inference about the fractal non-connectivity $r_1 = r_2 = 0$. Indeed, note that the variances of $\hat{r}_1$ and $\hat{r}_2$ are zero in the theorem under this assumption. To accommodate the case $r_1 = r_2 = 0$, we would need to restate Theorem 2.4 in the form of Theorem 2.3. We shall not do so for the shortness sake. But we shall indicate the corresponding result for $\hat{r}_1$ and $\hat{r}_2$. More specifically, we have

$$\sqrt{m} \begin{pmatrix} \frac{\hat{r}_1 - r_1}{\mathbb{1}_{\{r_1=0\}} + (\log N/m)\mathbb{1}_{\{r_1 \neq 0\}}} \\ \frac{\hat{r}_2 - r_2}{\mathbb{1}_{\{r_2=0\}} + (\log N/m)\mathbb{1}_{\{r_2 \neq 0\}}} \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \tilde{\Psi}_{r_1, r_2}) \quad (2.28)$$

with

$$\tilde{\Psi}_{r_1, r_2} = \begin{pmatrix} \frac{|G|(\nu_0-1)^2}{2\nu_0^2} 1_{\{r_1=0\}} + \frac{r_1^2}{2} 1_{\{r_1 \neq 0\}} & \frac{r_1 r_2}{2} \\ \frac{r_1 r_2}{2} & \frac{|G|}{2} 1_{\{r_2=0\}} + \frac{r_2^2}{2} 1_{\{r_2 \neq 0\}} \end{pmatrix}.$$

Then, we can perform a test for fractal non-connectivity based on

$$\hat{\xi}_{fc, N} = \frac{2m}{|\hat{G}|} \left(\frac{\hat{\nu}_0^2}{(\hat{\nu}_0 - 1)^2} \hat{r}_1^2 + \hat{r}_2^2 \right) \xrightarrow{d} \chi^2(2) \quad (2.29)$$

under H_0 .

3 Reduction of optimization problems

We discuss here several reductions of the optimization problems for the local Whittle estimation. It is a standard exercise to show that the optimization problem (1.7) can be reduced to that over ϕ, D only as

$$(\hat{\phi}, \hat{D}) = \underset{(\phi, D)}{\operatorname{argmin}} R(\phi, D) \quad (3.1)$$

with

$$R(\phi, D) = \log |\hat{\Omega}(D, \phi)| - 2(d_1 + d_2) \frac{1}{m} \sum_{j=1}^m \log \lambda_j, \quad (3.2)$$

where

$$\hat{\Omega}(\phi, D) = \Re \left(\frac{1}{m} \sum_{j=1}^m \Phi_{D, \phi}(\lambda_j) I_X(\lambda_j) \bar{\Phi}_{D, \phi}(\lambda_j) \right) \quad (3.3)$$

with $\Re$ denoting the real part (entry-wise). Moreover, $\hat{\Omega}$ in (1.7) is equal to $\hat{\Omega}(\hat{\phi}, \hat{D})$. For example, this reduction was used by Robinson (2008), and seems to be one reason why the asymptotics only for $\hat{\phi}$ and $\hat{D}$ were established.

It is perhaps less known but the optimization (3.1) can be reduced even further. Indeed, we show in Appendix C that the optimization (1.8) can be reduced to that over D only and that $\hat{\phi}$ can be expressed in terms of the estimated $\hat{D}$. More specifically, we have

$$\hat{D} = \underset{D}{\operatorname{argmin}} R(\hat{\phi}^e(D), D), \quad (3.4)$$

where

$$\hat{\phi}^e(D) = -\arctan \left(\frac{\sum_{j=1}^m \lambda_j^{d_1+d_2} \Im(I_{X,12}(\lambda_j))}{\sum_{j=1}^m \lambda_j^{d_1+d_2} \Re(I_{X,12}(\lambda_j))} \right) \quad (3.5)$$

and $I_{X,12}(\lambda_j)$ is the (1,2) entry of the periodogram $I_X(\lambda_j)$ in (1.9). Moreover $\hat{\phi} = \hat{\phi}^e(\hat{D})$. It is interesting to note that a preliminary estimator of ϕ given in Remark 3 of Robinson (2008), namely,

$$\hat{\phi}^{init} = -\arctan \left(\frac{\sum_{j=1}^m \Im(I_{X,12}(\lambda_j))}{\sum_{j=1}^m \Re(I_{X,12}(\lambda_j))} \right), \quad (3.6)$$

is the local Whittle estimator (3.5) with $d_1 = d_2 = 0$.

Turning to the case of fractional cointegration, similar calculations give the explicit form of the phase parameter estimator as

$$\hat{\phi}^e(D, \beta) = -\arctan \left(\frac{\sum_{j=1}^m \lambda_j^{d_1+d_2} \{\Im(I_{X,12}(\lambda_j)) - \beta I_{X,22}(\lambda_j)\}}{\sum_{j=1}^m \lambda_j^{d_1+d_2} \{\Re(I_{X,12}(\lambda_j)) - \beta I_{X,22}(\lambda_j)\}} \right), \quad (3.7)$$

in terms of the other model parameters. We omit details for the shortness sake.

Finally, there is another natural way to understand the reduced optimization problem (3.4) with the expressions (3.2) and (3.5) for other model parameters. Note that the reduced optimization problem (3.1) was obtained from (1.7) by eliminating the real-valued matrix Ω ; in particular, the fact that (3.3) has real entries is ensured by taking the real part $\Re$. One could expect that a similar reduction could also take place in eliminating not the real-valued matrix Ω but a complex-valued matrix G in (1.1), comprised of both Ω and ϕ . Indeed, we show in Appendix C that the optimization problem (3.4)–(3.5) is equivalent to

$$\hat{D} = \operatorname{argmin}_D \left\{ \log |\hat{G}(D)| - 2(d_1 + d_2) \frac{1}{m} \sum_{j=1}^m \log \lambda_j \right\}, \quad (3.8)$$

where

$$\hat{G}(D) = \frac{1}{m} \sum_{j=1}^m \lambda_j^D I_X(\lambda_j) \lambda_j^D. \quad (3.9)$$

Moreover, letting $\hat{G} = (\hat{g}_{jk})$ be defined from $\hat{\phi}, \hat{\Omega}$ in (1.7) as in the relations preceding (1.2) (that is, $\hat{g}_{11} = \hat{\omega}_{11}$, $\hat{g}_{22} = \hat{\omega}_{22}$, etc.), we have $\hat{G} = \hat{G}(\hat{D})$. That is, in deriving the local Whittle estimators, one may as well work with the specification (1.1), involving the complex-valued matrix G . These observations also suggest that the finding (3.5) should not be surprising, since $\hat{\phi}$ can be expressed in terms of $\hat{G}(\hat{D})$. In the case of fractional cointegration, the same relationship (3.8) holds for optimizing over D and B with $\hat{G}(D) = \frac{1}{m} \sum_{j=1}^m \lambda_j^D B I_X(\lambda_j) B' \lambda_j^D$.

4 Simulation study

In this section, we present a small simulation study examining our asymptotic results. We focus on the situations that highlight our contributions: estimation of the additional parameters $\omega_{11}, \omega_{22}, \omega_{12}$ and dealing with fractal (non-)connectivity. But first we define the processes used in the simulations.

4.1 Data generating processes

We consider the following data generating processes (DGPs) in the simulations discussed in Section 4.2 below and also in Section 5.

- (DGP1) Two-sided VARFIMA(0, D , 0) model: This model, introduced in Kechagias and Pipiras (2017), is defined as

$$X_n = \Delta_{c,D}(B)^{-1} \eta_n, \quad n \in \mathbb{Z}, \quad (4.1)$$

where $\{\eta_n\}_{n \in \mathbb{Z}}$ is a Gaussian white noise series with $\mathbb{E} \eta_n \eta_n' = \Sigma_\eta$, B is the backward shift operator and

$$\Delta_{c,D}(B) = (1 - B)^{-D} + (I - B^{-1})^{-D} \begin{pmatrix} c & 0 \\ 0 & -c \end{pmatrix} \quad (4.2)$$

with $c \in (-1, 1)$ and $D = \text{diag}(d_1, d_2)$. As shown in Kechagias and Pipiras (2017), there is a one-to-one correspondence between $c \in (-1, 1)$ and the phase parameter $\phi_c \in (-\pi/2, \pi/2)$, given by the relationship

$$\phi_c = -\arctan\left(\frac{a_1 \frac{c-1}{c+1} + a_2 \frac{1+c}{1-c}}{1 + a_1 a_2}\right), \quad a_j = \tan\left(\frac{\pi d_j}{2}\right).$$

The autocovariance function of the series is known in closed form, and the series can be generated exactly as in e.g. Helgason, Pipiras and Abry (2011). When $c = 0$, the two-sided VARFIMA(0, D , 0) model becomes the one-sided VARFIMA(0, D , 0) model commonly used in the literature on multivariate long memory.

- (DGP2) Fractally non-connected model: This is a model with the spectral density

$$f_X(\lambda) = \begin{pmatrix} \frac{\sigma_{11}^2}{2\pi} |1 - e^{-i\lambda}|^{-2d_1} & \frac{\sigma_{12}}{2\pi} (1 - e^{-i\lambda})^{-\delta_1} (1 - e^{i\lambda})^{-\delta_2} \\ \frac{\sigma_{12}}{2\pi} (1 - e^{-i\lambda})^{-\delta_2} (1 - e^{i\lambda})^{-\delta_1} & \frac{\sigma_{22}^2}{2\pi} |1 - e^{-i\lambda}|^{-2d_2} \end{pmatrix}, \quad (4.3)$$

where $0 < \delta_j < d_j < 1/2$, $j = 1, 2$. It is fractally non-connected since $\delta_j < d_j$. Note also that, since the entries of the matrix (4.3) are those of the spectral density of a one-sided VARFIMA(0, $\tilde{D}$, 0) series (with $\tilde{d} = d$ or δ), the autocovariance function of the series can be computed explicitly and the Gaussian series can be generated exactly following Helgason et al. (2011).

