The Averaged Periodogram Estimator for a Power Law in Coherency

Rebecca J. Sela
J.P. Morgan Chase and Co.
and
Clifford M. Hurvich
New York University

August 30, 2011

Abstract: We prove the consistency of the averaged periodogram estimator (APE) in two new cases. First, we prove that the APE is consistent for negative memory parameters, after suitable tapering. Second, we prove that the APE is consistent for a power law in the cross-spectrum and therefore for a power law in the coherency, provided that sufficiently many frequencies are used in estimation. Simulation evidence suggests that the lower bound on the number of frequencies is a necessary condition for consistency. For a Taylor series approximation to the estimator of the power law in the cross-spectrum, we consider the rate of convergence, and obtain a central limit theorem under suitable regularity conditions.

1 Introduction

The averaged periodogram estimator (APE) was first introduced by Robinson [1994] and was extended to estimate the memory parameters of multiple time series by Lobato [1997]. As far as we are aware, no one has applied the APE to estimating the memory parameter of a long memory time series in the case where the memory parameter may be negative. Furthermore, while Lobato applied the APE to multivariate time series, he did not estimate the power law in the cross-spectrum. This paper addresses both of these issues.

In this paper, we will focus on real-valued bivariate time series, $X_t = (x_{1t}, x_{2t})'$, with a spectral density matrix given by:

$$f(\lambda) = \begin{pmatrix} f_{11}(\lambda) & f_{12}(\lambda) \\ f_{21}(\lambda) & f_{22}(\lambda) \end{pmatrix}, \quad \lambda \in [-\pi, \pi]$$

where $f^*(\lambda) = f(\lambda)$ and $f(-\lambda) = \overline{f(\lambda)}$, with A^* denoting the conjugate transpose of a matrix A. In this paper, we assume that the individual series have long memory, so that the spectral densities obeys $f_{jj}(\lambda) \sim C_j|1 - e^{-i\lambda}|^{-2d_j}$ as $\lambda \to 0^+$, for $C_j > 0$, $d_j < \frac{1}{2}$, and $j = 1, 2$. We will sometimes use d_{jj} as an alternative notation for d_j.

The cross-spectrum, $f_{12}(\lambda)$, can be decomposed into the phase, $\phi(\lambda)$, the coherency, $\rho(\lambda)$, and terms involving the auto-spectra:

$$f_{12}(\lambda) \quad = \quad \sqrt{f_{11}(\lambda)f_{22}(\lambda)}\rho(\lambda)e^{i\phi(\lambda)} \tag{1.1}$$

where the coherency is a real, even function with $0 \le \rho(\lambda) \le 1$ and the phase is an odd function that we assume lies in the interval $(-\pi, \pi]$. In this paper, we will focus on the behavior of the coherency. Specifically, we assume that $|f_{12}(\lambda)| \sim C_{12}|1 - e^{-i\lambda}|^{-2d_{12}}$, where $d_{12} \le \frac{1}{2}(d_1 + d_2)$. (If $d_{12} > \frac{1}{2}(d_1 + d_2)$, the spectral density matrix would not be positive definite and the implied coherency would be greater than one.) Then, the coherency satisfies:

$$\rho(\lambda) = C_\rho \lambda^{-2d_\rho} + o(\lambda^{-2d_\rho}) \tag{1.2}$$

as $\lambda \to 0^+$, where $C_\rho > 0$ and $d_\rho = d_{12} - \frac{1}{2}(d_1 + d_2) \le 0$. When $d_\rho < 0$, we will refer to the time series as having *power law coherency*. The fact that d_{12} need not equal $\frac{1}{2}(d_1 + d_2)$ was mentioned by Lobato [1997, page 139], though he did not try to estimate d_{12}. We present one time-domain example of power law coherency in Section 2. For additional time-domain examples and a discussion of possible behaviors of the phase, see Sela [2010].

There is a large literature on semiparametric methods for estimating d_1, d_2, based on the periodogram in a neighborhood of zero. Estimation methods that can be applied to univariate series include the averaged periodogram estimator (APE) [Robinson, 1994, Lobato and Robinson, 1996], the log periodogram (GPH) estimator [Geweke and Porter-Hudak, 1983, Robinson, 1995a], and the Gaussian semiparametric estimator (GSE) [Kunsch, 1987, Robinson, 1995b]. Many of these methods have been studied for multivariate time series (Lobato [1997] for APE, Lobato [1999], Shimotsu [2007] for GSE), but the power laws being estimated have been only d_1, d_2, not d_{12} as far as we are aware.

In Section 2, we introduce a semiparametric long-memory time series model that allows for power law coherency, and present a time-domain example. In Section 3, we first show that the averaged periodogram estimator (APE) is consistent for $d_1, d_2 < 1/2$ and for d_{12}, under certain conditions. Then, we consider a Taylor series approximation to the proposed estimator of d_{12}, derive its rate of convergence and, in the case $d_{12} < 1/4$, obtain a central limit theorem. Unfortunately, as we will see in simulations in Section 4, high variability of the estimators in sample sizes typically used in practice makes power laws in the coherency hard to detect with the APE. In Section 5, we apply the APE to a bivariate time series of two components of the money supply.

2 Semiparametric model

We introduce a semiparametric model for a bivariate time series, $\{X_t\}$, that will allow for power law coherency (as well as many other types of behavior). Each $\{x_{jt}\}$, for $j = 1, 2$, is the sum of up to p component series, where each component series may have a different memory parameter. Making this representation explicit allows us to calculate the resulting behavior of the coherency in a neighborhood of zero frequency. Though our model is semiparametric in the sense that it specifies power laws only in a neighborhood of zero frequency, we do (for ease of presentation and proofs) in some of the assumptions below place conditions on certain functions that are global in that they hold on $[-\pi, \pi] - \{0\}$. As Chen and Hurvich [2006, page 2948] note in a related context, it may be possible (and would be desirable) to replace these assumptions by local versions, but we will not pursue this here.

We assume that $\{X_t\}$ has the infinite-order moving average representation

$$X_t = \sum_{r=-\infty}^{\infty} \psi_r \epsilon_{t-r} \tag{2.1}$$

where the real-valued $2 \times p$ matrices, ψ_r, are specified below and $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{pt})'$ is a p-variate, zero-mean series ($p \geq 2$) that satisfies the following:

Assumption 1 *$\{\epsilon_t\}$ is independent and identically distributed with:*

- *$Cov(\epsilon_t) = 2\pi\Sigma$, where Σ is symmetric and positive definite.*
- *$E\left(\epsilon_{kt}^4\right) < \infty$ for $k = 1, \ldots, p$.*

Allowing for more than two driving innovation series allows for straightforward descriptions of a rich variety of models. Other authors, including Hannan [1970] and Robinson [2008], have also allowed $p > 2$ in models for a bivariate series. As Robinson [2008] states, "This is natural if [the components of $\{X_t\}$] are seen as just two of a vector of related observations that are analyzed pairwise." Power law coherency is easily obtained in such a framework, since the entries of $\{X_t\}$ can then have some common long-memory components and some idiosyncratic components, with the memory in the common components weaker than that for any of the idiosyncratic components.

Equation (2.1) implies that $\{X_t\}$ is the output of passing $\{\epsilon_t\}$ through a linear filter with transfer function $\Psi(\lambda) = \sum_{r=-\infty}^{\infty} \psi_r e^{-i\lambda r}$, a $2 \times p$ matrix with entries $\Psi_{jk}(\lambda)$, for $j = 1, 2$ and $k = 1, \ldots, p$. For each (j, k), we generalize Chen and Hurvich [2006] and consider transfer functions, $\Psi_{jk}(\lambda)$ on $[-\pi, \pi]$, that can be written as:

$$\Psi_{jk}(\lambda) = (1 - e^{-i\lambda})^{-\delta_{jk}} \tau_{jk}(\lambda) e^{i\varphi_{jk}(\lambda)} \tag{2.2}$$

with $\tau_{jk}(\lambda), \delta_{jk}, \varphi_{jk}(\lambda)$ satisfying Assumptions 2-5 (Assumption 5 appears after some discussion of Assumptions 2-4). These assumptions relax those of Chen and Hurvich [2006] since we do not require for all j, k that $\delta_{jj} \geq \delta_{jk}$, or that $\tau_{jk}(0) > 0$, or that $\tau_{jk}(\lambda)$ and $\varphi_{jk}(\lambda)$ are differentiable at zero frequency.

Assumption 2 *For $j = 1, 2$ and $k = 1, \ldots, p$, $\tau_{jk}(\lambda)$ is a real, bounded, non-negative, continuous, even function on $[-\pi, \pi]$ that is differentiable on $[-\pi, \pi] - \{0\}$, and there exist positive constants C, A and Ξ such that $|\tau'_{jk}(\lambda)| \leq C\lambda^{-1+\Xi}$ for all $\lambda \in (0, A)$. Furthermore, either $\tau_{jk}(0) > 0$ or $\tau_{jk}(\lambda) \equiv 0$ for all $\lambda \in [0, \pi]$; for each j, $\tau_{jk}(0) > 0$ for at least one k.*

Assumption 3 *$\delta_{jk} < 1/2$ for all $j = 1, 2$ and $k = 1, \ldots, p$.*

Assumption 4 *$\varphi_{jk}(\lambda)$ is an odd, differentiable function on $[-\pi, \pi] - \{0\}$, where $\lim_{\lambda \to 0^+} \varphi_{jk}(\lambda)$ exists and $\varphi'_{jk}(\lambda)$ is continuous at 0 or there exist positive constants C, A and Ξ such that $|\varphi'_{jk}(\lambda)| \leq C\lambda^{-1+\Xi}$ for all $\lambda \in (0, A)$.*

The decomposition in Equation (2.2) separates the transfer function into three different operations that transform the series $\{\epsilon_{kt}\}$ into a component of $\{x_{jt}\}$. First, $(1 - e^{-i\lambda})^{-\delta_{jk}}$ is the fractional integration operator of order δ_{jk}. Because δ_{jk} varies with k, $\{x_{jt}\}$ consists of components with potentially different orders of integration. Second, $\tau_{jk}(\lambda)$ is either a filter that changes the short-memory properties of the resulting fractionally integrated or a filter that annihilates the component. Finally, $\varphi_{jk}(\lambda)$ changes the phase of $\{x_{jt}\}$ relative to the original $\{\epsilon_{kt}\}$. For example, if $\varphi_{jk}(\lambda) = -a\lambda$, for some real a, then the component of $\{x_{jt}\}$ that depends on $\{\epsilon_{kt}\}$ is lagged by a periods. More complicated phase shifts are also possible.

The spectral density of $\{X_t\}$ is simply:

$$f(\lambda) \quad = \quad \Psi(\lambda)\Sigma\Psi(\lambda)^*, \;\; \lambda \in [-\pi, \pi]$$

Using the representation in Equation (2.2), the autospectral densities, $f_1(\lambda)$ and $f_2(\lambda)$, are given for $j = 1, 2, \lambda \in [-\pi, \pi]$ by:

$$f_j(\lambda) \quad = \quad \sum_{k=1}^{p} \sum_{l=1}^{p} (1 - e^{-i\lambda})^{-\delta_{jk}} (1 - e^{i\lambda})^{-\delta_{jl}} \sigma_{kl} \tau_{jk}(\lambda) \tau_{jl}(\lambda) e^{i(\varphi_{jk}(\lambda) - \varphi_{jl}(\lambda))}$$

where σ_{kl} is the (k, l) element of Σ. The power laws in the auto-spectra are determined by the largest δ_{jk} that have non-zero coefficients, $\tau_{jk}(0)^2$, in the contribution to the sum above with $k = \ell$; the contribution for $k \neq \ell$ will not change the power law in the auto-spectrum. Thus, we define

$$d_j = \max_{k : \tau_{jk}(0) > 0} \delta_{jk} \;\; .$$

We will sometimes find it convenient to use the alternative notation d_{jj} for d_j, and $f_{jj}(\lambda)$ for $f_j(\lambda)$. The semiparametric model implies that, as $\lambda \to 0^+$:

$$f_j(\lambda) \quad \sim \quad C_j \lambda^{-2d_j} \tag{2.3}$$

where

$$C_j = \lim_{\lambda \to 0^+} \sum_{k=1}^{p} \sum_{l=1}^{p} \sigma_{kl} \tau_{jk}(\lambda) \tau_{jl}(\lambda) e^{i(\varphi_{jk}(\lambda) - \varphi_{jl}(\lambda))} \chi(\delta_{jk} = \delta_{jl} = d_j) \tag{2.4}$$

and χ is an indicator function. Note that $C_j > 0$ since it can be expressed as a quadratic form in the positive definite matrix, Σ, after expressing the indicator function in (2.4) as the product $\chi(\delta_{jk} = d_j)\chi(\delta_{jl} = d_j)$.

In describing the cross-spectrum and coherency, it will be convenient to separate the power law into its modulus and argument. When $\lambda \in (0, \pi]$, we use the identity $(1 - e^{-i\lambda}) = \left|2 \sin \frac{\lambda}{2}\right| e^{i(\pi - \lambda)/2}$ to rewrite Equation (2.2) as:

$$\Psi_{jk}(\lambda) \quad = \quad \left|2 \sin \frac{\lambda}{2}\right|^{-\delta_{jk}} \tau_{jk}(\lambda) e^{i(\varphi_{jk}(\lambda) - (\pi - \lambda)\delta_{jk}/2)} \tag{2.5}$$

Using Equations (2.2), (2.5), and the identity above, we find that for $\lambda \in (0, \pi]$, the cross-spectral density is given by:

$$\begin{aligned} f_{12}(\lambda) \quad &= \quad \sum_{k=1}^{p} \sum_{l=1}^{p} (1 - e^{-i\lambda})^{-\delta_{1k}} (1 - e^{i\lambda})^{-\delta_{2l}} \sigma_{kl} \tau_{1k}(\lambda) \tau_{2l}(\lambda) e^{i(\varphi_{1k}(\lambda) - \varphi_{2l}(\lambda))} \\ &= \quad \sum_{k=1}^{p} \sum_{l=1}^{p} \left|2 \sin \frac{\lambda}{2}\right|^{-\delta_{1k} - \delta_{2l}} \sigma_{kl} \tau_{1k}(\lambda) \tau_{2l}(\lambda) e^{i[\varphi_{1k}(\lambda) - \varphi_{2l}(\lambda) - (\pi - \lambda)\delta_{1k}/2 + (\pi - \lambda)\delta_{2l}/2]} \end{aligned}$$

In order to understand the power law behavior of the cross-spectrum, we decompose the sum above into a sum of terms where the power, $\delta_{1k} + \delta_{2l}$, is constant. To do this, partition the set of $\{(k, l) : k, l \in \{1, ..., p\}\}$ into sets $S_1, ..., S_{\tilde{Q}}$ such that $\delta_{1k} + \delta_{2l} = \delta_{1k'} + \delta_{2l'}$ if and only if $(k, l), (k', l')$ are in the same set. Define $d_{12}(\tilde{q})$ to be the value of $\frac{1}{2}(\delta_{1k} + \delta_{2l})$ for $(k, l) \in S_{\tilde{q}}$ for $\tilde{q} = 1, ..., \tilde{Q}$, with $d_{12}(\tilde{q}) > d_{12}(\tilde{q} + 1)$ for all $\tilde{q} = 1, ..., \tilde{Q} - 1$. Note that $d_{12}(1) = \frac{1}{2}(d_1 + d_2)$. Then, for $0 < \lambda < \pi$, we may write:

$$f_{12}(\lambda) \quad = \quad \sum_{\tilde{q}=1}^{\tilde{Q}} \left|2 \sin \frac{\lambda}{2}\right|^{-2d_{12}(\tilde{q})} s(\lambda; \tilde{q}) \tag{2.6}$$

where

$$s(\lambda; \tilde{q}) = \sum_{(k,l) \in S_{\tilde{q}}} \sigma_{kl} \tau_{1k}(\lambda) \tau_{2l}(\lambda) e^{i[\varphi_{1k}(\lambda) - \varphi_{2l}(\lambda) - (\pi - \lambda)\delta_{1k}/2 + (\pi - \lambda)\delta_{2l}/2]}. \tag{2.7}$$

