NONPARAMETRIC TEST FOR CAUSALITY
WITH LONG-RANGE DEPENDENCE[*]

by

Javier Hidalgo
London School of Economics and Political Science

Contents:

Abstract

This paper introduces a nonparametric Granger-causality test for covariance stationary linear processes under, possibly, the presence of long-range dependence. We show that the test is consistent and has power against contiguous alternatives converging to the parametric rate $T^{\frac{1}{2}}$. Since the test is based on estimates of the parameters of the representation of a *VAR* model as a, possibly, two-sided infinite distributed lag model, we first show that a modification of Hannan's (1963, 1967) estimator is root-T consistent and asymptotically normal for the coefficients of such a representation. When the data is long-range dependent this method of estimation becomes more attractive than Least Squares, since the latter can be neither root-T consistent nor asymptotically normal as is the case with short-range dependent data.

Keywords: Causality; long-range dependence; spectral analysis; distributed lag model; consistent test.

JEL Nos.: C12, C13, C32.

1.- INTRODUCTION

Since Granger (1969), when analyzing relationships between economic variables, the concept of causality has been a basic subject routinely invoked in the econometric/economic literature. Typically, tests for causality are commonly performed in the context of unrestricted vector autoregressive $(VAR(P))$ models with P a finite known positive number. See among others, Granger (1969) or Geweke (1982) when the data is short-range dependent, or for variables showing stochastic-trend behaviour, see Sims et al. (1990) or Toda and Phillips (1993). Some extensions are in Hosoya (1991) who analyses causality for stationary short-range dependent processes which do not necessarily have a VAR representation or Lütkepohl and Poskitt (1996), and references therein, who allow for an infinite order VAR model. However, the above models do not cover the so-called long-range dependent processes which have attracted immense attention in recent years in the econometric literature. The object of this paper is thus to introduce and analyze a causality test which allows for long-range dependence. In addition, the test does not rely on any specific finite parameterization of the model and covers processes which do not possess a finite order vector autoregressive moving average $(VARMA)$ representation such as Bloomfield's (1973) exponential model.

To fix ideas, let a bivariate observable covariance stationary vector $w_t = (y_t, x_t)'$ satisfy

$$A(L) w_t = \sum_{j=0}^{\infty} A_j w_{t-j} = \varepsilon_t, \quad t = 1, 2..., T,$$

where ε_t is a bivariate martingale difference sequence and A_0 is the identity matrix. The interest is in testing the null hypothesis H_0: $y_t \nRightarrow x_t$, that is y_t does not cause x_t, against the alternative hypothesis $H_1 : y_t \Longrightarrow x_t$, that is y_t causes x_t.

The main attributes of the test, described below, are; 1) it is nonparametric, that is, we do not impose any specific parametric model for the data, 2) it is consistent, 3)

it has power against $T^{-1/2}$ contiguous alternatives, and 4) its limit distribution can be obtained, say, from the distribution of the supremum of the standard Brownian motion. Thus, the paper extends previous work in two main directions. First, we allow for a general covariance stationary linear process and second, since it is nonparametric, we avoid the danger of possible misspecification.

Now we briefly discuss the main ideas of the test. Following Sims (1972) or Hosoya (1977), a test for H_0 is equivalent to testing whether the coefficients $c(j)$ are simultaneously equal to zero for all $j < 0$ in

$$y_t = \sum_{j=-\infty}^{\infty} c(j) x_{t-j} + u_t, \tag{1.1}$$

where, by construction, $E[u_t | x_s, -\infty < s < \infty] = 0$, and x_t and u_t are, possibly, long-range dependent processes. Alternatively, the null hypothesis H_0 is equivalent to

$$\sum_{j=-\infty}^{0} c(j-1) \cos(\pi j \lambda) = 0 \quad \forall \lambda \in [0,1],$$