- (DGP3) Fractionally cointegrated two-sided VARFIMA(0, D , 0) model: The spectral density $f_X(\lambda)$ of this model is such that $Bf_X(\lambda)B'$ coincides with the spectral density of the two-sided VARFIMA(0, D , 0) model, where B is defined in (1.11) and involves a cointegration parameter β . Note that this is equivalent to the spectral density of the series BX being that of the two-sided VARFIMA(0, D , 0) series Y , that is, $BX = Y$ or $X = B^{-1}Y$ with $B^{-1} = (1 \ \beta; 0 \ 1)$. For the underlying two-sided VARFIMA(0, D , 0) model, we take DGP1.
- (DGP4) Fractionally cointegrated fractally non-connected model: This model is defined in the same way as DGP3 but based on the fractally non-connected model DGP2 rather than the two-sided VARFIMA(0, D , 0) model DGP1.

Unless specified otherwise, the parameters used for each DGP are the following. DGP1 uses

$$d_1 = 0.2, \quad d_2 = 0.4, \quad \phi_c = 1.158 \ (c = 0.6), \quad \Omega = \begin{pmatrix} 0.371 & 0.097 \\ 0.097 & 0.157 \end{pmatrix}, \quad \Sigma_\eta = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix},$$

and DGP3 is generated from DGP1 by taking the cointegration parameter $\beta = 1$. For the fractally non-connected model DGP2, we take

$$d_1 = 0.3, \quad d_2 = 0.4, \quad \delta_1 = 0.1, \quad \delta_2 = 0.1, \quad \sigma_{11} = \sigma_{22} = 1, \quad \sigma_{12} = 0.2.$$

DGP4 is defined by fractionally cointegrating DGP2 with the cointegration parameter $\beta = 1$.

4.2 Simulation results

We use the following measures to evaluate the performance of the local Whittle estimators: the bias, the mean-squared error (MSE), the difference between the empirical variance and

Parameters	Measures $\times 10^2$	Number of frequencies m						
		$N^{.5}$	$N^{.55}$	$N^{.6}$	$N^{.65}$	$N^{.7}$	$N^{.75}$	$N^{.8}$
ω_{11}	Bias	5.46	2.87	2.06	1.62	1.75	2.27	2.54
	MSE	5.41	2.56	1.38	0.64	0.33	0.20	0.14
	Var Diff 1	0.83	0.18	0.21	0.04	0.00	0.00	-0.01
	Var Diff 2	0.11	-0.43	-0.24	-0.24	-0.17	-0.09	-0.04
ω_{12}	Bias	-2.39	-1.91	-1.76	-1.70	-1.56	-1.78	-2.04
	MSE	0.82	0.45	0.29	0.18	0.10	0.09	0.06
	Var Diff 1	0.53	0.29	0.19	0.11	0.05	0.04	0.01
	Var Diff 2	0.59	0.33	0.21	0.13	0.06	0.05	0.01
ω_{22}	Bias	0.33	-1.50	-2.66	-3.55	-4.37	-5.32	-6.42
	MSE	1.13	0.40	0.21	0.19	0.22	0.29	0.42
	Var Diff 1	0.60	0.12	0.02	0.01	0.00	0.00	0.00
	Var Diff 2	0.51	0.05	-0.03	-0.02	-0.01	-0.01	0.00
ϕ	Bias	-1.67	-1.69	-2.16	-3.65	-5.17	-7.37	-11.32
	MSE	11.56	7.34	4.48	3.17	2.42	2.11	2.38
	Var Diff 1	3.79	1.65	0.33	0.06	0.07	0.09	0.05
	Var Diff 2	3.79	1.65	0.33	0.06	0.07	0.09	0.05
d_1	Bias	-0.38	-0.36	-0.50	-0.59	-0.80	-1.19	-1.45
	MSE	1.03	0.68	0.47	0.33	0.23	0.17	0.13
	Var Diff 1	-0.02	-0.01	0.01	0.02	0.01	0.01	0.00
	Var Diff 2	0.29	0.16	0.10	0.07	0.04	0.03	0.01
d_2	Bias	1.98	2.42	3.22	4.38	5.80	8.08	11.78
	MSE	1.30	0.83	0.58	0.53	0.56	0.80	1.49
	Var Diff 1	0.21	0.08	0.02	0.03	0.01	0.00	0.01
	Var Diff 2	0.52	0.25	0.11	0.08	0.04	0.02	0.01

Table 1: Measures of interest $\times 10^2$ for DGP1 and its parameters. Var Diff 1 is the difference between the empirical variance and the variance determined by the information matrix (2.3), Var Diff 2 is the difference between the empirical variance and the asymptotic variance in Theorem 2.1.

the variance determined by the information matrix (2.3) (Var Diff 1), and the difference between the empirical variance and the asymptotic variance determined by Γ_p in Theorem 2.1 (Var Diff 2). All results are based on 500 replications. The sample size is $N = 1,000$, and $m = N^{.5}, N^{.55}, N^{.6}, N^{.65}, N^{.7}, N^{.75}, N^{.8}$ are used in the local Whittle estimation.

Table 1 shows the performance measures for DGP1. Note that all parameters are inferred reasonably well in terms of the bias and the variance differences. Note, however, that the bias and MSE for the phase parameter are calculated by considering a modification of the phase parameter estimator discussed in Section 5.2 below.

For the fractionally cointegrated model DGP2, the results are reported in Table 2. They are worse than those for the non-cointegrated model DGP1. First, the cointegration parameter β estimates show large bias when the number of frequencies m is small, and larger m is needed to get the smallest MSE. The asymptotic variance is calculated based on the information matrix (2.14) which depends on the phase parameter. We used the modification of Section 5.2 to estimate the phase parameter correctly. Second, Table 2 shows much larger MSEs when $m = N^{.5}$. This is due to very unstable estimates of the cointegration parameter β , affecting other parameters as well. However, as m increases, MSEs tend to decrease. This indicates that cointegration makes estimation worse, and requires a larger number of frequencies for better performance, and hence also a larger sample.

We also report in Table 3 the size and power of the tests for fractal non-connectivity proposed in Sections 2.3 and 2.4. For the size of the test based on ξ_N in (2.24), DGP2 is considered. It is observed that the size depends on the number of frequencies. Either $m = N^{.5}$ or $N^{.55}$ seems

Parameters	Measures $\times 10^2$	Number of frequencies m						
		$N^{.5}$	$N^{.55}$	$N^{.6}$	$N^{.65}$	$N^{.7}$	$N^{.75}$	$N^{.8}$
β	Bias	-1.32	-3.34	-6.29	-6.79	-7.21	-7.70	-9.08
	MSE	324.02	219.78	97.96	19.50	12.34	7.39	5.06
	Var Diff 1	317.88	212.51	90.62	13.81	8.06	4.27	2.82
	Var Diff 2	314.05	210.32	90.59	13.71	7.88	4.20	2.82
ω_{11}	Bias	116.56	61.59	29.75	7.95	5.41	4.38	4.29
	MSE	2250.05	829.11	302.76	2.56	1.32	0.52	0.33
	Var Diff 1	2105.02	786.90	292.03	1.04	0.63	0.13	0.05
	Var Diff 2	2103.01	785.96	291.56	0.80	0.47	0.06	0.03
ω_{12}	Bias	-4.55	-4.34	-3.06	-2.86	-2.11	-1.64	-1.49
	MSE	35.12	13.74	6.23	1.17	0.64	0.29	0.11
	Var Diff 1	34.30	13.20	5.95	1.00	0.54	0.23	0.08
	Var Diff 2	34.52	13.37	6.06	1.06	0.58	0.25	0.09
ω_{22}	Bias	0.27	-1.16	-2.51	-3.42	-4.28	-5.27	-6.35
	MSE	0.97	0.41	0.21	0.18	0.21	0.29	0.41
	Var Diff 1	0.48	0.13	0.01	0.00	0.00	0.00	0.00
	Var Diff 2	0.26	0.07	0.00	0.00	0.00	0.00	0.00
ϕ	Bias	-106.43	-104.04	-86.90	-72.88	-53.59	-37.44	-27.63
	MSE	181.42	185.81	161.43	140.05	102.78	60.80	27.46
	Var Diff 1	51.48	60.53	71.95	74.41	65.18	41.36	17.12
	Var Diff 2	64.54	74.45	83.15	84.61	72.29	45.51	18.94
d_1	Bias	-2.24	-1.77	-1.43	-1.33	-1.17	-1.32	-1.57
	MSE	1.16	0.77	0.54	0.38	0.25	0.18	0.13
	Var Diff 1	0.02	0.04	0.05	0.05	0.02	0.02	0.01
	Var Diff 2	0.39	0.22	0.15	0.10	0.05	0.03	0.02
d_2	Bias	1.53	2.05	2.97	4.15	5.59	7.90	11.47
	MSE	1.12	0.86	0.58	0.53	0.55	0.79	1.43
	Var Diff 1	0.01	0.11	0.03	0.04	0.03	0.02	0.01
	Var Diff 2	0.38	0.29	0.12	0.09	0.05	0.03	0.02

Table 2: Measures of interest $\times 10^2$ for DGP3 with $\beta = 1$ and its parameters.

	Test statistic	Number of frequencies						
		$N^{.5}$	$N^{.55}$	$N^{.6}$	$N^{.65}$	$N^{.7}$	$N^{.75}$	$N^{.8}$
size	$\hat{\xi}_N$	0.05	0.07	0.114	0.188	0.382	0.67	0.936
	$\hat{\xi}_{fc,N}$	0.04	0.022	0.028	0.024	0.028	0.042	0.068
power	$\hat{\xi}_N$	0.824	0.942	0.986	1	1	1	1
	$\hat{\xi}_{fc,N}$	0.954	0.958	0.978	0.994	1	1	1

Table 3: Size and power of the tests for fractal non-connectivity based on $\hat{\xi}_N$ and $\hat{\xi}_{fc,N}$.

to achieve a nominal significance level while the test is seriously oversized when larger frequencies are used. On the other hand, the power of the test against fractal connectivity is examined by using DGP1. The power is reasonably good for all frequencies considered. We also considered the test for fractal non-connectivity when the series is fractionally cointegrated, based on the test statistic $\xi_{fc,N}$ in (2.29). The model used in the simulation is DGP4 with $\beta = 1$. Compared to the non-cointegrated case, note that a nearly nominal size level is obtained regardless of any frequency used. The power for the cointegrated case is considered based on DGP3 with $\beta = 1$. It also shows an excellent power for all frequencies.