Define

$$s(0; \tilde{q}) := \lim_{\lambda \to 0^+} s(\lambda; \tilde{q}) = \sum_{(k,l) \in S_{\tilde{q}}} \sigma_{kl} \tau_{1k}(0) \tau_{2l}(0) \lim_{\lambda \to 0^+} \left(e^{i[\varphi_{1k}(\lambda) - \varphi_{2l}(\lambda) - (\pi - \lambda)\delta_{1k}/2 + (\pi - \lambda)\delta_{2l}/2]} \right).$$

To ensure that the power law behavior in the cross-spectral density is determined by the terms containing $\left|2\sin\frac{\lambda}{2}\right|^{-d_{12}(\tilde{q})}$ instead of by the $s(\lambda;\tilde{q})$, we make the following assumption:

Assumption 5 *There is at least one $\tilde{q}$ such that $s(0;\tilde{q}) \neq 0$. Let $\tilde{q}_0$ be the smallest such $\tilde{q}$. Then, we define:*

$$d_{12} = d_{12}(\tilde{q}_0) \tag{2.8}$$

Whenever $s(0;\tilde{q}) = 0$, we assume that there exist positive constants C, A and Ξ such that

$$|s(\lambda;\tilde{q})| \leq C\left(\lambda^{-2(d_{12}-d_{12}(\tilde{q}))+\Xi}\right) \ , \ \lambda \in (0, A).$$

Using this representation, we describe the power laws in the modulus of the cross-spectrum and the coherency in a neighborhood of zero frequency. As $\lambda \to 0^+$, the modulus of the cross spectrum and the coherency obey:

$$\begin{aligned} |f_{12}(\lambda)| &\sim C_{12}\lambda^{-2d_{12}} \\ \rho(\lambda) &\sim \frac{C_{12}}{\sqrt{C_1 C_2}}\lambda^{-2\left(d_{12}-\frac{1}{2}(d_1+d_2)\right)} \end{aligned} \tag{2.9}$$

where the power law of the cross-spectrum, d_{12}, is defined by Equation (2.8), C_1, C_2 are defined by Equation (2.3) and $C_{12} = |s(0;\tilde{q}_0)|$. Since $\delta_{1k} + \delta_{2l} \leq d_1 + d_2$, d_{12} is bounded above by $\frac{1}{2}(d_1 + d_2)$. In the case where $d_{12} < \frac{1}{2}(d_1 + d_2)$, power law coherency occurs. We define:

$$d_\rho = d_{12} - \frac{1}{2}(d_1 + d_2) \tag{2.10}$$

The only decay rate of the cross-spectral density that will not lead to power law coherency is $d_{12} = \frac{1}{2}(d_1 + d_2)$, that is, $d_\rho = 0$. Autoregressive fractionally integrated moving average (ARFIMA) models will have $d_\rho = 0$, and hence no power law coherency. Fractional cointegration of the components of $\{X_t\}$, though not ruled out in this paper, nevertheless implies that $d_\rho = 0$, and indeed that $\rho(0) = 1$, so once again there would be no power law coherency in this case. Under Assumption 5, power law coherency will occur if and only if $s(0;1) = 0$.

We present here a parametric model that has power law coherency. Assume that $d_3 < d_2 \leq d_1 < 1/2$ and that $\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}$ are independent white noise series with variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$, respectively. Define a bivariate time series by:

$$\begin{aligned} x_{1t} &= (1-L)^{-d_3}\epsilon_{3t} + (1-L)^{-d_1}\epsilon_{1t} \\ x_{2t} &= (1-L)^{-d_3}\epsilon_{3t} + (1-L)^{-d_2}\epsilon_{2t} \end{aligned} \tag{2.11}$$

where L is the backshift operator. As $\lambda \to 0^+$, we have

$$\begin{aligned} f_1(\lambda) &\sim \frac{\sigma_1^2}{2\pi}\lambda^{-2d_1} \\ f_2(\lambda) &\sim \frac{\sigma_2^2}{2\pi}\lambda^{-2d_2} \\ f_{12}(\lambda) &\sim \frac{\sigma_3^2}{2\pi}\lambda^{-2d_3} \\ \rho(\lambda) &\sim \frac{\sigma_3^2}{\sigma_1\sigma_2}\lambda^{-2(d_3-\frac{1}{2}(d_1+d_2))} \end{aligned}$$

In this model, $\{x_{1t}\}$ and $\{x_{2t}\}$ are long-memory time series with a common component that has a smaller memory parameter than either of the individual time series. If we instead had $d_3 > \max(d_1, d_2)$, then the two time series would be cointegrated. The two time series are correlated in the "short run" (for frequencies away from zero), but the strength of the relationship decays to zero at frequency zero. This occurs because the common component has a smaller memory parameter and is dwarfed by the more persistent idiosyncratic components at low frequencies. We therefore refer to the model as an *anti-cointegration* model. A simulated realization is shown as a time series in Figure 1. The long-term movements of the time series are not strongly related, since the levels drift separately with longer memory, but the short-term movements are related.

3 The averaged periodogram estimator

The averaged periodogram estimator estimates the integrated autospectrum in a neighborhood of frequency zero by averaging the periodogram near zero frequency. The APE is based on the periodogram matrix, defined as $I(\lambda_j) = J(\lambda_j)J(\lambda_j)^*$, where $\lambda_j = \frac{2\pi j}{n}$ $(j = 0, \ldots, n-1)$ is the j^{th} Fourier frequency, $J(\lambda_j) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} X_t e^{i\lambda_j t}$ is the discrete Fourier transform of the bivariate series and n is the sample size. We will assume that d_1, d_2 are known to lie in the interval $(-s-\frac{1}{2}, \frac{1}{2})$ where s is a non-negative integer; this may occur because the series have been differenced s times to remove possible non-stationarity and/or a polynomial trend. When the memory parameter is less than $-\frac{1}{2}$, the raw periodogram or cross-periodogram is not a good estimator of the spectral density because of leakage; thus, authors such as Velasco [1999], Hurvich and Chen [2000], and Hurvich et al. [2002] recommend tapering the series before computing the cross-periodogram. Our results assume that the taper of Hurvich and Chen [2000] is used. Using the notation of Hurvich et al. [2002], the first-order taper is given by:

$$h_t = \left(1 - e^{i2\pi t/n}\right), \quad t = 1, \ldots, n\ . \tag{3.1}$$

If $d_1, d_2 \in \left(-\frac{1}{2} - s, \frac{1}{2}\right)$, the order-$s$ taper h_t^s is used, where h_t^s is h_t raised to the power s. Then, the tapered discrete Fourier transform is $J(\lambda_j) = \frac{1}{\sqrt{2\pi n a_s}}\sum_{t=1}^{n} h_t^s X_t e^{it\lambda_j}$, with $a_s = \binom{2s}{s} = \frac{1}{n}\sum_{t=1}^{n}|h_t|^{2s}$, and the tapered periodogram is $I(\lambda_j) = J(\lambda_j)J(\lambda_j)^*$. This tapering reduces

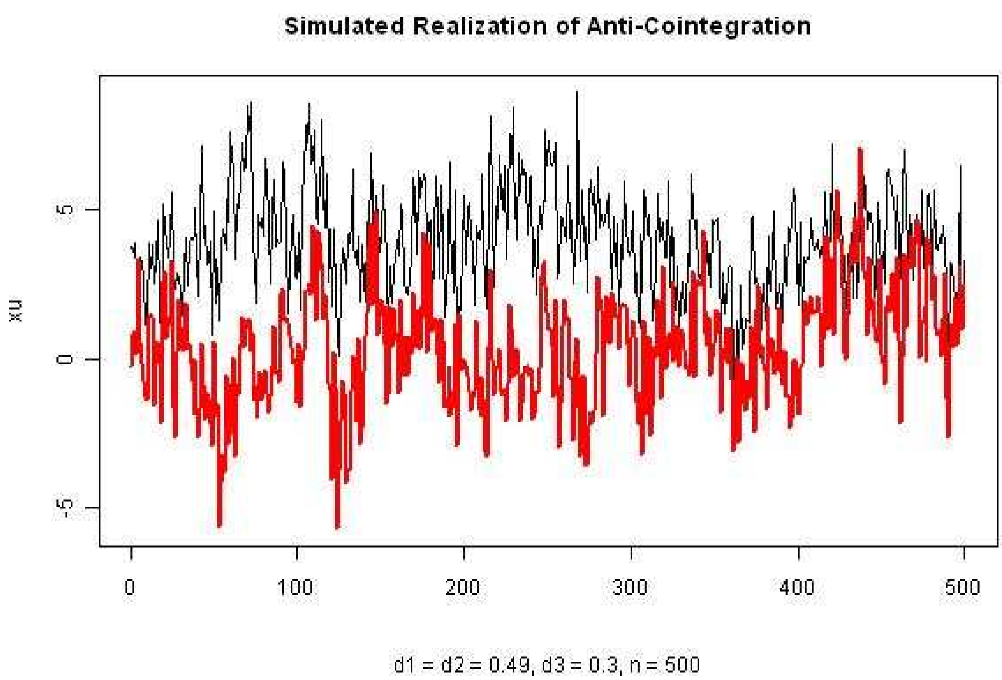

Figure 1: One simulated realization ($n = 500$) of the power law coherency model (2.11), with $d_1 = d_2 = 0.49$, and $d_3 = 0.3$.

the leakage at the low frequencies and therefore reduces the bias of estimators based on the periodogram.

In view of (2.9), it may seem reasonable to try to estimate d_{12} using linear regression of $\{\log|I_{12}(\lambda_j)|\}_{j=1}^m$ on $\{-\frac{1}{2}\log\lambda_j\}_{j=1}^m$ with $m^{-1}+m/n\to 0$, where I_{12}, the $(1,2)$ entry of I, is the cross-periodogram. One could also consider related estimators based on the log moduli of (say, contiguous) averages of the cross periodogram $\sum_{\ell=0}^L I_{12}(\lambda_{j+\ell})$ where $L\geq 0$ is a fixed integer. However, we believe that such estimators will be inconsistent for d_{12}, converging instead to $(d_1+d_2)/2$. To provide some intuition for this claim, we focus temporarily (in this paragraph only) on the anti-cointegration model, we set $s=0$, and we consider the Bartlett approximation to the periodogram matrix, $M=\Psi(\lambda)I_\epsilon\Psi^*(\lambda)$ where I_ϵ is the periodogram of $\{\epsilon_t\}_{t=1}^n$. A straightforward calculation reveals that

$$|\lambda_j|^{d_1+d_2}\left|\sum_{\ell=0}^L M_{12}(\lambda_{j+\ell})\right| = \left|\sum_{\ell=0}^L J_{\epsilon_1}(\lambda_{j+\ell})\overline{J_{\epsilon_2}(\lambda_{j+\ell})}\left(\frac{j}{j+\ell}\right)^{d_1+d_2}\right| + o_p(1) \ .$$

Thus,

$$\log\left|\sum_{\ell=0}^L M_{12}(\lambda_{j+\ell})\right| = -(d_1+d_2)\log|\lambda_j| + \eta_j$$

where the η_j are $O_p(1)$ and weakly dependent. This suggests that a regression of $\{\log|\sum_{\ell=0}^L I_{12}(\lambda_{j+\ell})|\}_{j=1}^m$ on $\{-\frac{1}{2}\log\lambda_j\}_{j=1}^m$ would converge to $(d_1+d_2)/2$ and not to d_{12}. A small simulation study for $L=0, L=1$ supports this conclusion. We considered 250 replications of the anti-cointegration model with $d_1=d_2=0.2$, $d_3=d_{12}=0$, $m=n^{0.5}$. For $n=512$, the regression estimator of d_{12} described above yielded a mean value (t-statistic) of 0.124 ($t=15.32$) when $L=0$ and 0.123 ($t=15.19$) when $L=2$. For $n=8192$ the mean value and t-statistic, both for $L=1$ and $L=2$, were 0.161 ($t=47.14$). These results suggest that the estimator is not approaching the true value $d_{12}=0$, and are consistent with the claim that the estimator is instead approaching $(d_1+d_2)/2=0.2$.

Another possibility would be to construct a version of the Gaussian Semiparametric estimator to estimate d_{12}. The argument presented above, however, suggests that once again such an estimator would be inconsistent, unless it explicitly accounts for power law behavior in all $2p$ entries of Ψ, and also accounts for the p^2 entries of Σ. Such an estimator would be cumbersome in practice, and would also require choice of the unknown p.

In the sequel, we adapt the averaged periodogram estimator (APE) of Robinson [1994], which he applied in the univariate case when $0<d<\frac{1}{2}$. In the univariate case, the APE is given by:

$$\hat{d} \quad = \quad \frac{1}{2} - \frac{\log(\hat{F}(q\lambda_m)/\hat{F}(\lambda_m))}{2\log q}$$

where $\hat{F}(\lambda) = \frac{2\pi}{n}\sum_{j=1}^{[n\lambda/2\pi]} I(\lambda_j)$, m is a bandwidth (depending on n) in $\{1,\cdots,\lfloor n/2\rfloor\}$ and $q\in(0,1)$ is fixed. Lobato and Robinson [1996] derived additional results about the limiting distribution of a suitably standardized version of $\hat{d}$ under various conditions, showing that it is normal when $0 < d < 1/4$ and non-normal for $1/4 < d < 1/2$. Lobato [1997] applied the averaged periodogram estimator to estimating d_1, d_2 in the multivariate case. While he did note (on page 139) the possibility of power law coherency, he was focused on estimating the memory parameters of the auto-spectra, not a power law in the coherency in a neighborhood of 0; in his Condition C1, he required that $\rho(0) > 0$. In contrast, we are particularly interested in estimating the power law of the cross-spectral density. Thus, we extend the consistency results of Robinson [1994] and Lobato [1997] in two ways: to the case in which $d < 0$ and to the estimation of d_{12}. The following lemma provides an expression for the integrated cross (or auto) spectral density.

Lemma 1 *Define* $F_{ab}(\lambda) = \int_0^\lambda f_{ab}(\theta)d\theta$ *for* $\lambda \in (0, \pi]$. *Under Assumptions 1-5,*

$$F_{12}(\lambda) \sim \frac{1}{1-2d_{12}} s(0;\tilde{q}_0)\lambda^{1-2d_{12}} \quad , \quad \lambda \to 0^+,$$

and for $j = 1, 2$,

$$F_{jj}(\lambda) \sim \frac{1}{1-2d_j} C_j \lambda^{1-2d_j} \quad , \quad \lambda \to 0^+.$$

In order to prove some results about the averaged periodogram estimator (with or without tapering), we make use of an additional global assumption on the product $\tau_{jk}(\lambda)e^{i\varphi_{jk}(\lambda)}$, based on Definition 2 of Hurvich et al. [2002, page 316]. From the discussion provided there, it follows that this condition would hold, for example, in the power law coherency example presented above .

Definition 1 *For some* $\mu > 1, \gamma \in (1, 2]$, *let* $\mathcal{L}^*(\mu, \gamma)$ *be the set of functions* $u(\lambda)$, *continuous on* $[0, \pi]$ *and differentiable on* $(0, \pi]$, *such that for all* $0 < |x|, |y| < \pi$,

$$\begin{aligned}
\frac{\max_{0\le z\le\pi}|u(z)|}{\min_{0\le z\le\pi}|u(z)|} &\le \mu \\
\frac{|u(x)-u(y)|}{\min_{0\le z\le\pi}|u(z)|} &\le \mu\frac{|y-x|}{\min(|x|,|y|)} \\
\frac{|u'(x)-u'(y)|}{\min_{0\le z\le\pi}|u(z)|} &\le \mu\frac{|y-x|^{\gamma-1}}{[\min(|x|,|y|)]^{\gamma}}
\end{aligned}$$

Assumption 6 *For all* j, k, *either* $\tau_{jk}(\lambda) = 0$ *for all* $\lambda \in [0, \pi]$, *or* $\tau_{jk}(\lambda)e^{i\varphi_{jk}(\lambda)} \in \mathcal{L}^*(\mu, \gamma)$ *for some* $\mu > 1, \gamma \in (1, 2]$.

The next assumption requires that the data have been tapered to a sufficiently high order.

Assumption 7 *$d_j, \delta_{jk} \in \left(-\frac{1}{2} - s, \frac{1}{2}\right)$, for some non-negative integer, s, and for all $j \in \{1, 2\}$, $k \in \{1, \cdots, p\}$, and the data are tapered of order s.*

Finally, we will consider consistency results under two different assumptions about the growth of the number of frequencies used in estimation as the sample size grows. The first is standard [for example, Robinson, 1994, Condition B] and is sufficient for the estimation of the memory parameters of the auto-spectra. The second is less common.