or

$$S^*(\mu) = \mathrm{Re}\left(\int_0^\mu \left(\sum_{j=-\infty}^{0} c(j-1) e^{-i\pi j \lambda}\right) d\lambda\right) = 0 \quad \forall \mu \in [0,1],$$

where $\mathrm{Re}(a)$ denotes the real part of a complex number a. Therefore the hypothesis testing can be described as

$$H_0 : S^*(\mu) = 0 \quad \forall \mu \in [0,1] \text{ against } H_1 : S^*(\mu) \neq 0 \text{ in } \Delta \subset [0,1] \tag{1.2}$$

where Δ has Lebesgue measure greater than zero.

Given estimates of $c(j)$, say $\widehat{c}(j)$, and using Riemann's discrete approximation of integrals by sums, $S^*(\mu)$ can be estimated by

$$S_T(\mu) = \mathrm{Re}\left(\frac{1}{M} \sum_{p=1}^{[M\mu]} \left(\sum_{j=-\infty}^{0} \widehat{c}(j-1) e^{-ij\lambda_p}\right)\right) \tag{1.3}$$

2

where $\lambda_p = \pi p/M$, $p = 1, ..., M$, and $M = M(T)$ is a number which increases slowly with T, that is $M^{-1} + MT^{-1} \to 0$. The test can thus be based on whether or not $S_T(\mu)$ is significantly different than zero for all $\mu \in [0, 1]$ by the implementation of a functional of $S_T(\mu)$, say a Kolmogorov-Smirnov test.

The remainder of the paper is organized as follows. In the next section, we describe and motivate the estimation technique of the coefficients $c(j)$, whose statistical properties are given in Section 4. Section 3 describes the statistical framework. Also, since our statistics and proofs are based on the spectral density matrix estimator of w_t and the cross-spectrum between x_t and u_t, we give some useful statistical properties of them. In Section 5, we study the properties of (1.3) and we provide the test for Granger-causality and its implementation. In Section 6, we discuss the choice of M and how the results of Sections 4 and 5 are extended to general p-dimensional data. Finally, Section 7 gives a summary of the paper. The proofs of the results are confined to Mathematical Appendix A which apply some technical Lemmas given in Mathematical Appendix B.

2.- ESTIMATION PROCEDURE AND ITS MOTIVATION

In this section we describe the estimation technique of the coefficients $c(j)$ in (1.1) and discuss why it is more desirable than Least Squares (LSE) estimates in the presence of long-range dependence. To accomplish this, we adopt the frequency domain approach whose merits have been proven to be numerous and applied in several contexts. For example, it was used when testing for causality in the pioneer work by Granger (1969), or in Geweke (1986) to analyze the neutrality of money. Recently, Hosoya (1995) has shown the usefulness of spectral analysis in contrast to time domain methods when analyzing the causality among economic variables, see also Geweke (1982). Another example is in the efficient estimation of the parameters in a regression model, see the cornerstone work by Hannan (1963), and extended to

3

more general models useful in econometrics by Hannan and Terrell (1973) and recently by Robinson (1991) who also allowed for data dependent smoothing in the spectral estimation. One rationale of the approach lies in that only minimal conditions of the process are needed, like stationarity, and no explicit assumptions on its dynamic specification are thus required.

In the frequency domain, the lag structure given in (1.1) is described by the frequency response function $C(\lambda) = \sum_{j=-\infty}^{\infty} c(j) e^{-ij\lambda}$, so $c(j)$ is interpreted as the jth Fourier coefficient of $C(\lambda) = f_{xx}^{-1}(\lambda) f_{yx}(\lambda)$, that is,

$$c(j) = (2\pi)^{-1} \int_0^{2\pi} C(\lambda) e^{ij\lambda} d\lambda, \tag{2.1}$$

where $f_{yx}(\lambda)$ and $f_{xx}(\lambda)$ are the indicated elements of the spectral density matrix, $f_{ww}(\lambda)$, of w_t defined from the relationship

$$E\left((w_1 - Ew_1)(w_{j+1} - Ew_1)'\right) = \int_{-\pi}^{\pi} f_{ww}(\lambda) e^{-ij\lambda} d\lambda \quad j = 0, \pm 1, \pm 2, ...$$