5 Local Whittle plots

Local Whittle plots are plots of parameter estimates or quantities derived from these as functions of the tuning parameter m in (1.8), usually also supplemented by confidence intervals if appropriate. We introduce here a local Whittle plot in connection to fractal connectivity (Section 5.1) and discuss some issues behind a local Whittle plot for a phase parameter ϕ (Section 5.2). In Section 5.3, we also list the local Whittle plots that we suggest to examine for a bivariate series. These plots are then used with data illustrations in Section 6.

5.1 Case of fractal (non-)connectivity

The quantity that we suggest to consider in connection to fractal (non-)connectivity is

$$\rho^2 = \frac{\omega_{12}^2}{\omega_{11}\omega_{22}} = \frac{r_1^2 + r_2^2}{\omega_{11}\omega_{22}}, \quad (5.1)$$

and to examine its local Whittle plot based on

$$\hat{\rho}^2 = \frac{\hat{\omega}_{12}^2}{\hat{\omega}_{11}\hat{\omega}_{22}} = \frac{\hat{r}_1^2 + \hat{r}_2^2}{\hat{\omega}_{11}\hat{\omega}_{22}}. \quad (5.2)$$

Both ρ^2 and $\hat{\rho}^2$ take values in $[0, 1]$. Under H_0 in (2.22) (that is, fractal non-connectivity), we have

$$m\hat{\rho}^2 = \frac{m}{|\hat{\Omega}|}(\hat{r}_1^2 + \hat{r}_2^2) = \frac{\hat{\xi}_N}{2} \xrightarrow{d} \frac{\chi^2(2)}{2} \quad (5.3)$$

and under the alternative (that is, fractal connectivity), by using Theorem 2.1 (and the subsequent discussion) and the delta method, we have

$$\sqrt{m}(\hat{\rho}^2 - \rho^2) \xrightarrow{d} \mathcal{N}(0, \sigma_\rho^2), \quad (5.4)$$

where

$$\sigma_\rho^2 = \frac{2\omega_{12}^2|\Omega|^2}{\omega_{11}^3\omega_{22}^3}. \quad (5.5)$$

A local Whittle plot tracking $\hat{\rho}^2$ could then be supplemented by both a confidence interval based on (5.4)–(5.5), and by a critical value based on (5.3). A similar approach can be taken in the case of fractional cointegration, but in view of (2.29), based on

$$\rho_{fc}^2 = \frac{\nu_0^2 r_1^2 + (\nu_0 - 1)^2 r_2^2}{(\nu_0 - 1)^2 \omega_{11}\omega_{22}}, \quad \hat{\rho}_{fc}^2 = \frac{\hat{\nu}_0^2 \hat{r}_1^2 + (\hat{\nu}_0 - 1)^2 \hat{r}_2^2}{(\hat{\nu}_0 - 1)^2 \hat{\omega}_{11}\hat{\omega}_{22}}.$$

The asymptotic result in (5.3) then remains the same but the asymptotic variance in (5.5) has to be replaced accordingly by using Theorem 2.2 and the delta method.

The above approach is illustrated in Figure 1 for both fractally non-connected (left plot) and connected (middle plot) cases. For the non-connected case, we consider DGP2 in Section 4.1 and, for the connected case, we consider DGP1. The quantities $\hat{\rho}^2$ are plotted with 95% confidence intervals based on (5.4)–(5.5), and the dashed line represents the critical value for fractal non-connectivity based on (5.3). Note that the critical values also depend on the number of frequencies. For the fractally non-connected case (left plot), it is observed that $\hat{\rho}^2$ are below the critical values, hence we do not reject H_0 at the 5% significance level. On the other hand, in the middle plot, the estimated $\hat{\rho}^2$ are well above the critical values, hence suggesting that the data follow a fractally connected model.

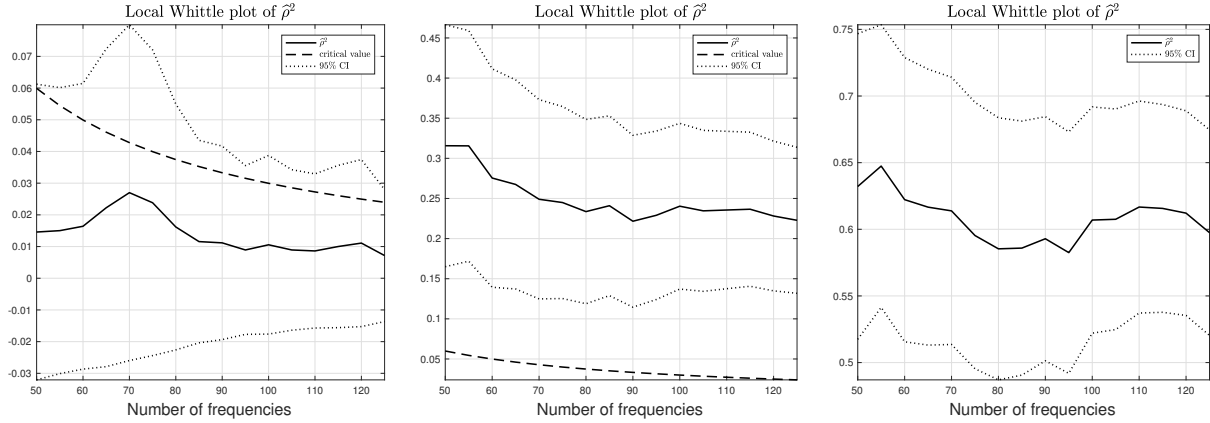


Figure 1: The local Whittle plot for $\hat{\rho}^2$ (in solid line) with 95% confidence intervals (in dotted line) under H_1 together with critical values (in dashed line) under H_0 . Left: fractally non-connected model. Middle: fractally connected model. Right: fractionally cointegrated model.

Remark 5.1 A somewhat bothersome feature of the results presented in this work is that the cases of fractional non-cointegration and fractional cointegration are treated separately. It is supposedly understood that a fractionally cointegrated model could be tried first and if $\beta = 0$ cannot be rejected, a fractionally non-cointegrated model could then be refit. A related question is then how fractional cointegration could be detected by working with a fractionally non-cointegrated model and the local Whittle estimation (1.7)–(1.8). In the non-cointegrated model formulation, fractional cointegration corresponds to (1.12) but recall also that (1.12) has to be excluded from the assumptions on the model parameters in Theorems 2.1 and 2.2. A more specific question is then whether Theorems 2.1 and 2.2 can be extended to accommodate the case (1.12).

We do not answer the latter question here but may do so in a future work. We nevertheless suggest that it may still be interesting to examine informally a local Whittle plot for the quantity ρ^2 in (5.1) even in connection to fractional cointegration. By (1.12), the quantity ρ^2 becomes

$$\rho^2 = 1 \quad (5.6)$$

in the case of fractional cointegration, and the local Whittle plot for ρ^2 should be indicative of (5.6). For example, the right panel of Figure 1 illustrates this point for the cointegrated model DGP3. This is by far a formal test but if $\hat{\rho}^2$ appears to be quite smaller than 1, there is no evidence a priori to reject a fractionally non-cointegrated model. Some of these points are illustrated further on real data in Section 6. We also stress that the confidence intervals are not valid for a fractionally cointegrated model but are included in line with other plots.

5.2 Case of phase parameter

We shall argue here that in some cases, the local Whittle plot of the phase parameter, to be referred to as the local Whittle phase plot, requires more careful interpretation, and we shall suggest its modification. To illustrate our thinking, consider DGP1 in Section 4.1 with sample size $N = 500$. Figure 2, left panel, shows a corresponding local Whittle phase plot. Observe that for some values of the tuning parameter m , the phase estimate is far from the true value but close to the boundary value $-\pi/2$. This observation should, in principal, not be surprising: the phase parameter ϕ enters into the model (1.2) and the likelihood (3.2)–(3.3) through a periodic function, and other related

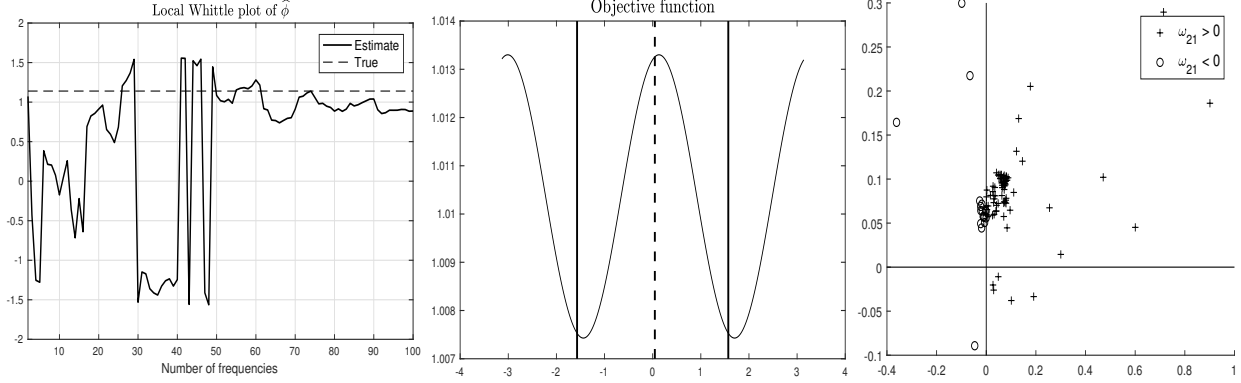


Figure 2: Left: Local Whittle phase plot with true value in dashed horizontal line. Middle: Objective function $R(\hat{D}, \phi)$ with $m = 35$. Right: The estimates of $g_{21} = \omega_{21}e^{i\phi}$, for a range of values of m , with those having estimated $\omega_{21} < 0$ in circle.

phase plots (e.g. the spectral phase plots for multivariate time series) have a similar feature. There are nevertheless a number of issues that, in our view, require further attention.