Assumption 8

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

as $n \to \infty$.

Assumption 9 *As $n \to \infty$:*

- *If $1 - \frac{\gamma}{2} < d_1 + d_2 < 1$,*

$$\frac{n^{\frac{-2d_\rho}{1-2d_{12}}}}{m} \to 0$$

- *If $d_1 + d_2 \le 1 - \frac{\gamma}{2}$,*

$$\frac{n^{\frac{-2d_\rho}{\gamma/2-2d_\rho}}}{m} \to 0.$$

Notice that the growth rates above change continuously, since $\frac{-2d_\rho}{1-2d_{12}} = \frac{2d_\rho}{2d_\rho-\gamma/2}$ when $d_1 + d_2 = 1 - \frac{\gamma}{2}$. The required growth rates (as a power of n) for three choices of $d_1 + d_2$ are shown in Figure 2. Unlike most assumptions on the growth rate of m in the context of long memory, Assumption 9 requires a lower bound on the growth rate of m. (Hurvich et al. [2005, Equation (3.9)] is one other paper that requires a lower bound on the growth rate of m.) In practice, this assumption is likely to be problematic because it depends on the unknown d_{12} and d_ρ, two of the very quantities that we wish to estimate. Larger growth rates of m are generally associated with increased finite-sample bias in estimation. Furthermore, theorems establishing limiting normality of the estimated memory parameter generally require that the growth rate of m be bounded above, with a tighter bound when the spectral density is less smooth. [See, for example, Lobato and Robinson, 1996, Condition C3.] These opposing requirements are likely to cause problems for the averaged periodogram estimator for the cross-spectral density when d_ρ is very negative.

In the case $a \neq b$, Assumption 9 is needed for the Bartlett-type approximation (3.4), while Assumptions 10–12 below are needed for sums of the scaled innovation periodogram matrix to be good approximations to corresponding sums of the cross spectral density in (3.5). Once again, Assumptions 10–12 impose lower bounds on the growth rate of m.

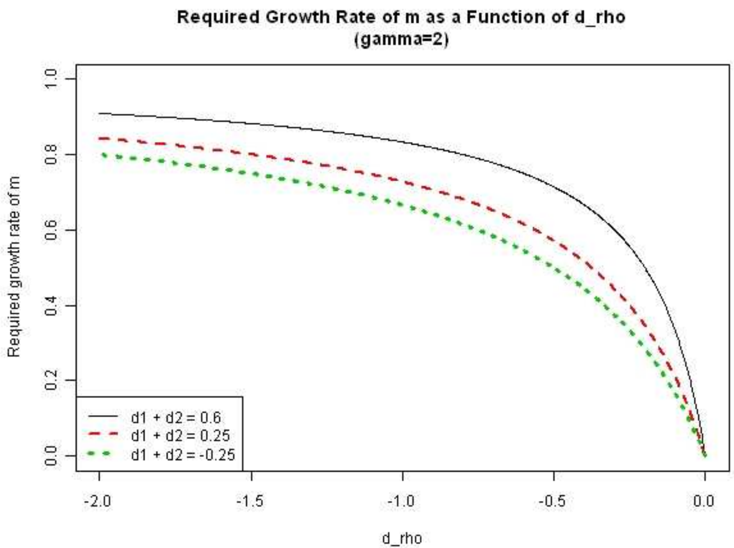

Figure 2: Minimum growth rate of m required by Assumption 9 as a function of d_ρ.

Assumption 10 *If $-2d_{11} - 2d_{22} < -1$, as $n \to \infty$:*

$$\frac{n^{\frac{d_{12}-2d_\rho}{1-d_{12}}}}{m} \to 0 \ .$$

Assumption 11 *If $-2d_{11} - 2d_{22} = -1$, as $n \to \infty$:*

$$(\log(m))^{\frac{1}{1-d_{12}}} \ \frac{n^{\frac{d_{12}-2d_\rho}{1-d_{12}}}}{m} \to 0 \ .$$

Assumption 12 *As $n \to \infty$:*

$$\frac{n^{\frac{-4d_\rho}{1-4d_\rho}}}{m} \to 0 \ .$$

Theorem 2 *If Assumptions 1–8 hold, for $a, b \in \{1, 2\}$,*

$$\hat{F}_{ab}(\lambda_m) - F_{ab}(\lambda_m) \quad = \quad o_p(\lambda_m^{1-d_a-d_b}). \tag{3.2}$$

If $a \neq b$, under the assumptions above together with Assumptions 9–12, we also have:

$$\hat{F}_{ab}(\lambda_m) - F_{ab}(\lambda_m) \quad = \quad o_p(\lambda_m^{1-2d_{ab}}). \tag{3.3}$$

When $a = b$, the theorem extends the result of Robinson [1994] to the case where $d_a < 0$ and tapering may be used. In that case, the second part of the theorem is irrelevant and no lower bound on m is necessary beyond the condition that $m \to \infty$. When the theorem is applied to the estimation of the cross-spectrum, power law coherency will affect the choice of m, with a more negative d_ρ placing a more stringent requirement on m.

Proof. As in Chen and Hurvich [2006], define $\tilde{j} = j + \frac{s}{2}$ to be the shifted Fourier frequency. Define $I_\epsilon(\lambda)$ to be the tapered periodogram of ϵ_t, with (a,b) element $I_{\epsilon,ab}(\lambda)$. Let $\Psi_a(\lambda)$ be the a^{th} row of $\Psi(\lambda)$. Generalizing the proofs of Robinson [1994, Theorem 1] and Lobato [1997, Theorem 1], we decompose the difference between the estimated averaged periodogram and the true averaged periodogram as:

$$\begin{aligned}
\hat{F}_{ab}(\lambda_m) - F_{ab}(\lambda_m) &= \frac{2\pi}{n}\sum_{j=1}^{m}\Big(I_{ab}(\lambda_j) - \Psi_a(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_b^*(\lambda_{\tilde{j}}))\Big) &(3.4)\\
&+\frac{2\pi}{n}\sum_{j=1}^{m}\Big(\Psi_a(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_b^*(\lambda_{\tilde{j}}) - f_{ab}(\lambda_j)\Big) &(3.5)\\
&+\frac{2\pi}{n}\sum_{j=1}^{m} f_{ab}(\lambda_j) - F_{ab}(\lambda_m) &(3.6)
\end{aligned}$$

Lemma 6 shows that the first term is $o_p(\lambda_m^{1-d_{aa}-d_{bb}})$ under Assumptions 1-8 and $o_p(\lambda_m^{1-2d_{ab}})$ if we also include Assumption 9. Lemma 8 shows that the second term is $o_p(\lambda_m^{1-d_{aa}-d_{bb}})$ under Assumptions 1–5, 7–8 and is $o_p(\lambda_m^{1-2d_{ab}})$ if we also include Assumptions 10, 11 and 12. Lemma 9 shows that the last term is $o(\lambda_m^{1-2d_{ab}})$ and therefore $o_p(\lambda_m^{1-d_{aa}-d_{bb}})$. ■

The result in Equation (3.3) can be used directly in Theorem 3 of Robinson [1994] to show that the averaged periodogram estimator is consistent for the memory parameter of the cross-spectral density under Assumption 9. We estimate the memory parameter of the cross-spectral density by using the fact from Lemma 1 that

$$F_{12}(\lambda) \sim \frac{1}{1-2d_{12}} s(0;\tilde{q}_0)\lambda^{1-2d_{12}} \;\;,\;\; \lambda \to 0^+.$$

Because $s(0;\tilde{q}_0)$ is complex, there are two possible ways to estimate d_{12} based on $F_{12}(\lambda)$. First, one could apply Robinson's theorem to the modulus of $F_{12}(\lambda)$. Second, one could apply the theorem separately for the real and imaginary parts of $F_{12}(\lambda)$ and take a weighted average based on $\varphi_{ab}(0)$. We recommend using the modulus; simulations have shown that it performs somewhat better than taking an average, even in the infeasible case where the optimal weights are known.

Theorem 3 *Define, for a fixed $q \in (0,1)$, and for $a, b \in \{1,2\}$,*

$$\hat{d}_{ab} \quad = \quad \frac{1}{2} - \frac{\log\left(|\hat{F}_{ab}(q\lambda_m)|/|\hat{F}_{ab}(\lambda_m)|\right)}{2\log q}.$$

For $a = b$, if Assumptions 1–8 hold,

$$\hat{d}_{aa}(\lambda_m) - d_{aa} \quad = \quad o_p(1). \tag{3.7}$$

For $a \neq b$, under the assumptions above together with Assumptions 9–12,

$$\hat{d}_{ab}(\lambda_m) - d_{ab} \quad = \quad o_p(1). \tag{3.8}$$

We may then estimate $\hat{d}_\rho = \hat{d}_{12} - \frac{1}{2}(\hat{d}_1 + \hat{d}_2)$, where $\hat{d}_{12}, \hat{d}_1, \hat{d}_2$ are estimated using the averaged periodogram estimator; $\hat{d}_\rho$ is consistent. Alternatively, $\hat{d}_1, \hat{d}_2$ could be estimated with another estimator that is consistent for univariate memory parameters. However, the convergence rate of $\hat{d}_\rho$ will generally depend on the worst convergence rate of the three estimators. To be precise, suppose that $\hat{d}_{12} = O_p(n^{-\alpha_{12}})$ and $\hat{d}_j = O_p(n^{-\alpha_j})$; then, $\hat{d}_\rho = O_p(n^{-\min(\alpha_{12},\alpha_1,\alpha_2)})$. As we will see in simulations in the next section and in Section 5, $\hat{d}_{12}$ and $\hat{d}_\rho$ are extremely variable in sample sizes typically used in practice.

Proof of Theorem 3. We have

$$\hat{d}_{ab} = \frac{1}{2} - \frac{1}{2\log q}\left(\log\left|\frac{\hat{F}_{ab}(q\lambda_m)}{F_{ab}(q\lambda_m)}\right| - \log\left|\frac{\hat{F}_{ab}(\lambda_m)}{F_{ab}(\lambda_m)}\right| + \log\left|\frac{F_{ab}(q\lambda_m)}{F_{ab}(\lambda_m)}\right|\right). \tag{3.9}$$

We have $\log\left|\frac{\hat{F}_{ab}(q\lambda_m)}{F_{ab}(q\lambda_m)}\right| = o_p(1)$ and $\log\left|\frac{\hat{F}_{ab}(\lambda_m)}{F_{ab}(\lambda_m)}\right| = o_p(1)$. For $a = b$, these results follow from (3.2) and the second part of Lemma 1, and for $a = 1$, $b = 2$ (under Assumptions 9–12) they follow from (3.3) and the first part of Lemma 1. Lemma 1 implies that $F_{ab}(q\lambda_m)/F_{ab}(\lambda_m) \sim q^{1-2d_{ab}}$ as $n \to \infty$ (making use of Assumptions 9–12 in the case $a \neq b$), so that

$$\log\left|\frac{F_{ab}(q\lambda_m)}{F_{ab}(\lambda_m)}\right| \to (1 - 2d_{ab})\log q.$$

Thus, from (3.9),

$$\begin{aligned} \hat{d}_{ab} \quad &= \quad \frac{1}{2} - \frac{1}{2\log q}\left[(1 - 2d_{ab})\log q + o_p(1)\right] \\ &= \quad d_{ab} + o_p(1). \end{aligned}$$

■

We next obtain rates of convergence and a central limit theorem for a Taylor series approximation to $\hat{d}_{12} - d_{12}$, under assumptions that include Assumptions 1− 12. Suppose that d_{12} is not equal to $-1/2$, and define

$$G(\lambda) = \frac{1}{1 - 2d_{12}}\, s(0; \tilde{q}_0)\, \lambda^{1-2d_{12}}$$

and

$$\tilde{R}(\lambda) = \left| \frac{\hat{F}_{12}(\lambda)}{G(\lambda)} \right|^2 - 1 \ .$$

Then

$$\hat{d}_{12} = \frac{1}{2} - \frac{\log\left(|\hat{F}_{12}(q\lambda_m)| / |\hat{F}_{12}(\lambda_m)| \right)}{2 \log q}$$

and

$$\hat{d}_{12} - d_{12} = -\frac{\log \left| \frac{\hat{F}_{12}(q\lambda_m)}{G(q\lambda_m)} \right|^2 - \log \left| \frac{\hat{F}_{12}(\lambda_m)}{G(\lambda_m)} \right|^2}{4 \log q} \ .$$

Arguing as in Lobato and Robinson [1996], pp. 307−308, we can approximate $\hat{d}_{12} - d_{12}$ by

$$\Delta_{m,n} = \frac{\tilde{R}(\lambda_m) - \tilde{R}(q\lambda_m)}{4 \log q} \ .$$

We will consider bounds on the rate of convergence of $\Delta_{m,n}$ to zero, and obtain a central limit theorem for a standardized version of $\Delta_{m,n}$. Since for any real A, B, C, D we have $|A + Bi|^2 - |C + Di|^2 = (A + C)(A - C) + (B + D)(B - D)$, it follows that

$$\begin{aligned} 4 \log q \, \Delta_{m,n} = & Re \left[\frac{\hat{F}_{12}(\lambda_m)}{G(\lambda_m)} + \frac{\hat{F}_{12}(q\lambda_m)}{G(q\lambda_m)} \right] Re \left[\frac{\hat{F}_{12}(\lambda_m)}{G(\lambda_m)} - \frac{\hat{F}_{12}(q\lambda_m)}{G(q\lambda_m)} \right] \\ & + Im \left[\frac{\hat{F}_{12}(\lambda_m)}{G(\lambda_m)} + \frac{\hat{F}_{12}(q\lambda_m)}{G(q\lambda_m)} \right] Im \left[\frac{\hat{F}_{12}(\lambda_m)}{G(\lambda_m)} - \frac{\hat{F}_{12}(q\lambda_m)}{G(q\lambda_m)} \right] \ . \end{aligned} \tag{3.10}$$

Since by (3.3) $\hat{F}_{12}(\lambda_m)/G(\lambda_m) \xrightarrow{p} 1$ and $\hat{F}_{12}(q\lambda_m)/G(q\lambda_m) \xrightarrow{p} 1$ we conclude that $\Delta_{m,n} \xrightarrow{p} 0$ at the same rate as

$$\tilde{\Delta}_{m,n} = \frac{\hat{F}_{12}(\lambda_m)}{G(\lambda_m)} - \frac{\hat{F}_{12}(q\lambda_m)}{G(q\lambda_m)} \ .$$

We next decompose $\tilde{\Delta}_{m,n}$ as

$$\tilde{\Delta}_{m,n} = \frac{2\pi}{nG(\lambda_m)} \sum_{j=1}^{m} I_{12}(\lambda_j) - \frac{2\pi}{nG(q\lambda_m)} \sum_{j=1}^{\lfloor qm \rfloor} I_{12}(\lambda_j)$$

$$= \frac{2\pi}{nG(\lambda_m)} \sum_{j=1}^{m} \Psi_1(\lambda_{\tilde{j}}) I_\epsilon(\lambda_j) \Psi_2^*(\lambda_{\tilde{j}}) - \frac{2\pi}{nG(q\lambda_m)} \sum_{j=1}^{\lfloor qm \rfloor} \Psi_1(\lambda_{\tilde{j}}) I_\epsilon(\lambda_j) \Psi_2^*(\lambda_{\tilde{j}}) + R(m, n, q)$$

where

$$R(m, n, q) = \frac{2\pi}{nG(\lambda_m)} \sum_{j=1}^{m} R_j - \frac{2\pi}{nG(q\lambda_m)} \sum_{j=1}^{\lfloor qm \rfloor} R_j$$

and $R_j = I_{12}(\lambda_j) - \Psi_1(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_2^*(\lambda_{\tilde{j}})$. From (7.5) and (7.9) in the proof of Lemma 6 below, we have under Assumption 9,

$$R(m,n,q) \quad = \quad \begin{cases} O_p(n^{-2d_\rho}m^{2d_{12}-1}) & \text{if} \ \ d_{11}+d_{22} > 1-\gamma/2 \\ O_p(m^{-\gamma/2}\lambda_m^{2d_\rho}) & \text{if} \ \ d_{11}+d_{22} < 1-\gamma/2 \end{cases} \tag{3.11}$$

which tends to zero under Assumption 9.