Due to this interpretation, Hannan (1963, 1967), see also Brillinger (1981), proposed to estimate $c(j)$ by the sample (discrete) analogue of (2.1),

$$\widetilde{c}(j) = \frac{1}{2M} \sum_{p=0}^{2M-1} \widehat{C}_p e^{ij\lambda_p}, \tag{2.2}$$

where $\widehat{C}_p = \widehat{f}_{xx,p}^{-1} \widehat{f}_{yx,p}$, and $\widehat{f}_{yx,p}$ and $\widehat{f}_{xx,p}$ are estimates of $f_{yx,p}$ and $f_{xx,p}$ respectively, where for a generic function $g(\lambda)$, g_p denotes $g(\lambda_p)$.

The motivation of the estimator in (2.2), coined by Sims (1974) as HI (Hannan's inefficient) estimator, is threefold. First is the ability to estimate the coefficients $c(j)$ irrespective of the number of lags specified in (1.1), which will be relevant when analyzing the properties of $S_T(\mu)$ defined in (1.3). Second, since there is no gain by exploiting the information on the covariance structure of the errors u_t, as Sims (1974) showed, the HI estimator becomes as efficient as the Generalized Least Squares (GLS)

4

estimator. This motivates the LSE of $c(j)$ given in Robinson (1979), although under stronger assumptions than those we want to impose in this paper.

Finally, the third motivation, which makes the estimate in (2.2) more appealing when the data is long-range dependent is as follows. Assume, for expositional simplicity, that model (1.1) is

$$y_t = \sum_{j=-r}^{q} c(j) x_{t-j} + u_t, \quad t = 1, ..., T, \tag{2.3}$$

where both q and r are finite and known a priory. When the data is short-range dependent, it is known that, under suitable conditions, the LSE is root-T consistent and asymptotically normal. However, under long-range dependence, as Robinson (1994) observed, when the joint long-range dependence in the regressor x_t and error u_t is sufficiently strong, that is the product of the spectral density functions of x_t and u_t is not integrable, the LSE is no longer root-T consistent and more importantly, it loses the central limit theorem property.

Motivated by this observation, Robinson and Hidalgo (1997) showed that a class of frequency-domain weighted LSE, including GLS (with parametric error spectral density function) as a special case, is root-T consistent, asymptotically normal and Gauss-Markov efficient in model (2.3). More generally, their results are also valid when $c(j)$ is known up to a set of parameters θ, that is $c(j) = c(j;\theta)$ for all j, in (1.1). The intuition why the estimator in Robinson and Hidalgo (1997) is root-T consistent and asymptotically normal is because the weighted function possesses a zero sufficiently strong to compensate for the singularity of the spectral density function induced by the joint long-range dependence of x_t and u_t. So, since $f_{xx}^{-1}(\lambda)$ possesses a zero at $\lambda = 0$, see C8 below, we can expect that $\widehat{f}_{xx,p}^{-1}$ becomes (asymptotically) a weighted function satisfying the conditions of Robinson and Hidalgo (1997). In fact, it is shown in Theorem 1 below that the modification of the HI estimator given in (4.1) achieves root-T consistency and asymptotic normality, so that the HI estimator is indeed a

desirable estimator.

It is worth mentioning that Hidalgo and Robinson (1999) have recently provided asymptotic justification of Hannan's (1963) GLS estimate when the spectral density $f_{uu}(\lambda)$ is unknown and both x_t and u_t are, possibly, long-range dependent. However, the last approach in the general framework of model (1.1) seems difficult to implement in empirical studies and based on the previous comments, it is expected that there is no gain in efficiency. It should be noted that when the singularities of $f_{uu}(\lambda)$ and $f_{xx}(\lambda)$ do not coincide, applying Robinson and Hidalgo's (1997) results, under suitable conditions, the LSE will be root-T consistent and asymptotically normal.