The shape of the local Whittle phase plot in Figure 2 can be explained from at least several angles. First, one could examine the objective function (3.2)–(3.3) of the local Whittle method as a function of the phase, keeping the two LRD parameters fixed at the estimated values. Moreover, it is instructive to plot this function not only for $\phi \in (-\pi/2, \pi/2)$ but for a wider range $\phi \in (-\pi, \pi)$. Figure 2, middle panel, depicts such a function for the same realization used in Figure 2, left plot, taking $m = 35$. The dashed vertical line indicates the initial value of the phase estimate in the optimization. The solid vertical lines are at $\pm\pi/2$. Note from the plot that the objective function has two local minima over $\phi \in (-\pi, \pi)$, having the same function value but separated exactly by the distance π on the horizontal axis. The right local minimum is close to the true value but is not selected since it falls outside the interval $(-\pi/2, \pi/2)$. For this value of m , the left local minimum falls in $(-\pi/2, \pi/2)$ and is selected for the phase estimate. We also note that if the initial estimate (dashed line) was further to the right, some optimization algorithms might yield the boundary point $\pi/2$ as the phase estimate (which we also observed for other realization but do not include the plots for the shortness sake).

Further insight into the local Whittle phase plot can be provided by considering the phase estimates along with the estimates of the parameter ω_{21} and, even better, by focusing on the estimates of $g_{21} = \omega_{21}e^{i\phi}$ in the complex plane. They are plotted in Figure 2, right plot, for the same realization used in Figure 2, left plot, and the same values of m . The points in the second and third quadrants are marked as circles in the plot. Since a phase belongs to $(-\pi/2, \pi/2)$ in our model parametrization, these points will have their estimated $\omega_{21} < 0$ and $\phi \in (-\pi/2, 0)$, in fact, with the corresponding phases close to $-\pi/2$ when the points are in the second quadrant.

The preceding discussion shows that the observed oscillations in the local Whittle phase plot are associated with transitions of the estimates of ω_{21} from one sign to the other. To avoid such sharp transitions, a natural modification would then consist of defining the phase estimates over $(-\pi, \pi)$ associated with a fixed sign of the estimates of ω_{21} . For example, to track positive $\hat{\omega}_{21}$, we modify the phase parameter estimator with values in the range $(-\pi, \pi)$ as

$$\hat{\phi}_+ = \begin{cases} \hat{\phi}, & \text{if } \hat{\omega}_{21} > 0, \\ \hat{\phi} + \pi, & \text{if } \hat{\omega}_{21} < 0, \quad -\pi/2 < \hat{\phi} \leq 0, \\ \hat{\phi} - \pi, & \text{if } \hat{\omega}_{21} < 0, \quad 0 < \hat{\phi} < \pi/2. \end{cases} \quad (5.7)$$

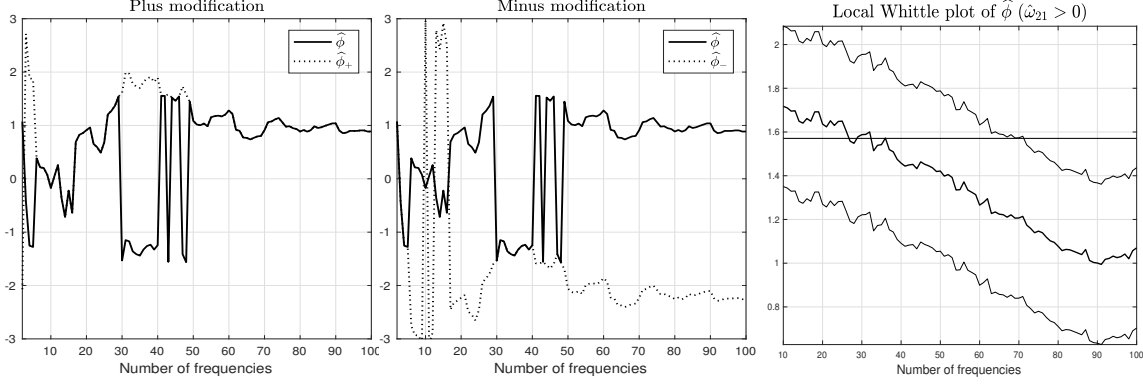


Figure 3: The local Whittle phase plots with modified phase estimates $\hat{\phi}_+$ (left) and $\hat{\phi}_-$ (middle). Right: The local Whittle phase plot with confidence intervals (in dashed lines). The horizontal line marks the value $\pi/2$.

That is, we are effectively using the parametrization on $g_{21} = \omega_{21}e^{i\phi_+}$ with $\omega_{21} > 0$ and $\phi_+ \in (-\pi, \pi)$ (see also Remark 5.2).

This modification is illustrated in Figure 3, left plot. The solid line represents the original phase estimates and the dashed line corresponds to the proposed modification. It is also possible to track negative $\hat{\omega}_{21}$ by considering

$$\hat{\phi}_- = \begin{cases} \hat{\phi}, & \text{if } \hat{\omega}_{21} < 0, \\ \hat{\phi} - \pi, & \text{if } \hat{\omega}_{21} > 0, \ 0 \leq \hat{\phi} < \pi/2, \\ \hat{\phi} + \pi, & \text{if } \hat{\omega}_{21} > 0, \ -\pi/2 < \hat{\phi} < 0, \end{cases} \quad (5.8)$$

as depicted in Figure 3, middle panel. Here, we are effectively using the parametrization $g_{21} = \omega_{21}e^{i\phi_-}$ with $\omega_{21} < 0$ and $\phi_- \in (-\pi, \pi)$.

In practice, it is not known a priori whether the true parameter ω_{21} is positive (or negative). Which modification of the local Whittle phase plot, positive or negative, should one use? Our practical suggestion is to decide by the majority rule of $\text{sign}(\hat{\omega}_{21})$ over some range of m , say $[N^{\cdot 5}]$ to $[N^{\cdot 8}]$. For instance, the local Whittle phase plot in Figure 2 points in favor of positive modification. Whichever modification is adopted, the user should keep in mind that the phase estimates outside $(-\pi/2, \pi/2)$ should actually be viewed as those in $(-\pi/2, \pi/2)$, after a suitable shift by $\pm\pi$ and a change of the sign of $\hat{\omega}_{21}$. As with all local Whittle plots, one reason to use a modified local Whittle phase plot is to assess the bias and the variability in the estimates, as to choose the tuning parameter m in balancing them. As shown in Tables 1 and 2, a modified local Whittle phase estimate combined with a phase confidence interval determined by Theorem 2.1 provides a good characterization of the variability of the (modified) phase parameter estimates. An example of a final modified local Whittle phase plot that we suggest to use is given in Figure 3, right panel, where confidence intervals for phase parameter are added to the positive phase modification.

Remark 5.2 With the discussion around (5.7)–(5.8), it might seem that starting with the parametrization $g_{21} = \omega_{21}e^{i\phi}$ where $\omega_{21} > 0$ and $\phi \in (-\pi, \pi)$ would resolve some of the issues discussed in this section. The latter parametrization, however, would bring its own issues. First, some of the formulas would be more involved to conform to the condition $\omega_{21} > 0$, for example, the expression (3.1). Second, similar oscillatory local Whittle phase plots would also be expected but now for phase parameters ϕ close to $\pm\pi$. A modification could then similarly be used by mapping some of the estimates in a periodic manner through $\pm 2\pi$.

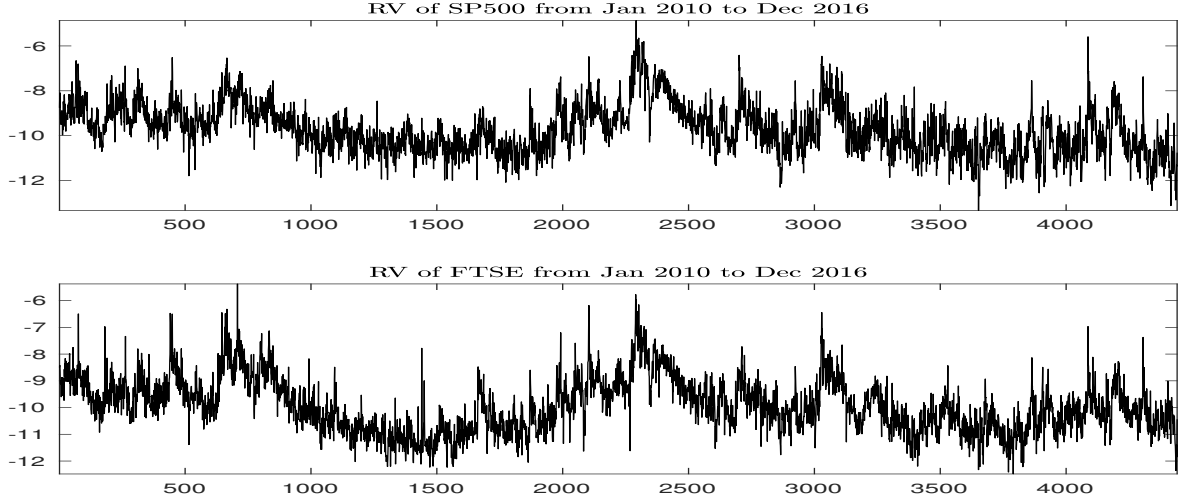


Figure 4: Time plots of RVs of SP and FTSE from Jan 2010 to Dec 2016.

5.3 Summary of plots

We suggest to examine the following 9 local Whittle plots of parameters and quantities of interest when having a bivariate series, possibly exhibiting long memory. The first 4 plots concern the fractionally non-cointegrated case and are the local Whittle plots of:

- $\hat{d}_1$ and $\hat{d}_2$;
- $\hat{\phi}$ (as discussed in Section 5.2);
- $\hat{r}_1$ and $\hat{r}_2$;
- $\hat{\rho}^2$ (as discussed in Section 5.1).