Defining $S_{12}(\lambda_j) = \Psi_1(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_2^*(\lambda_{\tilde{j}}) - f_{12}(\lambda_j)$, we have

$$\begin{aligned} \tilde{\Delta}_{m,n} = \frac{2\pi}{nG(\lambda_m)}\sum_{j=1}^{m} f_{12}(\lambda_j) - \frac{2\pi}{nG(q\lambda_m)}\sum_{j=1}^{\lfloor qm \rfloor} f_{12}(\lambda_j) \\ +S(m,n,q) + R(m,n,q), \end{aligned}$$

where

$$\begin{aligned} S(m,n,q) &= \frac{2\pi}{nG(\lambda_m)}\sum_{j=1}^{m} S_{12}(\lambda_j) - \frac{2\pi}{nG(q\lambda_m)}\sum_{j=1}^{\lfloor qm \rfloor} S_{12}(\lambda_j) \\ &= \left(\frac{2\pi}{nG(\lambda_m)} - \frac{2\pi}{nG(q\lambda_m)}\right)\sum_{j=1}^{\lfloor qm \rfloor} S_{12}(\lambda_j) + \frac{2\pi}{nG(\lambda_m)}\sum_{j=\lfloor qm \rfloor+1}^{m} S_{12}(\lambda_j) \ . \end{aligned}$$

By Lemma 7 and the Cauchy-Schwarz inequality,

$$\begin{aligned} E|S(m,n,q)|^2 &\le C\left(\frac{2\pi}{nG(\lambda_m)} - \frac{2\pi}{nG(q\lambda_m)}\right)^2 \sum_{j=1}^{\lfloor qm \rfloor}\sum_{k=1}^{\lfloor qm \rfloor} |\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{11}+d_{22})}\left(\frac{1}{n} + \chi(|j-k| \le s)\right) \\ &+C\left(\frac{2\pi}{nG(\lambda_m)} - \frac{2\pi}{nG(q\lambda_m)}\right)^2 \sum_{j=\lfloor qm \rfloor+1}^{m}\sum_{k=\lfloor qm \rfloor+1}^{m} |\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{11}+d_{22})}\left(\frac{1}{n} + \chi(|j-k| \le s)\right) \ . \end{aligned}$$

Since $G(\lambda_m) = C\lambda_m^{1-2d_{12}}$ and $G(q\lambda_m) = Cq^{1-2d_{12}}\lambda_m^{1-2d_{12}}$, we obtain

$$\begin{aligned} E|S(m,n,q)|^2 &\le C\lambda_m^{4d_{12}-2}\frac{1}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m} |\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{11}+d_{22})}\left(\frac{1}{n} + \chi(|j-k| \le s)\right) \\ &= O(\lambda_m^{4d_{12}-2}\frac{1}{n}\lambda_m^{2-2d_{11}-2d_{12}}) + \begin{cases} O(n^{-4d_\rho}m^{4d_{12}-2}) & \text{if} \ \ -2d_{11}-2d_{22} < -1 \\ O(n^{-4d_\rho}m^{4d_{12}-2}\log m) & \text{if} \ \ -2d_{11}-2d_{22} = -1 \\ O(\lambda_m^{4d_{12}-2d_{11}-2d_{22}}m^{-1}) & \text{if} \ \ -2d_{11}-2d_{22} > -1 \end{cases} \end{aligned}$$

by Equations (7.11), (7.14), (7.16), and (7.17). Since $d_\rho = d_{12} - \frac{1}{2}(d_{11} + d_{22})$ we have $4d_{12} - 2d_{11} - 2d_{22} = 4d_\rho$ and therefore

$$E|S(m,n,q)|^2 = O(\frac{1}{n}\lambda_m^{4d_\rho}) + \begin{cases} O(n^{-4d_\rho} m^{4d_{12}-2}) & \text{if } \ -2d_{11} - 2d_{22} < -1 \\ O(n^{-4d_\rho} m^{4d_{12}-2} \log m) & \text{if } \ -2d_{11} - 2d_{22} = -1 \\ O(\lambda_m^{4d_\rho} m^{-1}) & \text{if } \ -2d_{11} - 2d_{22} > -1 \end{cases}$$

Note that the $O(n^{-4d_\rho} m^{4d_{12}-2})$ term above goes to zero under Assumption 9, since the condition in the assumption's second part implies that in its first part. Also, the $O(\lambda_m^{4d_\rho} m^{-1})$ term above goes to zero by Assumption 12. Thus, we have

$$\tilde{\Delta}_{m,n} = T(m,n,q) + S(m,n,q) + R(m,n,q)$$

where a bound on $E|S(m,n,q)|^2$ is given above, a bound for $R(m,n,q)$ is given by (3.11), and where the deterministic bias term is given by

$$T(m,n,q) = \frac{2\pi}{nG(\lambda_m)} \sum_{j=1}^{m} f_{12}(\lambda_j) - \frac{2\pi}{nG(q\lambda_m)} \sum_{j=1}^{\lfloor qm \rfloor} f_{12}(\lambda_j).$$

To study the deterministic bias term we will consider the following assumption.

Assumption 13 *There exist $\beta \in (0,2]$ and nonzero constants C_1, C_2 (possibly complex-valued) such that, as $\lambda \to 0^+$,*

$$f_{12}(\lambda) = |\lambda|^{-2d_{12}}(C_1 + C_2|\lambda|^\beta) + o(|\lambda|^{-2d_{12}+\beta}).$$

Under Assumptions 8 and 13,

$$T(m,n,q) \sim \frac{1-2d_{12}}{s(0;q_0)} \left[\frac{C_1\lambda_m^{1-2d_{12}} + C_2\lambda_m^{1-2d_{12}+\beta}}{\lambda_m^{1-2d_{12}}} - \frac{C_1\lambda_{qm}^{1-2d_{12}} + C_2\lambda_{qm}^{1-2d_{12}+\beta}}{(q\lambda_m)^{1-2d_{12}}} \right]$$

$$= \frac{C_2(1-2d_{12})}{s(0;q_0)}(1-q^\beta)\lambda_m^\beta \ .$$

The following lemma provides the foundation for a central limit theorem for a rescaled version of $\tilde{\Delta}_{m,n}$ in the case $-2d_{11} - 2d_{22} > -1$ (which implies that $d_{12} < 1/4$).

Lemma 4 *Suppose that $-2d_{11} - 2d_{22} > -1$, and that $\{\epsilon_t\}$ is Gaussian. Define $\gamma_{m,n} = m^{2d_\rho - 1/2} n^{-2d_\rho}$. Under Assumptions 1 – 8 and Assumption 12, for any fixed integer $r \geq 3$,*

$$cum(\underbrace{S(m,n,q)/\gamma_{m,n}, \ldots, S(m,n,q)/\gamma_{m,n}}_{r \text{ terms}}) \to 0 \ .$$

Note that $R(m,n,q)/\gamma_{m,n} = o_p(1)$ under Assumption 9, since from (3.11)

$$R(m,n,q)/\gamma_{m,n} = \begin{cases} O_p(m^{-1/2+d_{11}+d_{22}}) & \text{if } \ d_{11}+d_{22} > 1-\gamma/2 \\ O_p(m^{1/2-\gamma/2}) & \text{if } \ d_{11}+d_{22} < 1-\gamma/2 \end{cases}$$

In view of this and (3.10), we have therefore established the following.

Theorem 5 *Under the assumptions of Lemma 4 together with Assumption 9, if the three quantities*

$$E[ReS(m,n,q)/\gamma_{m,n}]^2 \ \ E[ImS(m,n,q)/\gamma_{m,n}]^2 \ \ E[ReS(m,n,q)ImS(m,n,q)/\gamma_{m,n}^2]$$

tend to finite constants, not all zero, then $\gamma_{m,n}^{-1}[\tilde{\Delta}_{m,n} - T(m,n,q)]$ *is asymptotically complex normal with zero mean, and* $\gamma_{m,n}^{-1}[4\log q\Delta_{m,n} - ReT(m,n,q)]$ *is asymptotically (real-valued) normal with zero mean.*

The following assumption is a necessary and sufficient condition for $\gamma_{m,n}^{-1}T(m,n,q) \to 0$.

Assumption 14 *Assumption 13 holds and*

$$\frac{m^{\frac{2\beta+1-4d_\rho}{2\beta-4d_\rho}}}{n} \to 0 \ .$$

Note that Assumption 14 is an upper bound on the growth rate of m. If Assumptions 1–9, 12–14 hold, $-2d_{11} - 2d_{22} > -1$, (so that Assumptions 10 and 11 are irrelevant here) and the three limits in Theorem 5 exist and are not all zero, then $\gamma_{m,n}^{-1}\tilde{\Delta}_{m,n}$ is asymptotically complex normal with zero mean, and $4\log q\gamma_{m,n}^{-1}\Delta_{m,n}$ is asymptotically (real-valued) normal with zero mean. We omit the details on the evaluation of the three limits in Theorem 5. The values of these limits would determine the asymptotic variances of the limiting normal distributions described above. For the anti-cointegration model with $s=0$, a long but straightforward calculation reveals that under Assumptions 1–9, 12–14 together with the assumption that $d_1+d_2 < 1/2$, the asymptotic variance for $4\log q\,\gamma_{m,n}^{-1}\Delta_{m,n}$ is

$$(1/2)(2\pi)^2\sigma_1^2\sigma_2^2(1-2d_3)^2\sigma_3^{-4}\left[1+\cos\pi(d_1+d_2)\right]\left[q^{1-2(d_1+d_2)}(1-q^{2d_3-1})^2 + 1 - q^{1-2(d_1+d_2)}\right] \ .$$

Finally, we comment on the compatibility of the upper bound in Assumption 14 and the lower bounds in Assumptions 9 and 12, when $d_{11}+d_{22} < 1/2$. We focus on the case $\beta = 2$, $\gamma = 2$, as holds in the anti-cointegration example. Since $\gamma = 2$, it is easily shown that Assumption 12 is always stronger than Assumption 9. Thus, m must satisfy the two requirements

$$\frac{n^{\frac{-4d_\rho}{1-4d_\rho}}}{m} \to 0 \ , \quad \frac{m^{\frac{5-4d_\rho}{4-4d_\rho}}}{n} \to 0 \ .$$

If we assume that $m = n^\alpha$ where $\alpha \in (0, 1)$ then the two requirements become $\alpha > -4d_\rho/(1-4d_\rho)$ and $\alpha < (4-4d_\rho)/(5-4d_\rho)$. These two requirements can always be simultaneously satisfied for some choice of α since it is easily shown that for any $d_\rho < 0$, $-4d_\rho/(1-4d_\rho) < (4-4d_\rho)/(5-4d_\rho)$.

4 Simulation results

We now assess the performance of the APE in finite samples through simulation. We will test the APE for d_{12} and d_ρ in two cases:

- Fractionally Integrated Vector Autoregression (FIVAR): A $FIVAR(0, (d_1, d_2))$ model with the innovation variance of the vector autoregression equal to $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.
- Power law coherency: The anti-cointegration model given in Equation (2.11) with $d_3 = d_2 - b$, where $b = 0.1$ or $b = 0.5$. In this case, $d_{12} = d_2 - b$ and $d_\rho = \frac{1}{2}(d_2 - d_1) - b$.

All simulations are based on the algorithm described by Sela and Hurvich [2009]. We allow d_1, d_2, and the number of observations to vary. We choose $m = n^g$, where $g \in \{1/6, 1/3, 1/2, 2/3, 4/5\}$. In all cases, we use $q = \frac{1}{2}$, since Lobato [1997] showed that this choice worked well for a variety of values of d_1, d_2. All results are based on 1000 replications. To facilitate calculations using the fast Fourier transform, we chose sample sizes to be powers of 2. In particular we used powers of $7, 9, 11, 13, 15$, corresponding to $n = 128$, $n = 512$, $n = 2048$, $n = 8192$ and $n = 32768$, respectively.

4.1 APE for power law in auto-spectrum when $d < 0$

First, we describe the performance of the APE for a power law in the auto-spectrum when $d < 0$. In Figure 3, we plot the estimated values of memory parameters as n grows. Each panel shows a different choice of m. Because the averaged periodogram estimator for the power law in the auto-spectrum is consistent for any growth rate of m as long as $\frac{1}{m} + \frac{m}{n} \to 0$, the APE improves with the sample size in all panels. The APE is more variable for smaller growth rates, as one might expect. However, larger growth rates of m can lead to finite-sample bias in the APE. To see an example of this, we apply the APE to the auto-periodogram of a process that consists

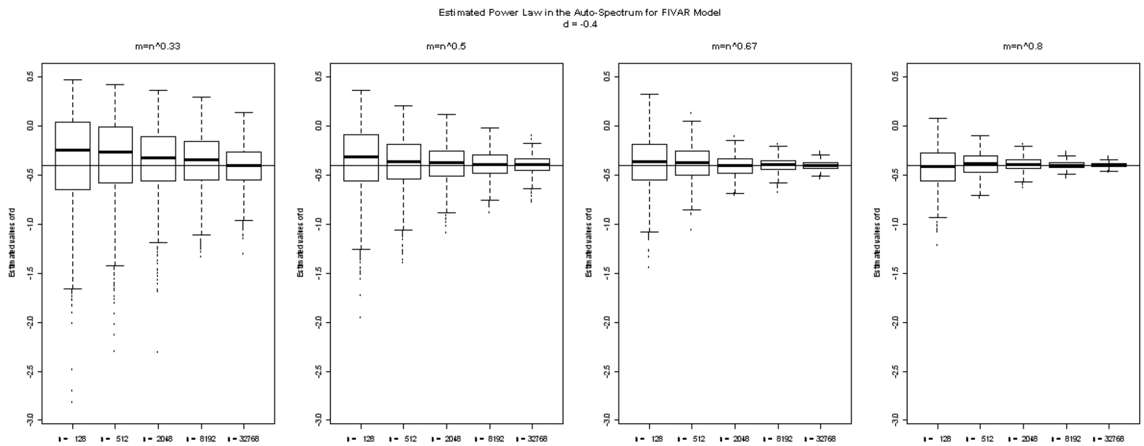

Figure 3: Estimated power law in the auto-spectrum when the true data-generating process is a FIVAR process with $d_1 = d_2 = -0.4$ and various values of n. In the four panels, m is equal to $n^{1/3}, n^{1/2}, n^{2/3}, n^{4/5}$.

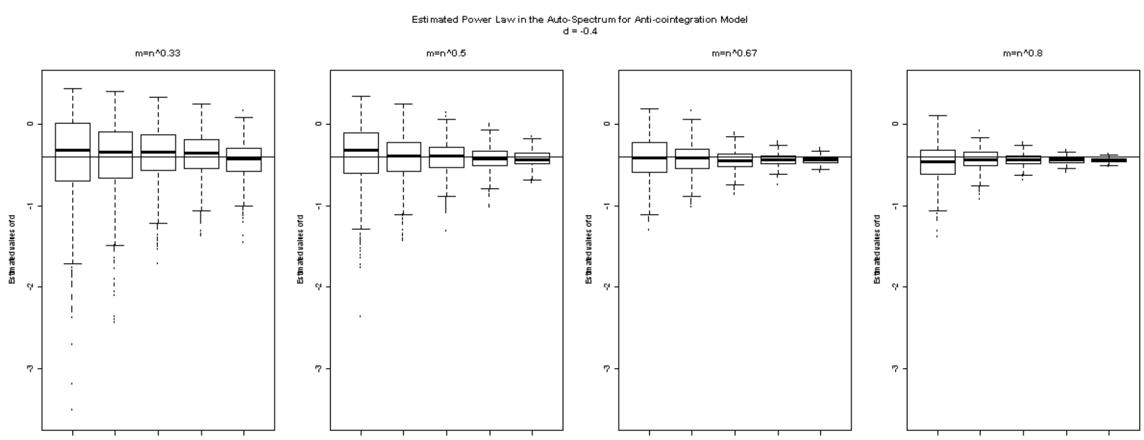

Figure 4: Estimated power law in the auto-spectrum when the true data-generating process is the power law coherency process with $d_1 = d_2 = -0.4$ and various values of n. In the four panels, m is equal to $n^{1/3}, n^{1/2}, n^{2/3}, n^{4/5}$.