3.- STATISTICAL FRAMEWORK AND ASYMPTOTIC PROPERTIES OF THE SPECTRAL DENSITY MATRIX ESTIMATOR

As the estimator in (2.2) and the test for H_0 (and their proofs) employ the estimate of the spectral density $f_{xx}(\lambda)$ and the cross-spectra $f_{yx}(\lambda)$ and $f_{ux}(\lambda)$, it seems appropriate to examine the properties of the spectral density matrix estimate of a, for example, $v-$dimensional covariance stationary linear process z_t. Let $f(\lambda)$ be the spectral density matrix of z_t defined from the relation

$$\gamma(j) = E\left((z_1 - Ez_1)(z_{j+1} - Ez_1)'\right) = \int_{-\pi}^{\pi} f(\lambda) e^{-ij\lambda} d\lambda, \quad j = 0, \pm 1, \pm 2, \dots \quad (3.1)$$

Let z_{tg} denote the gth element of z_t, the autocovariance (spectrum) function of z_{tg} by $\gamma_{gg}(j)(f_{gg}(\lambda))$ and the cross-covariance (cross-spectrum) of z_{tg} and z_{th} by $\gamma_{gh}(j)(f_{gh}(\lambda))$.

We wish to estimate $f(\lambda)$ on the basis of T observations $Z_T = \{z_t, t = 1, ..., T\}$. Writing

$$I(\lambda) = \frac{1}{2\pi T} \left(\sum_{t=1}^{T} z_t e^{it\lambda}\right) \left(\sum_{t=1}^{T} z_t e^{-it\lambda}\right)',$$

6

we estimate f by

$$\widehat{f}(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^{m} I\left(\overline{\lambda}_j + \lambda\right), \qquad (3.2)$$

where $m = [T/4M]$ with M as defined in Section 1 and $\overline{\lambda}_j = (2\pi j)/T$, $j = 0, \pm 1, ..., \pm [T/2]$.

To examine the properties of $\widehat{f}(\lambda)$, let us introduce

Condition C1 For $g = 1, ..., v$, there exist $C_g \in (0, \infty)$, $d_g \in [0, 1/2)$ and $\alpha \in (0, 2]$, such that

$$f_{gg}(\lambda) = C_g \lambda^{-2d_g} \left(1 + O\left(\lambda^\alpha\right)\right) \text{ as } \lambda \to 0+$$

and $f_{gg}(\lambda) > 0$ for all $\lambda \in [0, \pi]$.

Condition C2 For $g = 1, ..., v$, $f_{gg}(\lambda)$ is twice continuously differentiable in any open set outside the origin, and

$$\left| \frac{\partial^j}{\partial \lambda^j} f_{gg}(\lambda) \right| = O\left(\lambda^{-j-2d_g} \right) \text{ as } \lambda \to 0+ \text{ for } j = 1, 2.$$

Let us define the coherence between z_{tg} and z_{th} as $R_{gh}(\lambda) = f_{gh}(\lambda) / \left(f_{gg}^{1/2}(\lambda) f_{hh}^{1/2}(\lambda) \right)$.

Condition C3 For $g < h = 2, ..., v$, $|R_{gh}(\lambda)|$ is twice continuously differentiable in any open set outside the origin and for some $\beta \in (1, 2]$,

$$|R_{gh}(\lambda) - R_{gh}(0)| = O\left(\lambda^{\beta} \right) \text{ as } \lambda \to 0+.$$

Condition C4 $\{z_t\}$ is a covariance stationary linear process defined as

$$z_t = \sum_{j=0}^{\infty} \zeta_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \|\zeta_j\|^2 < \infty,$$

where ζ_0 is the identity matrix, ζ_j are $(v \times v)$ matrices and $\|D\|$ stands for the norm of the matrix D.