The other 5 plots concern the fractionally cointegrated case and are the local Whittle plots of:

- $\hat{\beta}$;
- $\hat{d}_1$ and $\hat{d}_2$;
- $\hat{\phi}$ (as discussed in Section 5.2);
- $\hat{r}_1$ and $\hat{r}_2$;
- $\hat{\rho}_{fc}^2$ (as discussed in Section 5.1).

An example of a figure with these 9 plots is Figure 5, discussed in greater detail in the next section.

6 Real data applications

The first data set concerns the so-called realized volatility (RV), obtained from the Oxford-Man Institute of Quantitative Finance² and processed by using the so-called realized kernels (RK) of Barndorff-Nielsen, Hansen, Lunde and Shephard (2009) to correct the possible bias of RV due to

² <http://realized.oxford-man.ox.ac.uk/>

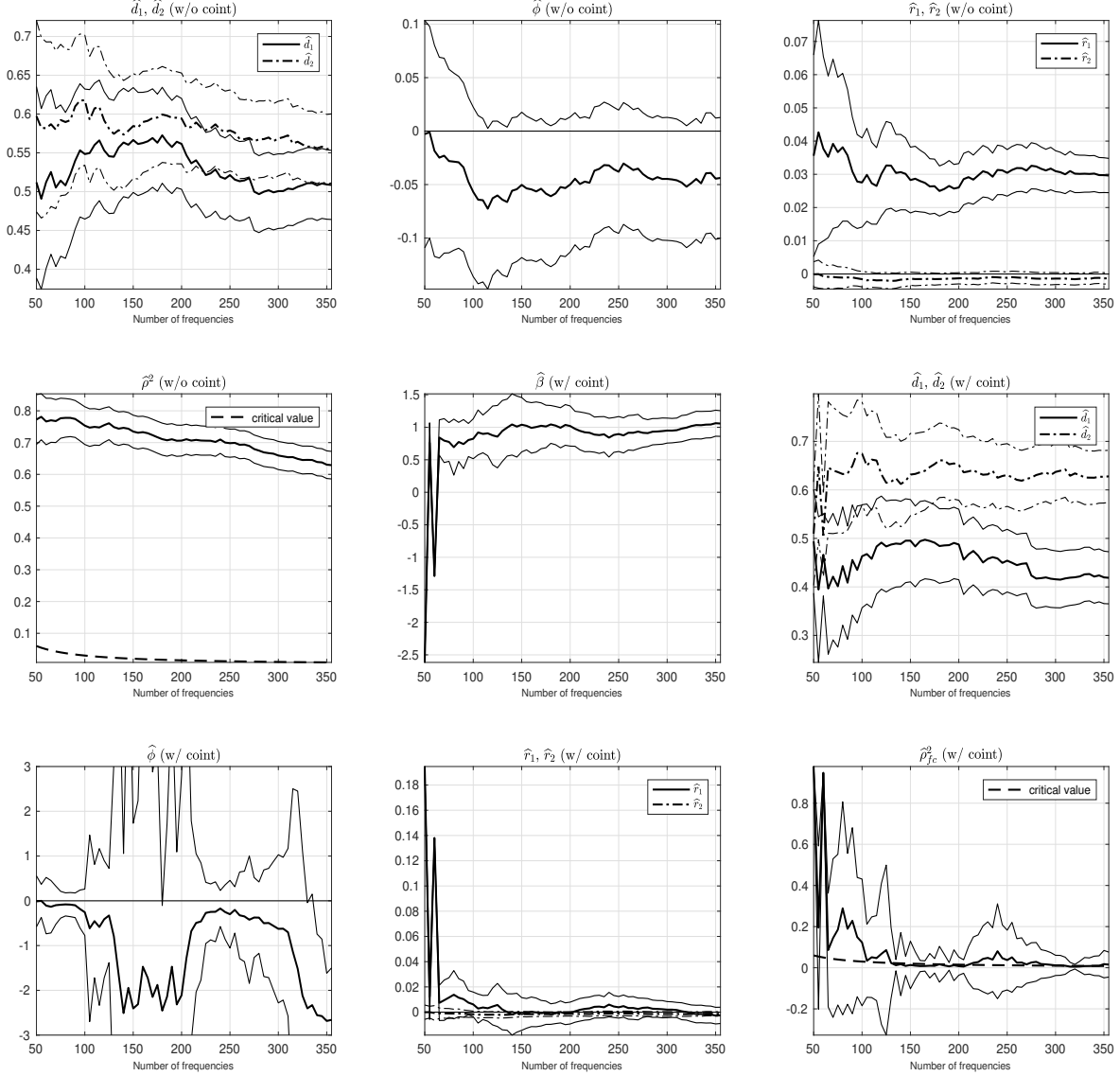


Figure 5: The local Whittle plots for SP and FTSE realized volatility.

microstructure dynamics. Two stock indices are considered, S&P 500 and FTSE. This data set aggregates 5-min within-day returns and covers 4414 trading days ranging from Jan 3, 2000 to Dec 30, 2016. The missing data due to different trading days are imputed by a linear interpolation. Figure 4 shows the time plots of RVs of the two stock indices.

Figure 5 presents the 9 local Whittle plots discussed in Section 5.3 for the RV data. A number of observations are in place. We first comment on the plots for the fractionally non-cointegrated case (with “w/o coint” in the title). In fact, these plots are consistent with what would be expected if the local Whittle analysis without fractional cointegration is applied to a fractionally cointegrated time series. Indeed, in this case, larger values (close to 1) of $\hat{\rho}^2$ are expected as discussed in Remark 5.1 and are also manifested in Figure 5. The values of $\hat{d}_1$ and $\hat{d}_2$ are also expected to be close to each other (cf. (1.12)) and appear such in the corresponding local Whittle plot of Figure 5. Finally, one also expects and has both the phase estimate $\hat{\phi}$ and hence $\hat{r}_2$ close to 0, as noted at the end of Remark 2.4.

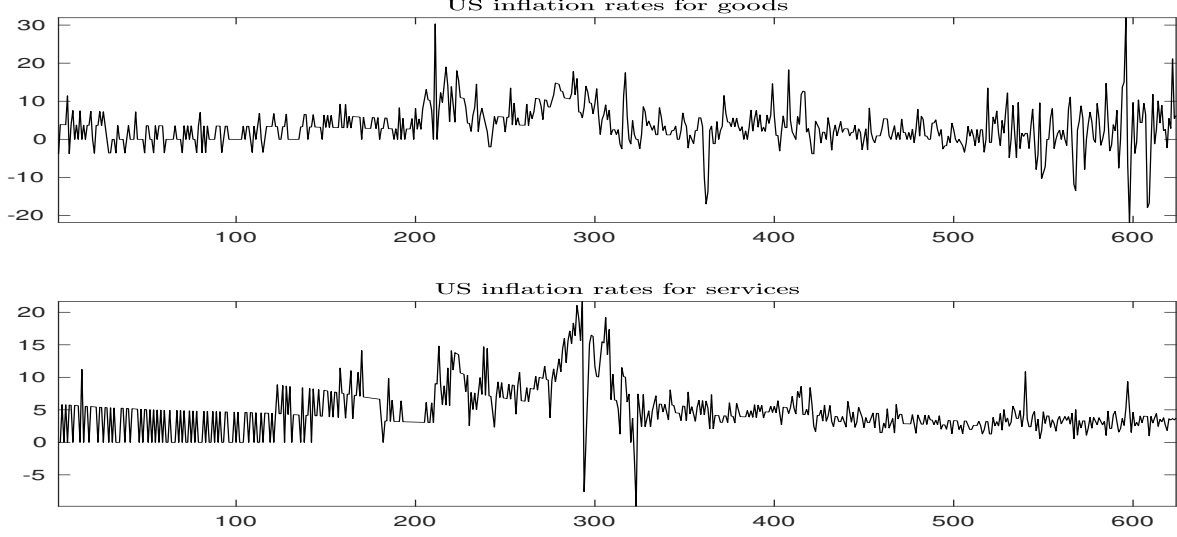


Figure 6: Time plots of the US inflation rates for goods and services from Feb 1956 to Jan 2008.

Turning to the plots for the fractionally cointegrated case (with “w/ coint” in the title), the local Whittle plot of $\hat{\beta}$ also suggests fractional cointegration with β around 1. The local Whittle plots of $\hat{d}_1$ and $\hat{d}_2$, as expected, suggest $d_1 < d_2$ though d_2 might actually be outside the stationarity regime $d_2 < .5$. The local Whittle plot of $\hat{\phi}$ is, in fact, modified as in Section 5.2, with several values of $\hat{\phi}$ falling below $-\pi/2$. The large confidence intervals for $\hat{\phi}$ at some numbers of m are due to a small estimate $\hat{\omega}_{11}$ (see the variance expression for $\hat{\phi}$ in Theorem 2.2). The local Whittle plots for both $\hat{r}_1$, $\hat{r}_2$ and $\hat{\rho}_{fc}^2$ suggest that the underlying model is non-connected; for example, note that many of the values of $\hat{\rho}_{fc}^2$ are below or around the critical value line in dashed. Thus, in conclusion, Figure 5 suggests that a fractionally cointegrated but fractally non-connected model is suitable for the RVs of S&P 500 and FTSE.

The second data set concerns the US monthly inflation rates for goods and services. These are defined as

$$g_n = 1200 \frac{CPI_n^c - CPI_{n-1}^c}{CPI_{n-1}^c}, \quad s_n = 1200 \frac{CPI_n^s - CPI_{n-1}^s}{CPI_{n-1}^s}, \quad n = 1, \dots, 624,$$

where $\{CPI_n^c\}$ and $\{CPI_n^s\}$ are the corresponding Consumer Price Indices for the period of January 1956–January 2008, and available from the Bureau of Labor Statistics. In connection to long memory, this bivariate series was analyzed in Baillie, Chung and Tieslau (1996), Doornik and Ooms (2004), Sela and Hurvich (2009), Sela (2010), Baillie and Morana (2012), Kechagias and Pipiras (2017). The time plots of the two series are given in Figure 6.