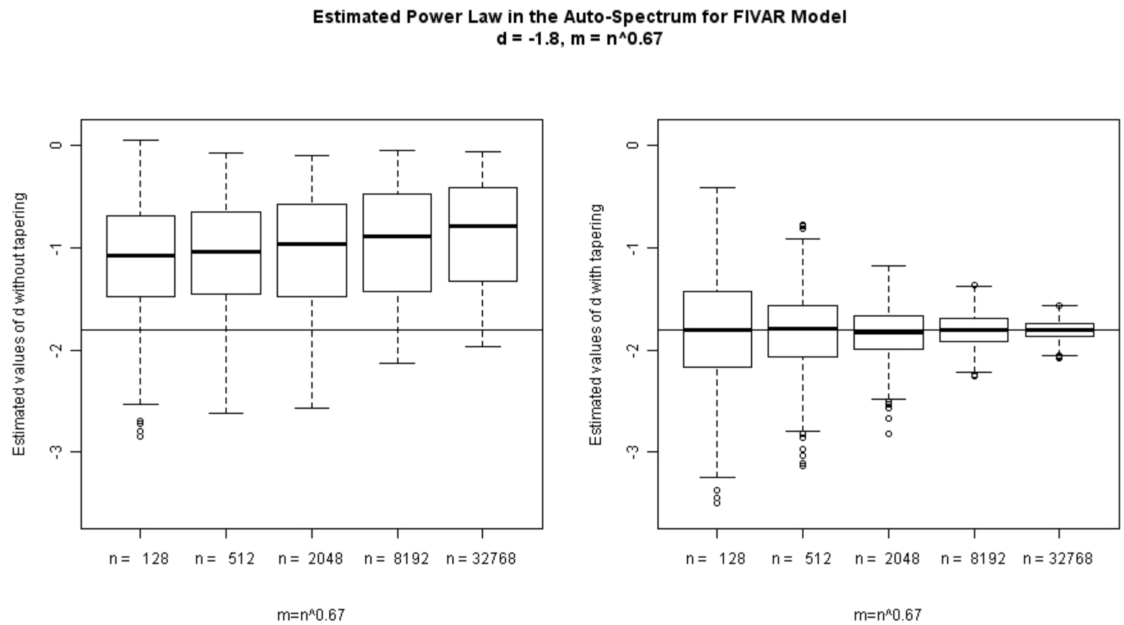

Figure 5: Estimated power law in the auto-spectrum when the true data-generating process is an $ARFIMA(0, -1.8, 0)$ process with varying n; $m = n^{2/3}$ in all cases. The data are not tapered in the left panel and are tapered in the right panel.

of two components. Figure 4 shows the finite-sample bias that can result when m grows too quickly.

Now, we consider a case where $d < -0.5$. In Figure 5, we plot the APE estimates of d when the true value is -1.8 as n grows. In the left panel, the data have not been tapered, and APE performs quite badly, appearing to be inconsistent. In the right panel, the data have been tapered of order 1, and the performance of the APE improves dramatically. This shows the importance of tapering when the series may be non-invertible.

4.2 Estimating a power law in the cross-spectrum

We now describe the performance of the APE for estimating a power law in the cross-spectral density. Here, we will focus on the case where $d_1 = 0.2, d_2 = 0$, and we use a taper with $s = 1$. In the case where we set $b = 0.1$ in a power law coherency model (which has $\gamma = 2$), $d_\rho = -0.2$, $d_{12} = -0.1$ and the lower bound on the growth rate required to ensure consistency is $\frac{-4d_\rho}{1-4d_\rho} = \frac{0.8}{1.8} = 0.444$, i.e., we require $m/n^{0.444} \to \infty$. When $b = 0.5$ and $d_\rho = -0.6$, so that $d_{12} = -0.5$, and the required growth rate is $\frac{-4d_\rho}{1-4d_\rho} = 2.4/3.4 = 0.7058$. Figure 6 presents

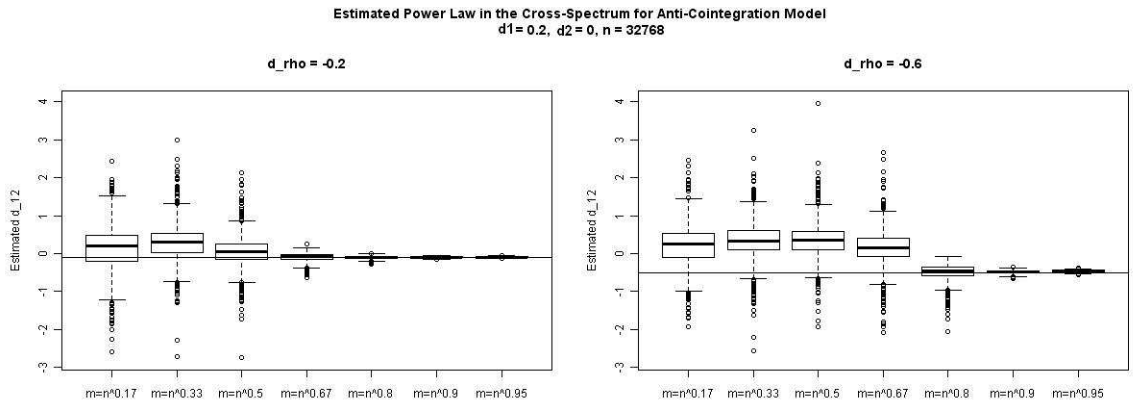

Figure 6: Estimated power law in the cross-spectral density, $\hat{d}_{12}$, using the averaged periodogram estimator when the true data-generating process has power law coherency, with $d_1 = 0.2, d_2 = 0, n = 32,768$ and varying choices of m. $d_\rho = -0.2$ in the left panel; $d_\rho = -0.6$ in the right panel. True values of d_{12} are -0.1 and -0.5, respectively.

boxplots of $\hat{d}_{12}$ in the anti-cointegration models when $n = 32,768$. Notice that the bias and variance of the estimators are much smaller for $m = n^{2/3}, n^{4/5}$ when $b = 0.1$ and for $m = n^{4/5}$ when $b = 0.5$, relative to smaller values of m. This occurs because the growth rates required by Assumption 12 differ in the two cases. The lower bound on the growth rate of m can also be seen by contrasting the two boxplots in Figure 7. The left boxplot shows a case in which m grows as $n^{1/2}$, which is less than the growth rate required for consistency; the bias and variability of $\hat{d}_{12}$ do not decrease with n. In contrast, the right boxplot shows that, when m grows more quickly (in this case, as $n^{4/5}$), the bias and the variance of the estimator decrease with n. Thus, the lower bound on the growth rate of m appears to be necessary for consistency.

Figure 8 present boxplots of the estimated values of d_ρ when $n = 32,768$, setting the estimated value to 0 when the $\hat{d}_{12} - \frac{1}{2}(\hat{d}_1 + \hat{d}_2) > 0$, since d_ρ cannot be positive. Figure 9 shows how $\hat{d}_\rho$ changes as n increases for different growth rates of m. As before, estimates of d_ρ based on $m = n^{1/2}$ have approximately constant bias and variability as n grows. When $m = n^{4/5}$, the variability and bias decrease as n increases, but the estimated values of d_ρ are biased upward and remain quite variable. This occurs because $\hat{d}_\rho$ depends on three different averaged periodogram estimators. Furthermore, since x_{1t} and x_{2t} contain two components with different memory parameters, the estimators of the memory parameters of the auto-spectra may be particularly badly behaved.

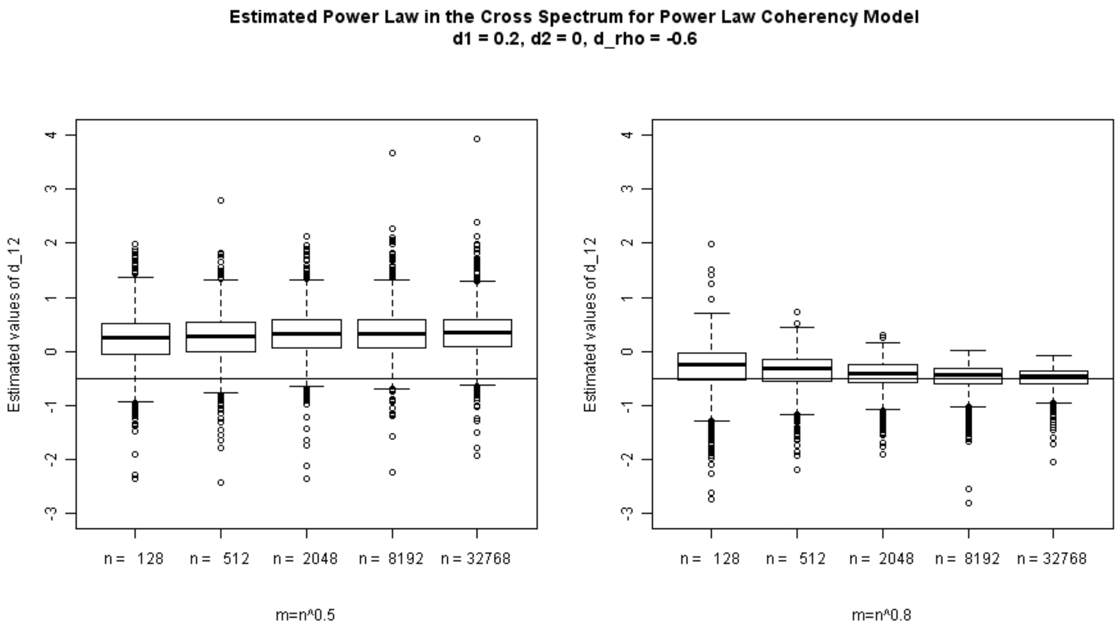

Figure 7: Estimated power law in the cross-spectrum when the true data-generating process has power law coherency with $d_1 = 0.2, d_2 = 0, d_\rho = -0.6$, and varying n. $m = n^{1/2}$ in the left panel and $m = n^{4/5}$ in the right panel.

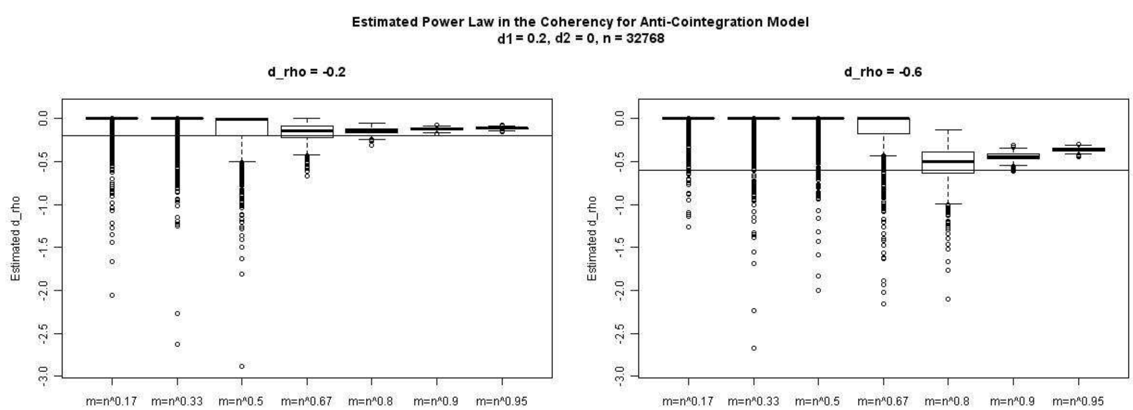

Figure 8: Estimated power law in the coherency using the averaged periodogram estimator when the true data-generating process has power law coherency with $d_1 = 0.2, d_2 = 0, n = 32,768$ and varying choices of m. $d_\rho = -0.2$ in the left panel; $d_\rho = -0.6$ in the right panel.

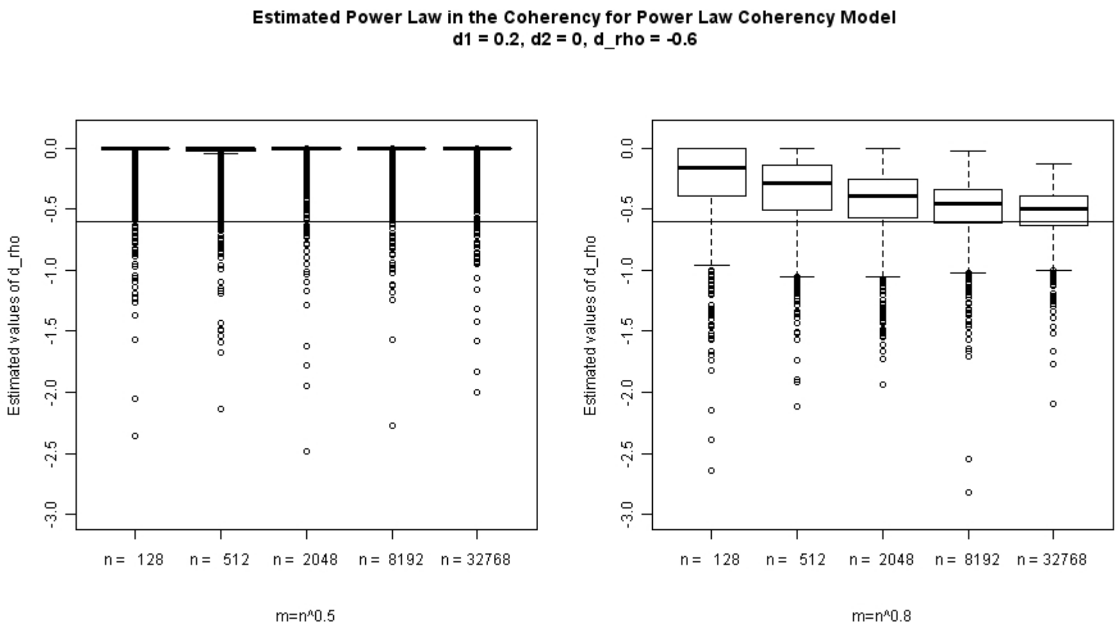

Figure 9: Estimated power law in the coherency when the true data-generating process has power law coherency with $d_1 = 0.2, d_2 = 0, d_\rho = -0.6$, and varying n. $m = n^{1/2}$ in the left panel and $m = n^{4/5}$ in the right panel.

5 Data analysis: Money supply growth

We examine monthly estimates of money stock from January 1959 to September 2009[1] (608 observations). We focus on two different supplies of money. M1 consists of easily accessible money, such as currency and demand deposits, while M2 consists of M1 together with forms of money that require more time to access, such as savings deposits and money market accounts. In order to remove the component common to M1 and M2, we focus on describing the relationship between M1 and M2 less M1. Both series have clear upward trends in their levels; we will work with the difference in logs, shown in Figure 10. The plot shows some common movements, such as a period of comovement in the late 1960's and mid-1970's. However, the long run movements are less related, with M1 growing faster in the mid-1980's and mid-1990's but M2 less M1 growing faster in other periods. The logarithms of the auto-periodograms (Figure 11) have an approximately linear relationship with the log frequency near frequency zero, suggesting that the individual series have long memory; it is unclear from this figure which series has a larger memory parameter. The GPH estimates suggest the presence of long memory in the individual series, though the estimates are quite sensitive to the number of frequencies used.

To estimate the coherency, we first smooth the periodogram, using the `spgram` function in R [R Development Core Team, 2008] with modified Daniell smoothers of widths $(21, 21)$. Though the smoothing is likely to be problematic close to the zero frequency because of the long memory, estimated coherency based on smoothing is consistent at frequencies away from singularities in the autospectra and cross-spectrum [Hidalgo, 1996]. The coherency of the two series is not significantly different from zero except at frequencies ranging from approximately $0.01(2\pi)$ to approximately $0.09(2\pi)$ (periods ranging from just under 1 year to just over 8 years). The coherency peaks just above 0.35 around frequency $0.06(2\pi)$, which corresponds to a period of just over 16 months. This suggests that only the longer run movements of M1 and M2-M1 (with periods greater than 1 year) are related. However, the coherency decreases toward zero at the zero frequency, suggesting power law coherency, so that very long-run movements are not related.

With this evidence for power law coherency, we can apply the averaged periodogram estimator for varying choices of m, holding q fixed at 0.5. The estimated values of the memory parameters of the individual series are in the range of the GPH estimates and vary less as m changes. However, the estimated power law in the cross-spectrum is always larger than the mean of the two auto-memory parameters, which appears to rule out power law coherency. As we saw in Section 4, identification of power law coherency is quite challenging even when $n = 8192$; in this case, $n = 608$. Thus, it is not clear that we can rule out power law coherency, even though $\hat{d}_\rho = 0$.