The local Whittle plots for this data set are presented in Figure 7. The key observation from these plots is that both fractionally non-cointegrated and fractionally cointegrated models seem suitable for the data. Note from the plots of $\hat{\rho}^2$ and $\hat{\rho}_{fc}^2$ that the models are suggested as fractally connected. Thus, in conclusion, a fractally connected model with or without fractional cointegration seems suitable. Kechagias and Pipiras (2017), in particular, advocate to use a model with fractional cointegration.

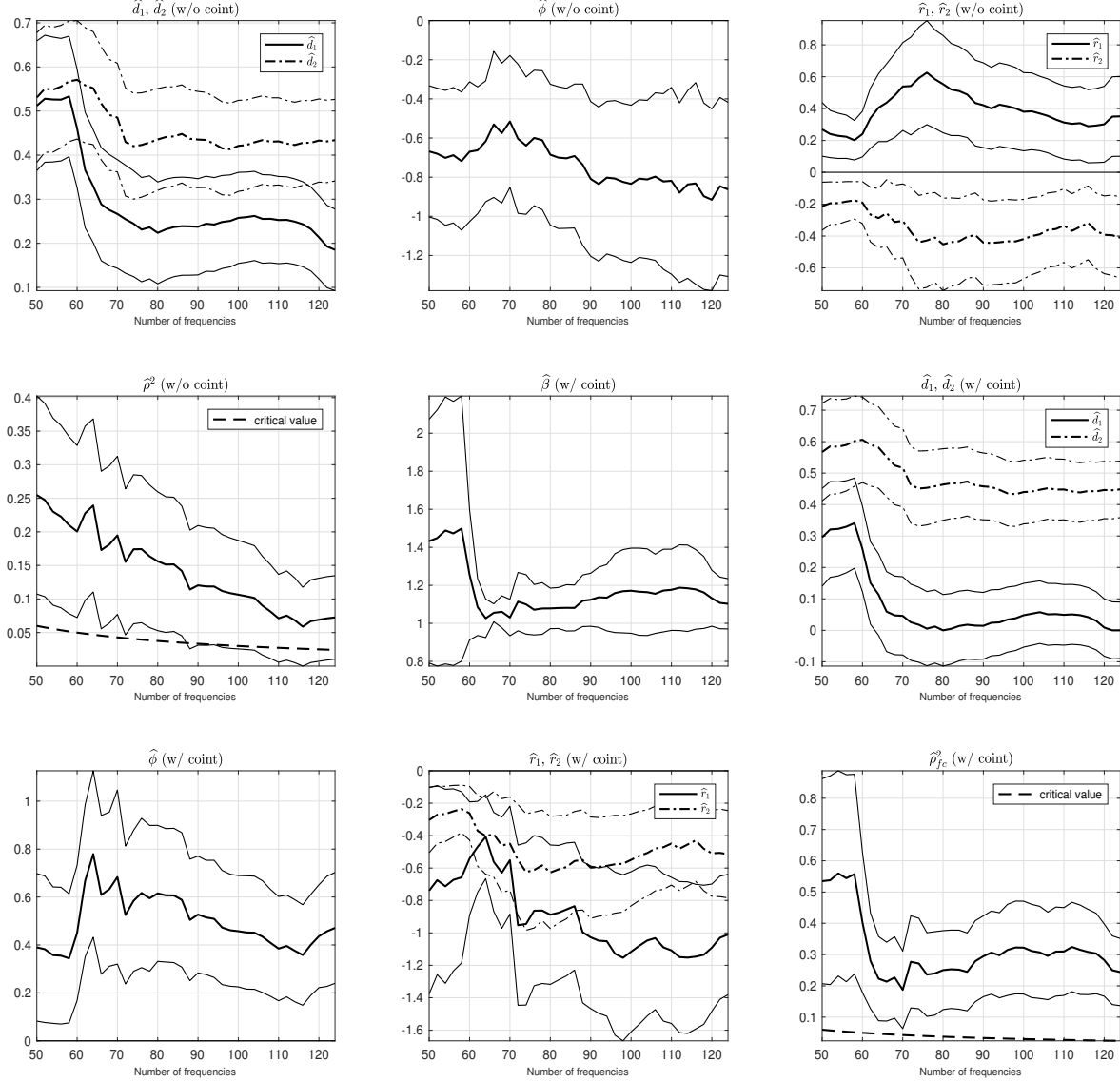


Figure 7: The local Whittle plots for the US inflation rates for goods and services.

7 Conclusions

In this work, we studied the asymptotics of all local Whittle estimators for bivariate stationary systems. We were motivated by the applications to fractal (non-)connectivity, and suggested to work in the “Complex” rather than “Polar” parametrization. We also discussed several new or modified local Whittle plots, and recommended to use and examined a list of these plots for real data.

Several questions raised in or related to this work could be examined in the future. As noted in Remark 2.4, it might be interesting to understand the local Whittle estimators (applied without cointegration) for fractionally cointegrated case. The conclusions about the RVs in Section 6 calls for appropriate fractionally cointegrated but fractally non-connected time series models. Another natural question is what happens for multivariate series in dimensions higher than 2.

A Proofs of Theorems 2.1 and 2.2

As noted in the beginning of Section 2, we focus on calculating the information matrix only, and shall make a number of simplifying replacements that are known to be justifiable. We consider first Theorem 2.1 in greater detail, and then only outline what is different for Theorem 2.2.

Observe from (1.8) that the negative log-likelihood is given as

$$Q(\theta) = \log |\Omega| - 2(d_1 + d_2) \frac{1}{m} \sum_{j=1}^m \log(\lambda_j) + \frac{1}{|\Omega|} T_X(\theta), \quad (\text{A.1})$$

where $|\Omega| = \omega_{11}\omega_{22} - \omega_{12}^2$ is the determinant of $|\Omega|$,

$$T_X(\theta) = \omega_{11}A_{22}^{(0)} + \omega_{22}A_{11}^{(0)} - 2\omega_{12}\Re(A_{12}^{(0)}e^{i\phi}), \quad (\text{A.2})$$

and we set

$$A_{pq}^{(k)} = \frac{1}{m} \sum_{j=1}^m \lambda_j^{d_p+d_q} I_{X,pq}(\lambda_j) (\log \lambda_j)^k, \quad p, q = 1, 2, \quad k = 0, 1, 2, \quad (\text{A.3})$$

where $I_{X,pq}$ is the (p, q) th entry of the periodogram I_X of X . The first derivative $(\partial/\partial\theta)Q(\theta)$ is given by

$$\begin{aligned} \frac{\partial Q}{\partial \omega_{11}} &= \frac{\omega_{22}}{|\Omega|} - \frac{\omega_{22}}{|\Omega|^2} T_X(\theta) + \frac{A_{22}^{(0)}}{|\Omega|}, & \frac{\partial Q}{\partial \omega_{12}} &= -\frac{2\omega_{12}}{|\Omega|} + \frac{2\omega_{12}}{|\Omega|^2} T_X(\theta) - \frac{2\Re(A_{12}^{(0)}e^{i\phi})}{|\Omega|}, \\ \frac{\partial Q}{\partial \omega_{22}} &= \frac{\omega_{11}}{|\Omega|} - \frac{\omega_{11}}{|\Omega|^2} T_X(\theta) + \frac{A_{11}^{(0)}}{|\Omega|}, & \frac{\partial Q}{\partial \phi} &= 2\omega_{12}\Im(A_{12}^{(0)}e^{i\phi}), \\ \frac{\partial Q}{\partial d_1} &= -2L + \frac{2\omega_{22}}{|\Omega|} A_{11}^{(1)} - \frac{2\omega_{12}}{|\Omega|} \Re(A_{12}^{(1)}e^{i\phi}), & \frac{\partial Q}{\partial d_2} &= -2L + \frac{2\omega_{11}}{|\Omega|} A_{22}^{(1)} - \frac{2\omega_{12}}{|\Omega|} \Re(A_{12}^{(1)}e^{i\phi}), \end{aligned}$$

where $L = m^{-1} \sum_{j=1}^m \log \lambda_j$ as appearing in (2.4).

The information matrix is obtained next by calculating the second derivative matrix of $Q(\theta)$ and taking expectation. We shall suppose for simplicity that $\mathbb{E}(I_{X,pq}(\lambda_j))$ can be replaced by its theoretical counterpart $\omega_{pq}e^{-i\phi_{pq}}\lambda_j^{-d_p-d_q}$, $p, q = 1, 2$ ($\phi_{pp} = 0, \phi_{12} = \phi = -\phi_{21}$), and hence

$$\mathbb{E}A_{pq}^{(0)} = \omega_{pq}e^{-i\phi_{pq}}, \quad \mathbb{E}A_{pq}^{(1)} = \omega_{pq}e^{-i\phi_{pq}}L, \quad \mathbb{E}A_{pq}^{(2)} = \omega_{pq}e^{-i\phi_{pq}}M, \quad (\text{A.4})$$

where, in addition, $M = m^{-1} \sum_{j=1}^m (\log \lambda_j)^2$ as in (2.4). Then, for example,

$$\begin{aligned} \frac{\partial^2 Q}{\partial \omega_{11}^2} &= -\frac{\omega_{22}^2}{|\Omega|^2} + \frac{2\omega_{22}^2}{|\Omega|^3} T_X(\theta) - \frac{2\omega_{22}}{|\Omega|^2} A_{22}^{(0)}, & \mathbb{E}\left(\frac{\partial^2 Q}{\partial \omega_{11}^2}\right) &= -\frac{\omega_{22}^2}{|\Omega|^2} + \frac{2\omega_{22}^2}{|\Omega|^3} 2|\Omega| - \frac{2\omega_{22}^2}{|\Omega|^2} = \frac{\omega_{22}^2}{|\Omega|^2}, \\ \frac{\partial^2 Q}{\partial \phi^2} &= \frac{2\omega_{12}}{|\Omega|} \Re(A_{12}^{(0)}e^{i\phi}), & \mathbb{E}\left(\frac{\partial^2 Q}{\partial \phi^2}\right) &= \frac{2\omega_{12}^2}{|\Omega|}, \\ \frac{\partial^2 Q}{\partial d_1^2} &= \frac{4\omega_{22}}{|\Omega|} A_{11}^{(2)} - \frac{2\omega_{12}}{|\Omega|} \Re(A_{12}^{(2)}e^{i\phi}), & \mathbb{E}\left(\frac{\partial^2 Q}{\partial d_1^2}\right) &= \frac{4\omega_{11}\omega_{22} - 2\omega_{12}^2}{|\Omega|} M, \end{aligned}$$

and so on. The information matrix is then given by

$$I(\theta) = \begin{pmatrix} \frac{\omega_{22}^2}{|\Omega|^2} & \frac{\omega_{12}^2}{|\Omega|^2} & -\frac{2\omega_{12}\omega_{22}}{|\Omega|^2} & 0 & -\frac{2\omega_{22}}{|\Omega|}L & 0 \\ & \frac{\omega_{11}^2}{|\Omega|^2} & -\frac{2\omega_{11}\omega_{12}}{|\Omega|^2} & 0 & 0 & -\frac{2\omega_{11}}{|\Omega|}L \\ & & \frac{2(\omega_{11}\omega_{22}+\omega_{12}^2)}{|\Omega|^2} & 0 & \frac{2\omega_{12}}{|\Omega|}L & \frac{2\omega_{12}}{|\Omega|}L \\ & & & \frac{2\omega_{12}^2}{|\Omega|} & 0 & 0 \\ & & & & \frac{2(2\omega_{11}\omega_{22}-\omega_{12}^2)}{|\Omega|}M & -\frac{2\omega_{12}^2}{|\Omega|}M \\ & & & & & \frac{2(2\omega_{11}\omega_{22}-\omega_{12}^2)}{|\Omega|}M \end{pmatrix}. \quad (\text{A.5})$$

Now, observe that, as $m \rightarrow \infty$,

$$L = \frac{1}{m} \sum_{j=1}^m (\log(j/m) + \log(2\pi m/N)) = \int_0^1 \log x dx + \log(2\pi m/N) + o(1) \sim \log(m/N), \quad (\text{A.6})$$

and similarly one can show (and as well known) that

$$M \sim (\log(m/N))^2, \quad M - L^2 = 1 + o(1). \quad (\text{A.7})$$

By taking the inverse of $I(\theta)$ and using these asymptotic orders of L, M and $M - L^2$ gives the asymptotic variance Γ as in (2.2). See also the discussion following Theorem 2.1.

Turning to fractional cointegration and Theorem 2.2, the negative log-likelihood can be written as (A.1) but replacing X by BX , that is,

$$Q_1(\theta_1) = \log |\Omega| - 2(d_1 + d_2) \frac{1}{m} \sum_{j=1}^m \log \lambda_j + \frac{1}{|\Omega|} T_{BX}(\theta_1), \quad (\text{A.8})$$

where $\theta_1 = (\beta, \theta)$ and $T_{BX}(\theta_1)$ is defined as in (A.2)–(A.3) but replacing X by BX . Since BX has the same spectral density as X in Theorem 2.1, we continue having (A.4). Moreover, since B and hence β appear only in BX , the information matrix for the parameters in θ is exactly the same as that given in (A.1). We thus need only to find the entries of the information matrix related to β . For example, we note that

$$\begin{aligned} I_{BX,11}(\lambda) &= I_{X,11}(\lambda) - 2\beta \Re(I_{X,12}(\lambda)) + \beta^2 I_{X,22}(\lambda), \\ I_{BX,12}(\lambda) &= I_{X,11}(\lambda) - \beta I_{X,12}(\lambda), \quad I_{BX,22} = I_{X,22}(\lambda). \end{aligned}$$

Then, for example,

$$\frac{\partial^2 Q_1}{\partial \beta^2} = \frac{1}{|\Omega|} \frac{\omega_{22}}{m} \sum_{j=1}^m \lambda_j^{2d_1} (2I_{X,22}(\lambda_j))$$

and

$$\mathbb{E} \left(\frac{\partial^2 Q_1}{\partial \beta^2} \right) = \frac{2\omega_{22}^2}{|\Omega|} \sum_{j=1}^m \lambda_j^{-2\nu_0} = \frac{2\omega_{22}^2}{|\Omega|} \frac{\lambda_m^{-2\nu_0}}{(1 - 2\nu_0)} + o(1),$$

where $\nu_0 = d_2 - d_1$, since

$$\frac{1}{m} \sum_{j=1}^m \lambda_j^{-2\nu_0} = \left(\frac{2\pi m}{N} \right)^{-2\nu_0} \int_0^1 x^{-2\nu_0} dx + o(1) = \frac{\lambda_m^{-2\nu_0}}{1 - 2\nu_0} + o(1).$$

The row of the information matrix associated with β can similarly be shown to be

$$I_{\beta}(\theta_1) = (I_{\beta} \ I_{\beta\omega_{11}} \ I_{\beta\omega_{22}} \ I_{\beta\omega_{12}} \ I_{\beta\phi} \ I_{\beta d_1} \ I_{\beta d_2}), \quad (\text{A.9})$$

where

$$\begin{aligned} I_{\beta} &= \frac{2\omega_{22}^2 \lambda_m^{-2\nu_0}}{|\Omega|(1-2\nu_0)}, \quad I_{\beta\omega_{11}} = 0, \quad I_{\beta\omega_{22}} = -\frac{2\omega_{12} \cos \phi}{|\Omega|} \frac{\lambda_m^{-\nu_0}}{1-\nu_0}, \quad I_{\beta\omega_{12}} = \frac{2\omega_{22} \cos \phi}{|\Omega|} \frac{\lambda_m^{-\nu_0}}{1-\nu_0}, \\ I_{\beta\phi} &= -\frac{2\omega_{12}\omega_{22} \sin \phi \lambda_m^{-\nu_0}}{|\Omega|(1-\nu_0)}, \quad I_{\beta d_1} = \frac{2\omega_{12}\omega_{22} \cos \phi \lambda_m^{-\nu_0}}{|\Omega|} \left(\frac{\nu_0}{(1-\nu_0)^2} - \frac{L}{1-\nu_0} \right) = -I_{\beta d_2}. \end{aligned}$$

The limiting covariance matrix Υ_{β} in Theorem 2.2 is obtained as the inverse of the information matrix.

B Proofs of Theorems 2.3 and 2.4

The proof of Theorem 2.3 is similar to that of Theorem 2.1. We present here only the negative log-likelihood and corresponding information matrix under Parametrization C. With a parameter vector $\theta = (g_{11}, g_{22}, r_1, r_2, d_1, d_2)$ and $|G| = g_{11}g_{22} - r_1^2 - r_2^2$, the negative log-likelihood is written as

$$P(\theta) = \log |G| - 2(d_1 + d_2) \frac{1}{m} \sum_{j=1}^m \log \lambda_j + \frac{1}{|G|} S_X(\theta), \quad (\text{B.1})$$

where

$$S_X(\theta) = g_{11}A_{22}^{(0)} + g_{22}A_{11}^{(0)} - 2r_1\Re(A_{12}^{(0)}) + 2r_2\Im(A_{12}^{(0)}).$$

Then, calculating the expectation of the second derivative of $P(\theta)$ and substituting

$$\mathbb{E}A_{pq}^{(0)} = g_{pq}, \quad \mathbb{E}A_{pq}^{(1)} = g_{pq}L, \quad \mathbb{E}A_{pq}^{(2)} = g_{pq}M \quad (\text{B.2})$$

gives the information matrix as

$$\begin{pmatrix} \frac{g_{22}^2}{|G|^2} & \frac{r_1^2+r_2^2}{|G|^2} & -\frac{2r_1g_{22}}{|G|^2} & -\frac{2r_2g_{22}}{|G|^2} & -\frac{2g_{22}}{|G|}L & 0 \\ & \frac{g_{11}^2}{|G|^2} & -\frac{2r_1g_{11}}{|G|^2} & -\frac{r_2g_{11}}{|G|^2} & 0 & -\frac{2g_{11}}{|G|}L \\ & & \frac{2g_{11}g_{22}+2r_1^2-2r_2^2}{|G|^2} & \frac{4r_1r_2}{|G|^2} & \frac{2r_1}{|G|}L & \frac{2r_1}{|G|}L \\ & & & \frac{2g_{11}g_{22}-2r_1^2+2r_2^2}{|G|^2} & \frac{2r_2}{|G|}L & \frac{2r_2}{|G|}L \\ & & & & \frac{2(2g_{11}g_{22}-r_1^2-r_2^2)}{|G|}M & -\frac{2(r_1^2+r_2^2)}{|G|}M \\ & & & & & \frac{2(2g_{11}g_{22}-r_1^2-r_2^2)}{|G|}M \end{pmatrix}. \quad (\text{B.3})$$

With fractional cointegration in Theorem 2.4 under Parametrization C, we can follow similar arguments as in the proof of Theorem 2.2. Indeed, replacing the periodogram in (B.1) by that of BX gives one extra parameter β . The information matrix for the parameter θ is exactly the same as in (B.3) and the information matrix associated with β can be shown to be:

$$\begin{aligned} I_{\beta} &= \frac{2g_{22}^2 \lambda_m^{-2\nu_0}}{|G|(1-2\nu_0)}, \quad I_{\beta g_{11}} = 0, \quad I_{\beta g_{22}} = -\frac{2r_1 \lambda_m^{-\nu_0}}{|G|(1-\nu_0)}, \quad I_{\beta r_1} = \frac{2g_{22} \lambda_m^{-\nu_0}}{|G|(1-\nu_0)}, \quad I_{\beta r_2} = 0, \\ I_{\beta d_1} &= -\frac{2r_1g_{22} \lambda_m^{-\nu_0}}{|G|} \left(\frac{\nu_0}{(1-\nu_0)^2} - \frac{L}{1-\nu_0} \right) = -I_{\beta d_2}. \end{aligned}$$