To provide further evidence that the averaged periodogram estimator may not be able to find evidence of a power law in the coherency in this dataset, we simulate 1000 datasets with

[1]Source: Federal Reserve, `http://www.federalreserve.gov/releases/h6/hist/h6hist1.txt`

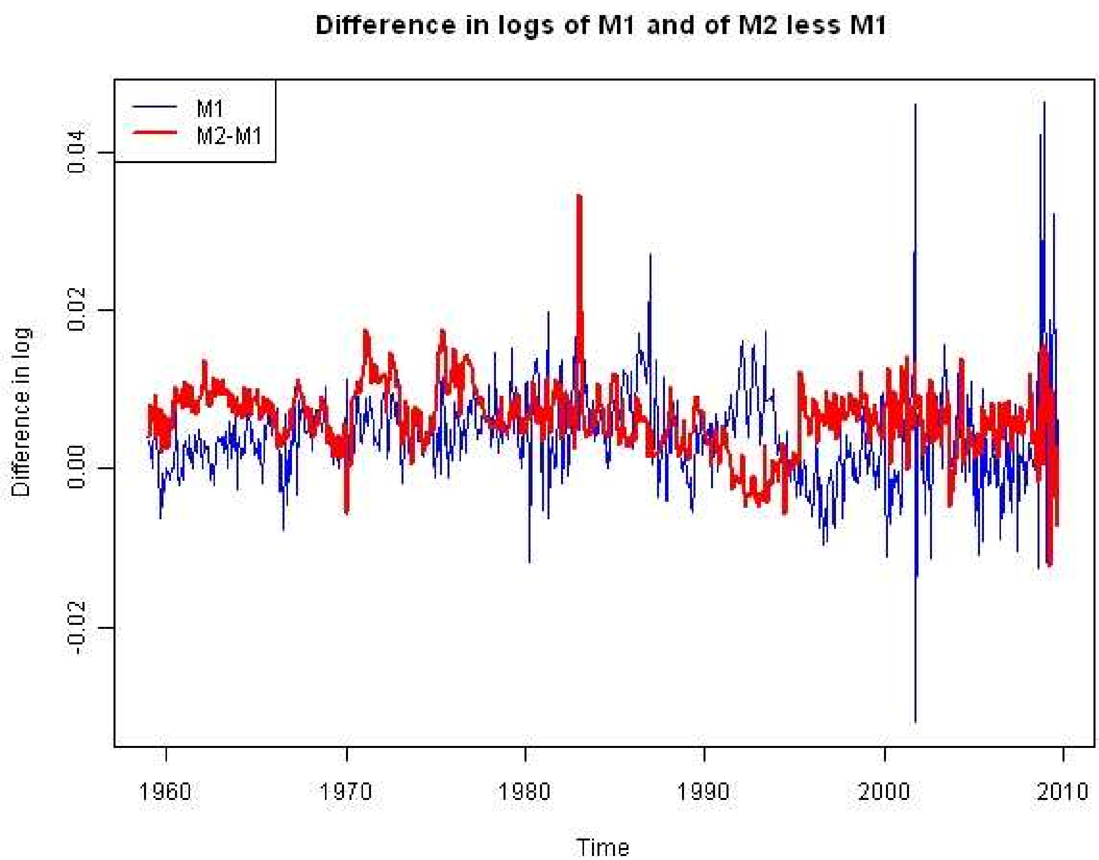

Figure 10: Time series of differences in logs of M1 and M2-M1.

m	$\hat{d}_{M1}$	$\hat{d}_{M2-M1}$	$\hat{d}_{M1,M2-M1}$
$n^{1/2} = 24$	0.347	0.375	0.409
$n^{3/5} = 46$	0.326	0.285	0.335
$n^{2/3} = 71$	0.335	0.300	0.425
$n^{3/4} = 122$	0.206	0.336	0.329
$n^{4/5} = 168$	0.269	0.379	0.358

Table 1: APE estimates of d_{M1}, d_{M2-M1}, and $d_{M1,M2-M1}$ for varying choices of m.

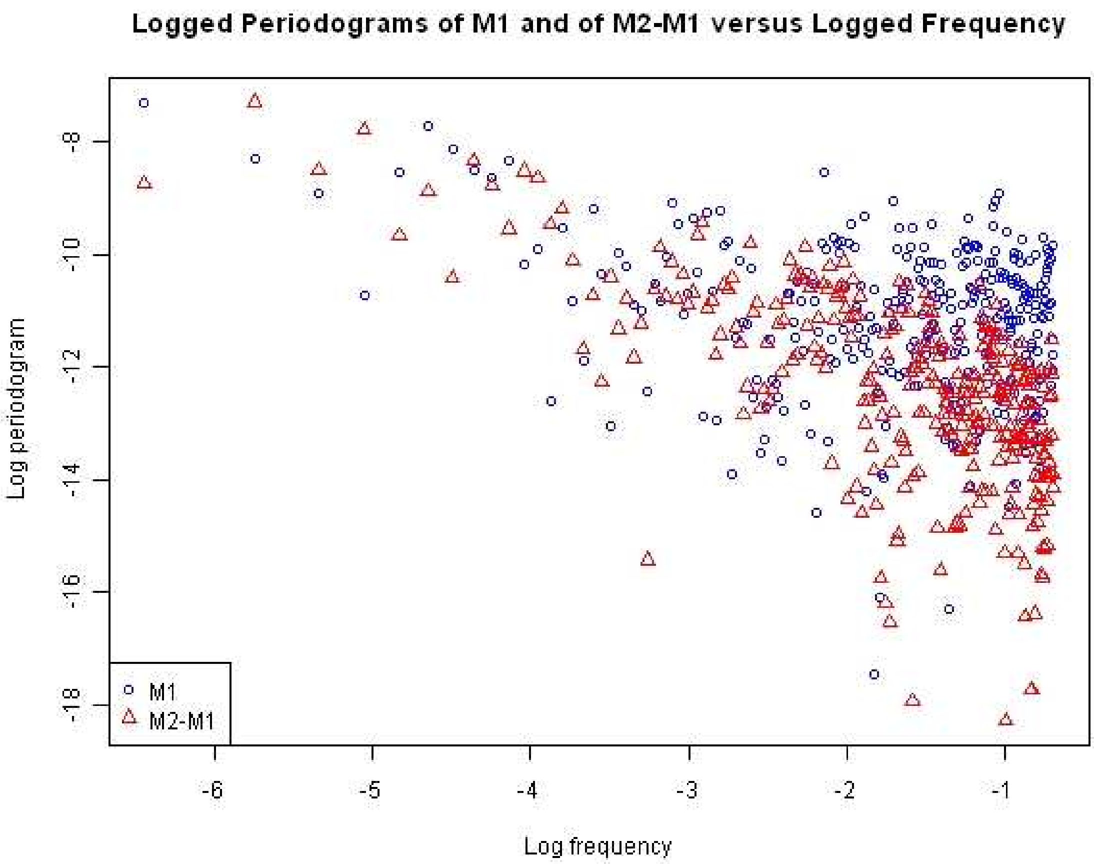

Figure 11: Log auto-periodograms of the differences in logs of M1 and M2-M1 versus the log frequency.

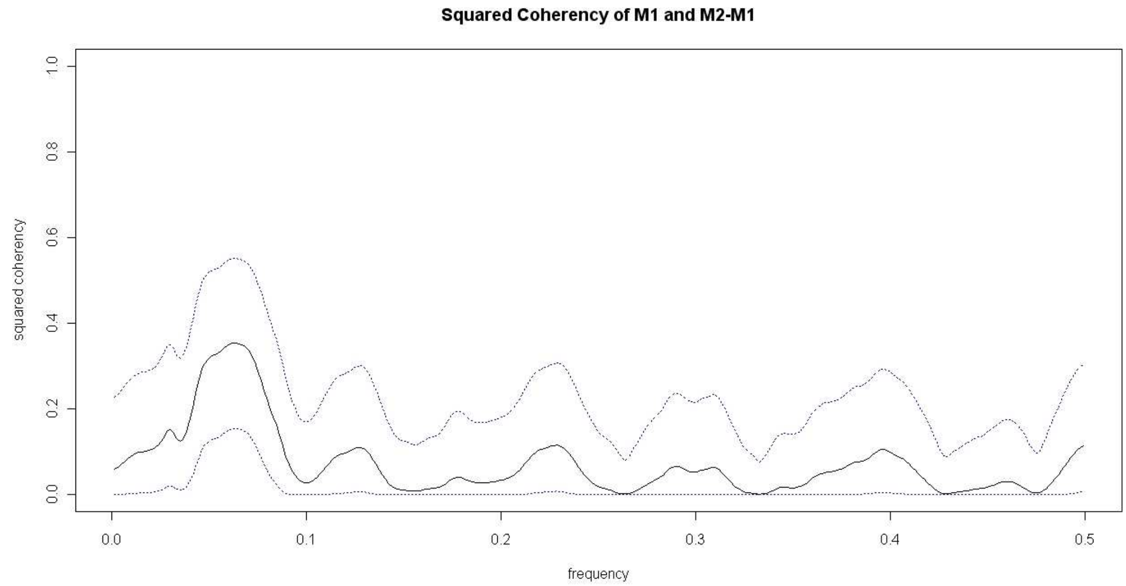

Figure 12: Estimated coherency of the differences in logs of M1 and M2-M1 (smoothed using $spans = (21, 21)$).

$n = 608, d_1 = 0.15, d_2 = 0.3$, and varying values of d_ρ, using the anti-cointegration model. For each dataset, we estimated d_ρ for the values of m used in Table 1. In Figure 13, we plot the estimated values of d_ρ for varying true values of d_ρ. The estimated values become more spread out for more negative values of d_ρ. In many cases with power law coherency, the proportion of the time that the point estimate is non-zero is under 0.5. In all cases, the estimates are biased upward, with the bias particularly pronounced for more negative values of d_ρ. This shows that the performance of the APE is quite poor in a sample of this size.

This dataset provides graphical evidence based on the smoothed periodogram that power law coherency may exist. However, the averaged periodogram estimator is not able to identify power law coherency in a sample of this size, as shown in simulation.

6 Conclusion

In this paper, we have discussed the possibility of a power law in coherency for bivariate long-memory time series, providing a parametric time-domain example. The average periodogram estimator provides a consistent estimator for the power law in the coherency, but can be quite variable in sample sizes considered in practice and requires that a sufficiently large number of

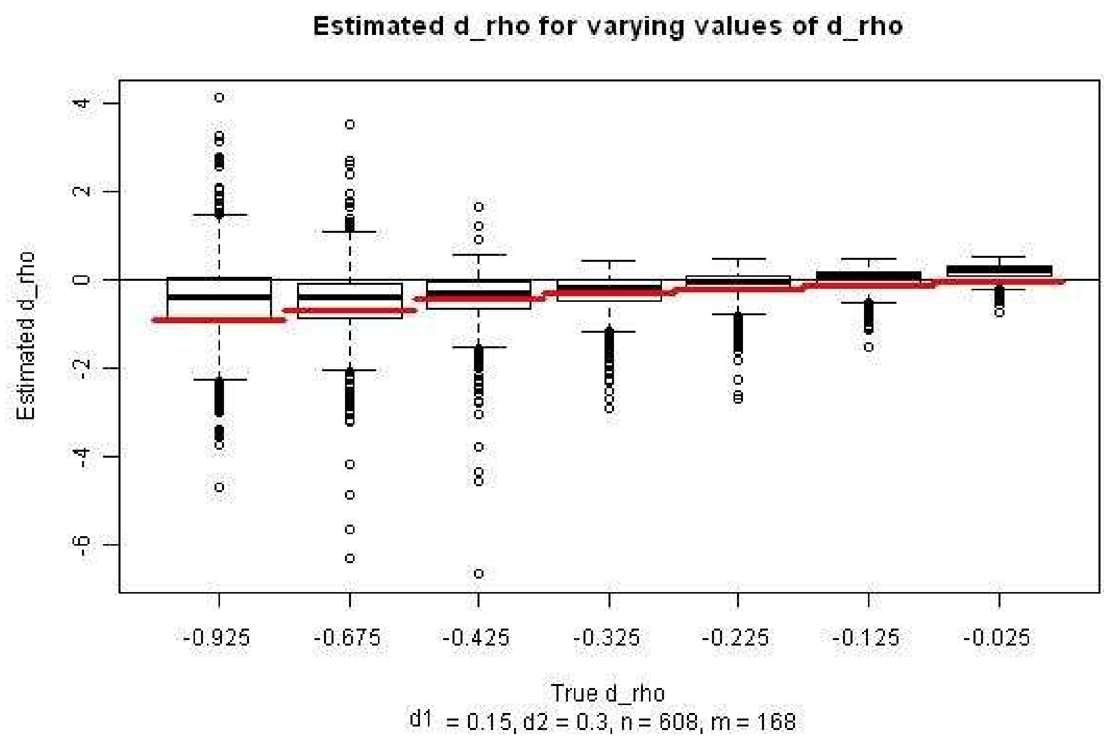

Figure 13: Estimated values of d_ρ in simulations for varying values of d_ρ.

frequencies be used. We have also proved that the APE can be applied to estimation of the power law in the autospectra when $d < 0$, provided that the data are tapered when necessary. We have applied our estimators to money stock measure data.

We have assumed a bivariate model for notational convenience. However, the multivariate case can be handled along the same lines. If the observed series $\{X_t\}$ is $q^* \times 1$ with $q^* \leq p$ then the Ψ matrices would be $q^* \times p$. The representation of the cross spectrum f_{ab} between components a and b of $\{X_t\}$ as a superposition of power law functions would be obtained just as in Section 2. The estimator of the power law, d_{ab}, would proceed by averaging the cross periodogram between components a and b. The proof of the properties of this estimator would proceed similarly to that given here for the bivariate case. In particular, no new difficulties would arise in the proof of Lemma 6 since the model of Chen and Hurvich [2006] is multivariate, and since the results used from Hurvich et al. [2002] are univariate. We thank an anonymous referee for motivating this discussion of the multivariate case.

The coherency can provide important insights into the relationships of two or more time series, which can help to understand the underlying mechanisms that generate them. Power laws in coherency arise naturally in the context of long-memory time series, and can affect the behavior of estimators of other quantities. Thus, the possibility of power law coherency should be considered in the analysis of bivariate or multivariate time series, and the APE provides a way to estimate the power law in the coherency in some cases.

7 Technical Lemmas

Proof of Lemma 1. By (2.6)− (2.8) and Assumption 5,

$$f_{12}(\lambda) = |2\sin(\lambda/2)|^{-2d_{12}} s(\lambda; \tilde{q}_0) + R(\lambda)$$

where $R(\lambda)$ is differentiable on $(0, \pi]$ and

$$\lim_{\lambda \to 0^+} R(\lambda)/\lambda^{-2d_{12}} = 0.$$

Define $s(\lambda) = f_{12}(\lambda)/|\lambda|^{-2d_{12}}$. Then

$$s(0) = \lim_{\lambda \to 0^+} s(\lambda) = s(0; \tilde{q}_0) \neq 0.$$

Using integration by parts, since $1 - 2d_{12} > 0$,

$$\begin{aligned} F_{12}(\lambda) &= \int_0^\lambda f_{12}(\theta) d\theta = \int_0^\lambda s(\theta)\theta^{-2d_{12}} d\theta \\ &= \frac{1}{1-2d_{12}} s(\lambda)\lambda^{1-2d_{12}} - \frac{1}{1-2d_{12}} \int_0^\lambda \theta^{1-2d_{12}} s'(\theta) d\theta. \end{aligned} \tag{7.1}$$

By Assumptions 2, 4 and 5, there exist positive constants C, A and Ξ such that $|s'(\theta)| \leq C\theta^{-1+\Xi}$ for all $\theta \in (0, A)$. It follows that for all $\lambda \in (0, A)$,

$$\frac{1}{\lambda^{1-2d_{12}}} \left| \int_0^\lambda \theta^{1-2d_{12}} s'(\theta) d\theta \right| \leq C\lambda^{\Xi}.$$

This, combined with (7.1) yields

$$F_{12}(\lambda) \sim \frac{1}{1-2d_{12}} s(0)\lambda^{1-2d_{12}} \quad , \quad \lambda \to 0^+.$$

The second part of the Lemma follows from Karamata's theorem, as (2.3) implies that $f_j(\lambda)$ is regularly varying at zero. ■

Lemma 6 *If $f_{ab}(\lambda)$ satisfies the conditions of Lemma 1, d_a, $d_b < 1/2$ for $a, b \in \{1, 2\}$ and Assumptions 1-8 hold,*

$$E\left(\frac{2\pi}{n}\sum_{j=1}^{m}\left|I(\lambda_j) - \Psi(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi^*(\lambda_{\tilde{j}})\right|\right) \quad = \quad o\left(\lambda_m^{1-d_{aa}-d_{bb}}\right)$$

If we further assume that Assumption 9 holds and $a \neq b$,

$$\left|E\left(\frac{2\pi}{n}\sum_{j=1}^{m}\left(I_{ab}(\lambda_j) - \Psi_a(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_b^*(\lambda_{\tilde{j}})\right)\right)\right| \quad = \quad o\left(\lambda_m^{1-2d_{ab}}\right)$$

Proof. A proof very similar to that of Lemma 18 of Chen and Hurvich [2006] shows that:

$$E\left(\left|I_{ab}(\lambda_j) - \Psi_a(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_b^*(\lambda_{\tilde{j}})\right|\right) \quad \leq \quad C\lambda_j^{-d_{aa}-d_{bb}} j^{-\gamma/2}$$

(Their Assumption 2 is similar to our Assumption 6, but uses $\Psi_{jk}^\dagger(\lambda)$ instead of $\tau_{jk}(\lambda)e^{i\varphi_{jk}(\lambda)}$. Because the results are proved for each (j, k) separately, we can allow δ_{jk} to vary with j, k when we apply the univariate results from Hurvich et al. [2002] in the proof of that lemma. Since our Assumption 2 allows $\tau_{au}(\lambda) \equiv 0$ (hence $\Psi_{au}(\lambda) \equiv 0$), $\lambda \in [0, \pi]$ for certain values of $u \in 1, \cdots, p$, we would restrict the double sum in (49) of Chen and Hurvich [2006] to those values of u, v such that $\tau_{au}(0) > 0$ and $\tau_{bv}(0) > 0$.)