C Proofs for optimization reduction

We prove here the statements made in Section 3. Though (3.4)–(3.5) follows from (3.8)–(3.9) as noted at the end of Section 3, we shall provide two independent proofs since they are not very long. To show (3.4)–(3.5), observe from (3.3) that

$$\widehat{\Omega}(\phi, D) = \Re \begin{pmatrix} m^{-1} \sum_{j=1}^m \lambda_j^{2d_1} I_{X,11}(\lambda_j) & m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,12}(\lambda_j) e^{i\phi} \\ m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,21}(\lambda_j) e^{-i\phi} & m^{-1} \sum_{j=1}^m \lambda_j^{2d_2} I_{X,22}(\lambda_j) \end{pmatrix}.$$

Hence,

$$\begin{aligned} |\widehat{\Omega}(\phi, D)| &= \left(m^{-1} \sum_{j=1}^m \lambda_j^{2d_1} I_{X,11}(\lambda_j) \right) \left(m^{-1} \sum_{j=1}^m \lambda_j^{2d_2} I_{X,22}(\lambda_j) \right) \\ &\quad - \Re \left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,12}(\lambda_j) e^{i\phi} \right) \Re \left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,21}(\lambda_j) e^{-i\phi} \right). \end{aligned}$$

Observe further that the latter term reduces to

$$\left(\left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} \Re(I_{X,12}(\lambda_j)) \right) \cos \phi - \left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} \Im(I_{X,12}(\lambda_j)) \right) \sin \phi \right)^2 =: (A \cos \phi - B \sin \phi)^2.$$

Then, for the objective function $R(\phi, D)$ in (3.2), solving

$$\frac{\partial R(\phi, D)}{\partial \phi} = 0$$

is equivalent to solving

$$-2(A \cos \phi - B \sin \phi)(-A \sin \phi - B \cos \phi) = 0. \quad (\text{C.1})$$

Since $\tan \phi = -B/A$ corresponds to the local minimum, we have

$$\widehat{\phi}^e = -\arctan \left(\frac{m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} \Im(I_{X,12}(\lambda_j))}{m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} \Re(I_{X,12}(\lambda_j))} \right).$$

In the case of fractional cointegration, note that (3.3) becomes

$$\widehat{\Omega}(\phi, D) = \Re \left(\frac{1}{m} \sum_{j=1}^m \Phi_{D,\phi}(\lambda_j) B I_X(\lambda_j) B' \overline{\Phi}_{D,\phi}(\lambda_j) \right). \quad (\text{C.2})$$

Since

$$B I_X(\lambda_j) B' = \begin{pmatrix} I_{X,11}(\lambda_j) - 2\beta \Re(I_{X,12}(\lambda_j)) + \beta^2 I_{X,22}(\lambda_j) & I_{X,12}(\lambda_j) - \beta I_{X,22}(\lambda_j) \\ \overline{I}_{X,12}(\lambda_j) - \beta \overline{I}_{X,22}(\lambda_j) & \overline{I}_{X,22}(\lambda_j) \end{pmatrix}, \quad (\text{C.3})$$

by using the periodogram (C.3) in the above proof gives the explicit estimator of the phase parameter as in (3.7).

Turning to (3.8)–(3.9), note that

$$|\widehat{G}(D)| = \left| \frac{1}{m} \sum_{j=1}^m \lambda_j^D I_X(\lambda_j) \lambda_j^D \right| = \left| \begin{pmatrix} m^{-1} \sum_{j=1}^m \lambda_j^{2d_1} I_{X,11}(\lambda_j) & m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,12}(\lambda_j) \\ m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,21}(\lambda_j) & m^{-1} \sum_{j=1}^m \lambda_j^{2d_2} I_{X,22}(\lambda_j) \end{pmatrix} \right|$$

$$\begin{aligned}
&= \left(m^{-1} \sum_{j=1}^m \lambda_j^{2d_1} I_{X,11}(\lambda_j) \right) \left(m^{-1} \sum_{j=1}^m \lambda_j^{2d_2} I_{X,22}(\lambda_j) \right) \\
&\quad - \left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,12}(\lambda_j) \right) \left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} I_{X,21}(\lambda_j) \right).
\end{aligned}$$

Since $I_{X,21}(\lambda_j) = \bar{I}_{X,12}(\lambda_j)$, the latter term becomes

$$\left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} \Re(I_{X,12}(\lambda_j)) \right)^2 + \left(m^{-1} \sum_{j=1}^m \lambda_j^{d_1+d_2} \Im(I_{X,12}(\lambda_j)) \right)^2 =: A^2 + B^2.$$

Therefore,

$$|\widehat{\Omega}(D, \widehat{\phi}^e)| - |\widehat{G}(D)| = -(A \cos \widehat{\phi}^e - B \sin \widehat{\phi}^e)^2 + (A^2 + B^2) = (A \sin \widehat{\phi}^e + B \cos \widehat{\phi}^e)^2 = 0$$

from (C.1). Thus, the parameter D estimates are identical for (3.4) and (3.8) which implies that D can be estimated by using either Parametrization P or Parametrization C.

Acknowledgments

We would like to thank the anonymous Referee for providing many useful comments and suggestions that helped improving significantly the presentation. The work of the first author was supported in part by the Basic Science Research Program from the National Research Foundation of Korea (NRF-2017R1A1A1A05000831, NRF-2019R1F1A1057104). The third author was supported in part by the NSF grant DMS-1712966.

References

- Achard, S., Bassett, D. S., Meyer-Lindenberg, A. and Bullmore, E. (2008), ‘Fractal connectivity of long-memory networks’, *Physical Review E* **77**, 036104.
- Baillie, R. T., Chung, C.-F. and Tieslau, M. A. (1996), ‘Analysing inflation by the fractionally integrated ARFIMA-GARCH model’, *Journal of Applied Econometrics* **11**(1), 23–40.
- Baillie, R. T. and Morana, C. (2012), ‘Adaptive ARFIMA models with applications to inflation’, *Economic Modelling* **29**(6), 2451–2459.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2009), ‘Realized kernels in practice: trades and quotes’, *Econometrics Journal* **12**(3), C1–C32.
- Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013), *Long-Memory Processes: Probabilistic Properties and Statistical Methods*, Springer.
- Doornik, J. A. and Ooms, M. (2004), ‘Inference and forecasting for ARFIMA models with an application to US and UK inflation’, *Studies in Nonlinear Dynamics & Econometrics* **8**(2).
- Giraitis, L., Koul, H. L. and Surgailis, D. (2012), *Large Sample Inference for Long Memory Processes*, Imperial College Press, London.
- Helgason, H., Pipiras, V. and Abry, P. (2011), ‘Fast and exact synthesis of stationary multivariate Gaussian time series using circular embedding’, *Signal Processing* **91**(5), 1123–1133.
- Kechagias, S. and Pipiras, V. (2017), Identification, estimation and applications of a bivariate long-range dependent time series model with general phase. Preprint.

- Kristoufek, L. (2013), ‘Mixed-correlated ARFIMA processes for power-law cross-correlations’, *Physica A: Statistical Mechanics and its Applications* **392**(24), 6484–6493.
- Nielsen, F. S. (2011), ‘Local Whittle estimation of multi-variate fractionally integrated processes’, *Journal of Time Series Analysis* **32**(3), 317–335.
- Nielsen, M. Ø. (2007), ‘Local Whittle analysis of stationary fractional cointegration and the implied realized volatility relation’, *Journal of Business & Economic Statistics* **25**(4), 427–446.
- Nielsen, M. Ø. and Shimotsu, K. (2007), ‘Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach’, *Journal of Econometrics* **141**(2), 574–596.
- Pipiras, V. and Taqqu, M. S. (2017), *Long-Range Dependence and Self-Similarity*, Cambridge University Press, Cambridge.
- Robinson, P. M. (1995), ‘Log-periodogram regression of time series with long range dependence’, *The Annals of Statistics* **23**(3), 1048–1072.
- Robinson, P. M. (2008), ‘Multiple local Whittle estimation in stationary systems’, *The Annals of Statistics* **36**(5), 2508–2530.
- Robinson, P. M., ed. (2003), *Time Series With Long Memory*, Advanced Texts in Econometrics, Oxford University Press, Oxford.
- Sela, R. J. (2010), Three essays in econometrics: multivariate long memory time series and applying regression trees to longitudinal data, PhD thesis, New York University.
- Sela, R. J. and Hurvich, C. M. (2009), ‘Computationally efficient methods for two multivariate fractionally integrated models’, *Journal of Time Series Analysis* **30**(6), 631–651.
- Shimotsu, K. (2007), ‘Gaussian semiparametric estimation of multivariate fractionally integrated processes’, *Journal of Econometrics* **137**(2), 277–310.
- Shimotsu, K. (2012), ‘Exact local Whittle estimation of fractionally cointegrated systems’, *Journal of Econometrics* **169**(2), 266–278.
- Wendt, H., Didier, G., Combexelle, S. and Abry, P. (2017), ‘Multivariate Hadamard self-similarity: Testing fractal connectivity’, *Physica D: Nonlinear Phenomena* **356–357**, 1–36.
- Wendt, H., Scherrer, A., Abry, P. and Achard, S. (2009), Testing fractal connectivity in multivariate long memory processes, in ‘Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on’, IEEE, pp. 2913–2916.
- Wolfram Research (2017), ‘Mathematica, Version 11.2’. Champaign, IL.

Changryong Baek
 Dept. of Statistics
 Sungkyunkwan University
 25-2, Sungkyunkwan-ro, Jongno-gu
 Seoul, 110-745, Korea
 crbaek@skku.edu

Stefanos Kechagias
 SAS Institute
 100 SAS Campus Drive
 Cary, NC 27513, USA
 Stefanos.Kechagias@sas.com

Vladas Pipiras
 Dept. of Statistics and Operations Research
 UNC at Chapel Hill
 CB#3260, Hanes Hall
 Chapel Hill, NC 27599, USA
 pipiras@email.unc.edu