Then, with C being an arbitrary non-zero constant that may change from one line to the next, we find the expected modulus of the sum in (3.4):

$$E\left(\frac{2\pi}{n}\sum_{j=1}^{m}\left|I_{ab}(\lambda_j) - \Psi_a(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi_b^*(\lambda_{\tilde{j}})\right|\right) \tag{7.2}$$

$$\leq \frac{C}{n}\sum_{j=1}^{m}\lambda_j^{-d_{aa}-d_{bb}}j^{-\gamma/2} \quad = \quad Cn^{-1+d_{aa}+d_{bb}}\sum_{j=1}^{m}j^{-d_{aa}-d_{bb}-\gamma/2} \tag{7.3}$$

Based on the value of $-d_{aa}-d_{bb}-\gamma/2$, we have three cases:

Case 1: $-d_{aa}-d_{bb}-\gamma/2 < -1$. In this case, $\sum_{j=1}^{m} j^{-d_{aa}-d_{bb}-\gamma/2} = O(1)$. Because $d_{aa}+d_{bb} < 1$,

$$\begin{aligned} n^{-1+d_{aa}+d_{bb}}\sum_{j=1}^{m}j^{-d_{aa}-d_{bb}-\gamma/2} &= O\left(n^{-1+d_{aa}+d_{bb}}\right) \\ &= O\left(\lambda_m^{1-d_{aa}-d_{bb}}m^{d_{aa}+d_{bb}-1}\right) &(7.4) \\ &= O\left(\lambda_m^{1-2d_{ab}}n^{-2d_\rho}m^{2d_{ab}-1}\right) &(7.5) \end{aligned}$$

Since $d_{aa}+d_{bb}<1$ by stationarity, Equation (7.4) is $o\left(\lambda_m^{1-d_{aa}-d_{bb}}\right)$. If Assumption 9 holds and $a \neq b$, then Equation (7.5) is $o\left(\lambda_m^{1-2d_{ab}}\right)$.

Case 2: $-d_{aa}-d_{bb}-\gamma/2 = -1$. Then, $\sum_{j=1}^{m} j^{-d_{aa}-d_{bb}-\gamma/2} = \sum_{j=1}^{m}\frac{1}{j} = O(\log(m))$, and we have:

$$\begin{aligned} n^{-1+d_{aa}+d_{bb}}\sum_{j=1}^{m}j^{-d_{aa}-d_{bb}-\gamma/2} &= O\left(n^{-1+d_{aa}+d_{bb}}\log(m)\right) \\ &= O\left(\lambda_m^{1-d_{aa}-d_{bb}}m^{d_{aa}+d_{bb}-1}\log(m)\right) &(7.6) \\ &= O\left(\lambda_m^{1-2d_{ab}}n^{-2d_\rho}m^{2d_{ab}-1}\log(m)\right) &(7.7) \end{aligned}$$

As before, $d_{aa}+d_{bb}<1$ by stationarity, so $m^{d_{aa}+d_{bb}-1}\log(m) = o(1)$ and Equation (7.6) is $o\left(\lambda_m^{1-d_{aa}-d_{bb}}\right)$. If Assumption 9 holds, then Equation (7.7) is $o\left(\lambda_m^{1-2d_{ab}}\right)$.

Case 3: $-d_{aa}-d_{bb}-\gamma/2 > -1$. In this case, we rewrite:

$$\begin{aligned} Cn^{-1+d_{aa}+d_{bb}}\sum_{j=1}^{m}j^{-d_{aa}-d_{bb}-\gamma/2} &= O\left(\frac{1}{m^{\gamma/2}}\lambda_m^{1-d_{aa}-d_{bb}}\right) &(7.8) \\ &= O\left(\frac{\lambda_m^{2d_\rho}}{m^{\gamma/2}}\lambda_m^{1-2d_{ab}}\right) &(7.9) \end{aligned}$$

Since $m \to \infty$, (7.8) is $o(\lambda_m^{1-d_{aa}-d_{bb}})$. If $\frac{n^{\frac{2d_\rho}{2d_\rho-\gamma/2}}}{m} \to 0$, then $m^{-\gamma/2}\lambda_m^{2d_\rho} = o(1)$ and (7.9) is $o(\lambda_m^{1-2d_{ab}})$.

■

The next lemma is closely related to Lemma 19 of Chen and Hurvich [2006], generalizing the result to the case where ϵ_t is non-Gaussian. In certain cases, the bound on $|E\left(S_{ab}(\lambda_j)S_{ab}(\lambda_k)\right)|$ could replace $-d_{aa}-d_{bb}$ by $-2d_{ab}$; however, this will happen only for values of the fourth cumulants and $E\left(J_{\epsilon,u_1}(\lambda_j)\overline{J_{\epsilon,v_2}(\lambda_j)}\right)E\left(J_{\epsilon,u_2}(\lambda_k)\overline{J_{\epsilon,v_1}(\lambda_k)}\right)$ that preserve the power law coherency properties.

Lemma 7 *Let $I_{\epsilon,u,v}(\lambda_j)$ be the (u,v) element of the cross-periodogram of p-variate white noise. Let $\lambda_{\tilde{j}}$ be the shifted Fourier frequency. Let $1 \le j,k \le n/2$. Define*

$$S(\lambda_j) \;=\; \Psi(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi^*(\lambda_{\tilde{j}}) - f(\lambda_j)$$

Let $S_{ab}(\lambda_j)$ be the (a,b) element of $S(\lambda_j)$. Then under Assumptions 1-5 and Assumption 7,

$$\left|E\left(S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right)\right| \;\le\; \begin{cases} C|\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{aa}+d_{bb})} + O\left(\frac{1}{n}|\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{aa}+d_{bb})}\right) & |j-k| \le s \\ O\left(\frac{1}{n}|\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{aa}+d_{bb})}\right) & |j-k| > s \end{cases}$$

Proof. Following Chen and Hurvich [2006], we write:

$$\begin{aligned} E\left(S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right) \;&=\; \sum_{u_1=1}^{p}\sum_{u_2=1}^{p}\sum_{v_1=1}^{p}\sum_{v_2=1}^{p}\Psi_{au_1}(\lambda_{\tilde{j}})\overline{\Psi_{au_2}(\lambda_{\tilde{k}})}\,\overline{\Psi_{bv_1}(\lambda_{\tilde{j}})}\Psi_{bv_2}(\lambda_{\tilde{k}}) \\ &\quad\times E\left[\left(I_{\epsilon,u_1v_1}(\lambda_j)-\sigma_{u_1v_1}\right)\left(\overline{I_{\epsilon,u_2v_2}(\lambda_k)}-\sigma_{u_2v_2}\right)\right] \end{aligned} \tag{7.10}$$

and

$$\begin{aligned} E\left(\left(I_{\epsilon,u_1v_1}(\lambda_j)-\sigma_{u_1v_1}\right)\left(\overline{I_{\epsilon,u_2v_2}(\lambda_k)}-\sigma_{u_2v_2}\right)\right) \;&=\; cum\left(J_{\epsilon,u_1}(\lambda_j),\overline{J_{\epsilon,u_2}(\lambda_k)},\overline{J_{\epsilon,v_1}(\lambda_j)},J_{\epsilon,v_2}(\lambda_k)\right) \\ &\quad + E\left(J_{\epsilon,u_1}(\lambda_j)J_{\epsilon,v_2}(\lambda_k)\right)E\left(\overline{J_{\epsilon,u_2}(\lambda_j)J_{\epsilon,v_1}(\lambda_k)}\right) \\ &\quad + E\left(J_{\epsilon,u_1}(\lambda_j)\overline{J_{\epsilon,u_2}(\lambda_k)}\right)E\left(\overline{J_{\epsilon,v_1}(\lambda_j)}J_{\epsilon,v_2}(\lambda_k)\right). \end{aligned}$$

Note that

$$\left|E\left(J_{\epsilon,u_1}(\lambda_j)J_{\epsilon,v_2}(\lambda_k)\right)E\left(\overline{J_{\epsilon,u_2}(\lambda_j)J_{\epsilon,v_1}(\lambda_k)}\right) + E\left(J_{\epsilon,u_1}(\lambda_j)\overline{J_{\epsilon,u_2}(\lambda_k)}\right)E\left(\overline{J_{\epsilon,v_1}(\lambda_j)}J_{\epsilon,v_2}(\lambda_k)\right)\right|$$

$$= C\chi(|j-k| \le s)\ .$$

Next, we compute the cumulant:

$$\begin{aligned} & cum\left(J_{\epsilon,u_1}(\lambda_j),\overline{J_{\epsilon,u_2}(\lambda_k)},\overline{J_{\epsilon,v_1}(\lambda_j)},J_{\epsilon,v_2}(\lambda_k)\right) \\ =\; & cum\left(\frac{1}{\sqrt{2\pi na_s}}\sum_{t=1}^{n}h_t^s\epsilon_{u_1,t}e^{it\lambda_j},\frac{1}{\sqrt{2\pi na_s}}\sum_{t=1}^{n}\overline{h_t^s}\epsilon_{u_2,t}e^{-it\lambda_k},\frac{1}{\sqrt{2\pi na_s}}\sum_{t=1}^{n}\overline{h_t^s}\epsilon_{v_1,t}e^{-it\lambda_j},\frac{1}{\sqrt{2\pi na_s}}\sum_{t=1}^{n}h_t^s\epsilon_{v_2,t}e^{it\lambda_k}\right) \\ =\; & \frac{1}{(2\pi a_s)^2n^2}cum(\epsilon_{u_1,1},\epsilon_{u_2,1},\epsilon_{v_1,1},\epsilon_{v_2,1})\sum_{t=1}^{n}|h_t^s|^4 = O\left(\frac{1}{n}\right) \end{aligned}$$

by Assumption 1.

Substituting these results into Equation (7.10), we find that:

$$\begin{aligned}\left|E\left(S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right)\right| &= O\left(\left(\chi(|j-k|\leq s)+\frac{1}{n}\right)\sum_{u_1=1}^{p}\sum_{u_2=1}^{p}\sum_{v_1=1}^{p}\sum_{v_2=1}^{p}\left|\Psi_{au_1}(\lambda_{\tilde{j}})\overline{\Psi_{au_2}(\lambda_{\tilde{k}})}\,\overline{\Psi_{bv_1}(\lambda_{\tilde{j}})}\Psi_{bv_2}(\lambda_{\tilde{k}})\right|\right)\\ &= O\left(\left(\chi(|j-k|\leq s)+\frac{1}{n}\right)|\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-(d_{aa}+d_{bb})}\right)\end{aligned}$$

■

Lemma 8 *Let $S(\lambda_j)=\Psi(\lambda_{\tilde{j}})I_\epsilon(\lambda_j)\Psi^*(\lambda_{\tilde{j}})-f(\lambda_j)$. Under Assumptions 1–5, 7–8,*

$$E\left(\frac{4\pi^2}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m}S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right) = o\left(\lambda_m^{2-2d_{aa}-2d_{bb}}\right).$$

If $a\neq b$ under the assumptions above together with Assumptions 10, 11, 12,

$$E\left(\frac{4\pi^2}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m}S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right) = o\left(\lambda_m^{2-4d_{ab}}\right).$$

Proof. Applying Lemma 7, the expected squared modulus of the sum in (3.5) is:

$$E\left(\frac{4\pi^2}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m}S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right)\leq\frac{4\pi^2}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left|E\left(S_{ab}(\lambda_j)\overline{S_{ab}(\lambda_k)}\right)\right|$$

$$=O\Bigg(\frac{1}{n^3}\sum_{j=1}^{m}\sum_{k=1}^{m}\lambda_{\tilde{j}}^{-d_{aa}-d_{bb}}\lambda_{\tilde{k}}^{-d_{aa}-d_{bb}} \tag{7.11}$$

$$+\frac{1}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m}|\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-d_{aa}-d_{bb}}\chi(|j-k|\leq s)\Bigg). \tag{7.12}$$

We first consider (7.11). Since $d_{aa}+d_{bb}<1$,

$$\begin{aligned}\frac{1}{n^3}\sum_{j=1}^{m}\sum_{k=1}^{m}\lambda_{\tilde{j}}^{-d_{aa}-d_{bb}}\lambda_{\tilde{k}}^{-d_{aa}-d_{bb}} &= \frac{1}{n}\left(\frac{1}{n}\sum_{j=1}^{m}\lambda_{\tilde{j}}^{-d_{aa}-d_{bb}}\right)^2\\ &= O\left(\frac{1}{n}\lambda_m^{2-2d_{aa}-2d_{bb}}\right)\end{aligned}$$

which is $o(\lambda_m^{2-2d_{aa}-2d_{bb}})$ under Assumption 8, and is $o(\lambda_m^{2-4d_{ab}})$ under Assumption 12 since

$$1+\frac{1}{4d_\rho}<\frac{-4d_\rho}{1-4d_\rho} \quad .$$

Next, we consider (7.12),

$$\begin{aligned}\frac{1}{n^2}\sum_{j=1}^{m}\sum_{k=1}^{m}|\lambda_{\tilde{j}}\lambda_{\tilde{k}}|^{-d_{aa}-d_{bb}}\chi(|j-k|\leq s) &= O\left(\frac{1}{n^2}\sum_{j=1}^{m}\lambda_{\tilde{j}}^{-2d_{aa}-2d_{bb}}\right)\\ &= O\left(n^{-2+2d_{aa}+2d_{bb}}\sum_{j=1}^{m}j^{-2d_{aa}-2d_{bb}}\right).\end{aligned}$$

As in Lemma 6, there are three cases, now based on the value of $-2d_{aa}-2d_{bb}$.

Case 1: $-2d_{aa}-2d_{bb}<-1$. In this case, $\sum_{j=1}^{m}j^{-2d_{aa}-2d_{bb}}=O(1)$ as $m\to\infty$. Thus,

$$\begin{aligned}n^{-2+2d_{aa}+2d_{bb}}\sum_{j=1}^{m}j^{-2d_{aa}-2d_{bb}} &= O\left(n^{-2+2d_{aa}+2d_{bb}}\right)\\ &= O\left(\lambda_m^{2-2d_{aa}-2d_{bb}}m^{2d_{aa}+2d_{bb}-2}\right) \qquad (7.13)\\ &= O\left(\lambda_m^{2-4d_{ab}}n^{-4d_\rho}m^{4d_{ab}-2}\right). \qquad (7.14)\end{aligned}$$

Under Assumption 8, the expression in Equation (7.13) is $o(\lambda_m^{2-2d_{aa}-2d_{bb}})$ because $d_{aa}+d_{bb}<1$ and $m\to\infty$. The expression in Equation (7.14) is $o(\lambda_m^{2-2d_{ab}})$ under Assumption 10.

Case 2: $-2d_{aa}-2d_{bb}=-1$. In this case, $\sum_{j=1}^{m}j^{-2d_{aa}-2d_{bb}}=O(\log(m))$ as $m\to\infty$. Then,

$$\begin{aligned}n^{-2+2d_{aa}+2d_{bb}}\sum_{j=1}^{m}j^{-2d_{aa}-2d_{bb}} &= O\left(n^{-2+2d_{aa}+2d_{bb}}\log(m)\right)\\ &= O\left(\lambda_m^{2-2d_{aa}-2d_{bb}}m^{2d_{aa}+2d_{bb}-2}\log(m)\right) \qquad (7.15)\\ &= O\left(\lambda_m^{2-4d_{ab}}n^{-4d_\rho}m^{4d_{ab}-2}\log(m)\right). \qquad (7.16)\end{aligned}$$

Under Assumption 8, the expression in Equation (7.15) is $o(\lambda_m^{2-2d_{aa}-2d_{bb}})$ because $d_{aa}+d_{bb}<1$ and $m\to\infty$. The expression in Equation (7.16) is $o(\lambda_m^{2-2d_{ab}})$ under Assumption 11.

Case 3: $-2d_{aa}-2d_{bb}>-1$. Then, we have

$$\begin{aligned}n^{-2+2d_{aa}+2d_{bb}}\sum_{j=1}^{m}j^{-2d_{aa}-2d_{bb}} &= O\left(\frac{1}{n}\lambda_m^{1-2d_{aa}-2d_{bb}}\right)\\ &= O\left(\frac{1}{m}\lambda_m^{2-2d_{aa}-2d_{bb}}\right). \qquad (7.17)\end{aligned}$$

which is $o(\lambda_m^{2-2d_{aa}-2d_{bb}})$ under Assumption 8, and is $o(\lambda_m^{2-4d_{ab}})$ under Assumption 12.

■

Lemma 9 *Under Assumptions 1-5,*

$$\frac{2\pi}{n}\sum_{j=1}^{m} f_{ab}(\lambda_j) - F_{ab}(\lambda_m) \quad = \quad o(F_{ab}(\lambda_m)).$$

Proof. We consider in detail the case $a = 1$, $b = 2$, as the other cases can be proved similarly. Taylor's Theorem yields for $\lambda \in [\lambda_{j-1}, \lambda_j)$,

$$f_{12}(\lambda) = f_{12}(\lambda_j) + (\lambda - \lambda_j) f_{12}'(\xi_j)$$

where ξ_j is between λ and λ_j, and therefore in $[\lambda_{j-1}, \lambda_j)$. We have

$$\begin{aligned}
\frac{2\pi}{n}\sum_{j=1}^{m} f_{12}(\lambda_j) - F_{12}(\lambda_m) &= \sum_{j=1}^{m}\int_{\lambda_{j-1}}^{\lambda_j} [f_{12}(\lambda_j) - f_{12}(\lambda)]d\lambda \\
&= \sum_{j=1}^{m}\int_{\lambda_{j-1}}^{\lambda_j} (\lambda_j - \lambda) f_{12}'(\xi_j) d\lambda = \sum_{j=1}^{m} f_{12}'(\xi_j)\left(\frac{2\pi}{n}\frac{2\pi j}{n} - \frac{\lambda_j^2}{2} + \frac{\lambda_{j-1}^2}{2}\right) \\
&= \frac{2\pi^2}{n^2}\sum_{j=1}^{m} f_{12}'(\xi_j).
\end{aligned}$$

As in the proof of Lemma 1, we can write $f_{12}(\lambda) = |\lambda|^{-2d_{12}} s(\lambda)$ where $\lim_{\lambda \to 0^+} s(\lambda) = s(0; q_0) \neq 0$, and there exist constants C, A and Ξ such that $|s'(\theta)| \leq C\theta^{-1+\Xi}$ for all $\theta \in (0, A)$. Differentiability implies that s is continuous, therefore bounded, on $(0, A)$. Since $m/n \to 0$ we have for sufficiently large m that

$$\begin{aligned}
\left|\frac{2\pi}{n}\sum_{j=1}^{m} f_{12}(\lambda_j) - F_{12}(\lambda_m)\right| &\leq \frac{2\pi^2}{n^2}\sum_{j=1}^{m} |f_{12}'(\xi_j)| \\
&\leq \frac{2\pi^2}{n^2}\sum_{j=1}^{m} |\xi_j|^{-2d_{12}} C |\xi_j|^{-1+\Xi} + Const\frac{2\pi^2}{n^2}\sum_{j=1}^{m} |\xi_j|^{-2d_{12}-1} \\
&\leq Const\frac{1}{n^2}\sum_{j=1}^{m} |\xi_j|^{-2d_{12}-1}
\end{aligned}$$

If $-2d_{12}-1 \geq 0$ (that is, $d_{12} \leq -1/2$), then $|\xi_j|^{-2d_{12}-1} \leq |\lambda_j|^{-2d_{12}-1}$ and so it suffices to show that $\frac{1}{n^2}\sum_{j=1}^{m} |\lambda_j|^{-2d_{12}-1} = o(F_{12}(\lambda_m))$. It suffices to show this even in the case $-2d_{12} - 1 < 0$, since $n^{-1}f_{12}(\lambda_1) = o(F_{12}(\lambda_m))$ and $\sum_{j=2}^m |\lambda_{j-1}|^{-2d_{12}-1} \leq \sum_{j=1}^m |\lambda_j|^{-2d_{12}-1}$. Therefore, we define

$$A(m,n) = \frac{1}{n^2}\sum_{j=1}^m (j/n)^{-2d_{12}-1}.$$

Case I: $-2d_{12} - 1 < -1$**.**

Then $|A(m,n)| \leq Const\, n^{-2}(1/n)^{-2d_{12}-1} = O(n^{-1+2d_{12}})$, so that $|A(m,n)/F_{12}(\lambda_m)| = O(m^{2d_{12}-1}) \to 0$, since $m \to \infty$ and $2d_{12} - 1 < 0$.

Case II: $-2d_{12} - 1 > -1$**.**

Then $|A(m,n)| \leq Const\ n^{-2}n^{1+2d_{12}}m^{-2d_{12}} = O(n^{2d_{12}-1}m^{-2d_{12}})$, so that $|A(m,n)/F_{12}(\lambda_m)| = O(m^{-1}) \to 0$.

Case III: $-2d_{12} - 1 = -1$**.**

Here, we have $d_{12} = 0$ and hence $f_{12}(\lambda) = s(\lambda)$. A similar argument to that presented at the beginning of this proof yields

$$\begin{aligned} \left| \frac{2\pi}{n}\sum_{j=1}^m f_{12}(\lambda_j) - F_{12}(\lambda_m) \right| &\leq \frac{2\pi^2}{n^2}\sum_{j=1}^m |f_{12}'(\lambda_j)| \leq Const\ \frac{1}{n^2}\sum_{j=1}^m |\lambda_j|^{-1+\Xi} \\ &= O(n^{-2}n^{1-\Xi}m^{\Xi}) = O(n^{-1-\Xi}m^{\Xi}) \\ &= o(F_{12}(\lambda_m)). \end{aligned}$$

■

8 Additional Proofs

Proof of Lemma 4: We have $S(m,n,q) = \sum_{j=1}^m \alpha_j S_{12}(\lambda_j)$ where

$$\alpha_j = \begin{cases} \frac{2\pi}{nG(\lambda_m)} - \frac{2\pi}{nG(q\lambda_m)} & \text{if } j \leq \lfloor qm \rfloor \\ \frac{2\pi}{nG(\lambda_m)} & \text{if } j > \lfloor qm \rfloor \end{cases}$$

Thus,

$$\mathrm{cum}[\underbrace{S(m,n,q),\ldots,S(m,n,q)}_{r\text{ terms}}] = \sum_{j_1=1}^m \cdots \sum_{j_r=1}^m \alpha_{j_1}\cdots\alpha_{j_r} cum[S_{12}(\lambda_{j_1}),\cdots,S_{12}(\lambda_{j_r})]\ .$$

Now,

$$cum[S_{12}(\lambda_{j_1}),\cdots,S_{12}(\lambda_{j_r})]=\sum_{u_1=1}^{p}\cdots\sum_{u_r=1}^{p}\sum_{v_1=1}^{p}\cdots\sum_{v_r=1}^{p}\Psi_{1u_1}(\lambda_{\tilde{j_1}})\overline{\Psi_{2v_1}(\lambda_{\tilde{j_1}})}\cdots\Psi_{1u_r}(\lambda_{\tilde{j_r}})\overline{\Psi_{2v_r}(\lambda_{\tilde{j_r}})}$$

$$\cdot\, cum[I_{\epsilon,u_1v_1}(\lambda_{j_1}),\cdots,I_{\epsilon,u_rv_r}(\lambda_{j_r})]\ .$$

Consider partitions of the table

$$\begin{array}{cc} J_{\epsilon,u_1}(\lambda_{j_1}) & \overline{J_{\epsilon,v_1}(\lambda_{j_1})} \\ J_{\epsilon,u_2}(\lambda_{j_2}) & \overline{J_{\epsilon,v_2}(\lambda_{j_2})} \\ \vdots & \vdots \\ J_{\epsilon,u_r}(\lambda_{j_r}) & \overline{J_{\epsilon,v_r}(\lambda_{j_r})} \end{array}$$

Following Brillinger [1981] (p. 20) we say that two elements of the partition *hook* if they both have an entry on the same row of the table. Moreover, two elements of the partition *communicate* if they are linked by a sequence of pairs that hook. Finally, the partition is *indecomposable* if all elements communicate. Now, Brillinger [1981] Theorem 2.3.2 (p. 21) implies that if we denote the entries of the above table by X_{ij} then

$$cum[I_{\epsilon,u_1v_1}(\lambda_{j_1}),\cdots,I_{\epsilon,u_rv_r}(\lambda_{j_r})]=\sum_{\nu} cum(X_{ij}\,;\,i,j\in\nu_1)\cdots cum(X_{ij}\,;\,i,j\in\nu_p) \tag{8.1}$$

where the sum is over all indecomposable partitions $\nu=\nu_1\cup\cdots\cup\nu_p$ of the table. Due to Gaussianity, the only partitions yielding a nonzero contribution to (8.1) are those whose elements all have exactly two entries. So we must have $p=r$ and none of the elements $\nu_1,\cdots\nu_r$ of the partition can be a row of the table, on account of indecomposability. Arguing as in the proof of Lemma 7, and in view of (8.1), we see that $cum[I_{\epsilon,u_1v_1}(\lambda_{j_1}),\cdots,I_{\epsilon,u_rv_r}(\lambda_{j_r})]$ is bounded by a constant (not depending on r, n, or on $j_1,\cdots,j_r$) and this cumulant is exactly zero if $\max_{K,L\in\{1,\cdots,r\}}|j_K-j_L|>s$. Thus,

$$|\mathrm{cum}[\underbrace{S(m,n,q),\ldots,S(m,n,q)}_{r\text{ terms}}]|\le C\cdot\sum_{j_1=1}^{m}\cdots\sum_{j_r=1}^{m}|\lambda_{j_1}\cdots\lambda_{j_r}|^{-d_{11}-d_{22}}|\alpha_{j_1}\cdots\alpha_{j_r}|\chi(\max_{K,L\in\{1,\cdots,r\}}|j_K-j_L|\le s)$$

Since $|\alpha_j|\le Cn^{-1}\lambda_m^{2d_{12}-1}=Cm^{2d_{12}-1}n^{-2d_{12}}$ we have

$$|\mathrm{cum}[\underbrace{S(m,n,q),\ldots,S(m,n,q)}_{r\text{ terms}}]|\le C(m^{2d_{12}-1}n^{-2d_{12}})^r\sum_{j=1}^{m}\lambda_j^{-r(d_{11}+d_{22})}$$

$$\le C(m^{2d_{12}-1}n^{-2d_{12}})^r n^{r(d_{11}+d_{22})}\cdot\begin{cases}1 & \text{if } \ -r(d_{11}+d_{22})<-1\\ m^{-r(d_{11}+d_{22})+1} & \text{if } \ -r(d_{11}+d_{22})\ge -1\end{cases}$$

We also have

$$|\mathrm{cum}(\underbrace{S(m,n,q)/\gamma_{m,n},\ldots,S(m,n,q)/\gamma_{m,n}}_{r \text{ terms}})| = \gamma_{m,n}^{-r}|\mathrm{cum}[\underbrace{S(m,n,q),\ldots,S(m,n,q)}_{r \text{ terms}}]|$$

Case I: $-r(d_{11}+d_{22}) < -1$**.** Then

$$|\mathrm{cum}(\underbrace{S(m,n,q)/\gamma_{m,n},\ldots,S(m,n,q)/\gamma_{m,n}}_{r \text{ terms}})| \le C(m^{2d_\rho-1/2}n^{-2d_\rho})^{-r}(m^{2d_{12}-1}n^{-2d_{12}})^r n^{r(d_{11}+d_{22})}$$

$$= C(m^{d_{11}+d_{22}-1/2})^r$$

which tends to zero since one of the assumptions of the lemma is that $d_{11}+d_{22} < 1/2$.

Case II: $-r(d_{11}+d_{22}) \ge -1$**.** Then

$$|\mathrm{cum}(\underbrace{S(m,n,q)/\gamma_{m,n},\ldots,S(m,n,q)/\gamma_{m,n}}_{r \text{ terms}})| \le C(m^{d_{11}+d_{22}-1/2})^r m^{-r(d_{11}+d_{22})+1}$$

$$= m^{-r/2+1} \to 0$$

since $r \ge 3$. $\Box$

ACKNOWLEDGEMENT

The authors thank the Associate Editor and two anonymous referees for their comments that led to substantial improvements in the paper.

References

David R. Brillinger. *Time Series: Data Analysis and Theory.* Holden-Day, Inc., 1981.

Willa W. Chen and Clifford M. Hurvich. Semiparametric estimation of fractional cointegrating subspaces. *Annals of Statistics*, 34(6):2939–2979, 2006.

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

E.J. Hannan. *Multiple Time Series.* John Wiley and Sons, Inc., 1970.

J. Hidalgo. Spectral analysis for bivariate time series with long memory. *Econometric Theory*, 12:773–792, 1996.

Clifford M. Hurvich and Willa W. Chen. An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis*, 21(2):156–180, 2000.

Clifford M. Hurvich, Eric Moulines, and Philippe Soulier. The FEXP estimator for potentially non-stationary linear time series. *Stochastic Processes and their Applications*, 97:307–340, 2002.

Clifford M. Hurvich, Eric Moulines, and Philippe Soulier. Estimating long memory in volatility. *Econometrica*, 73:1283–1328, 2005.

H. R. Kunsch. Statistical aspects of self-similar processes. In Yu Prohorov and V. V. Sazanov, editors, *Proceedings of the First World Congress of the Bernoulli Society 1*, pages 67–74, Utrecht, 1987. VNU Science Press.

I. N. Lobato. A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, 90:129–153, 1999.

I. N. Lobato and P.M. Robinson. Averaged periodogram estimation of long memory. *Journal of Econometrics*, 73:303–324, 1996.

Ignacio N. Lobato. Consistency of the averaged cross-periodogram in long memory time series. *Journal of Time Series Analysis*, 18(2):137–155, 1997.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL `http://www.R-project.org`.

Peter M. Robinson. Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23(3):1048–1072, 1995a.

Peter M. Robinson. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23:1630–1661, 1995b.

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

Peter M. Robinson. Multiple local Whittle estimation in stationary systems. *Annals of Statistics*, 36(5):2508–2530, 2008.

Rebecca J. Sela. Three essays in econometrics: Multivariate long memory time series and applying regression trees to longitudinal data. `http://pages.stern.nyu.edu/~rsela/DissertationSummary.pdf`, 2010.

Rebecca J. Sela and Clifford M. Hurvich. Computationally efficient gaussian maximum likelihood methods for vector ARFIMA models. *Journal of Time Series Analysis*, 30(6):631–651, 2009.

Katsumi Shimotsu. Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics*, 137:277–310, 2007.

Carlos Velasco. Non-stationary log-periodogram regression. *Journal of Econometrics*, 91:325–371, 1999.