Condition C5 $\{\varepsilon_t\}$ is a stochastic process with finite fourth moments, where $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = E(\varepsilon_t \varepsilon_t') = \Xi$ a.s., $E(\varepsilon_{tj_1} \varepsilon_{tj_2} \varepsilon_{tj_3} | \mathcal{F}_{t-1}) = \mu_{3, j_1, j_2, j_3}$ such that $\left| \mu_{3, j_1, j_2, j_3} \right| < \infty$ for all j_1, j_2, j_3 where $\mathcal{F}_t$ is the σ-algebra of events generated by $\varepsilon_s, s \leq t$, and the joint fourth cumulant of $\varepsilon_{t_i j_i}$, $j_i = 1, ..., v$ and $i = 1, ..., 4$ satisfies

$$cum\left(\varepsilon_{t_1 j_1}, \varepsilon_{t_2 j_2}, \varepsilon_{t_3 j_3}, \varepsilon_{t_4 j_4} \right) = \begin{cases} \kappa_{j_1, j_2, j_3, j_4}, & t_1 = t_2 = t_3 = t_4, \\ 0, & otherwise, \end{cases}$$

with $\kappa = \max_{j_i = 1, ..., v, i = 1, ..., 4} \left| \kappa_{j_1, j_2, j_3, j_4} \right| < \infty$.

Condition C6 $|(\partial / \partial \lambda) \eta(\lambda)| = O(|\eta(\lambda)| / \lambda)$ as $\lambda \to 0+$, where

$$\eta(\lambda) = \sum_{j=0}^{\infty} \zeta_j e^{ij\lambda},$$

8

which is twice continuously differentiable in any open set outside the origin, and $f_{gg}^{-1/2}(\lambda)\,\eta_g(\lambda)$, $g = 1, ..., v$, is a non-zero finite vector where $\eta_g(\lambda)$ denotes the gth row of $\eta(\lambda)$.

Condition C7 $M^2/T + 1/M \to 0$.

Prior to stating some results on the spectral matrix estimator, $\widehat{f}(\lambda)$, some discussion about our conditions is in order. Conditions C1-C3 deal with the behaviour of $f(\lambda)$. For frequencies $\lambda \to 0+$, they are the same used elsewhere by, say, Robinson (1995a, b) and thus, his comments apply here, while for frequencies λ in any open set outside the origin, they are standard. Conditions C4 and C5 are restrictive in the linearity they impose, but not otherwise. Although Condition C5 can be weakened to cover more general processes for the innovation ε_t as in Hosoya and Taniguchi (1982), we keep it in its present form to avoid extra notational complications of the rather lengthy proofs of some of the results.

Examples of processes satisfying C1-C5 are as follows. Let ξ_t be a $v-$dimensional unobservable covariance stationary linear process which possesses a continuous and bounded away from zero spectral density matrix and consider the filter

$$z_t = \sum_{j=0}^{\infty} G(j)\,\xi_{t-j}. \tag{3.3}$$

Let $G_g(\lambda)$ denote the gth row of the matrix $G(\lambda) = \sum_{j=0}^{\infty} G(j)\,e^{ij\lambda}$ such that $G_g(\lambda)\,\lambda^{d_g}$ tends to a non-zero finite vector as $\lambda \to 0+$, for $g = 1, ..., v$. For instance, let ξ_t be a stationary invertible vector autoregressive moving average ($VARMA$) process with iid innovations and let each z_{tg} be formed by separate fractional integration of the corresponding ξ_t element, so that

$$G(\lambda) = diag\left(\left(1 - e^{i\lambda}\right)^{-d_1}, ..., \left(1 - e^{i\lambda}\right)^{-d_v}\right).$$

Then C1-C5 hold. In the particular case of z_t being scalar, its spectral density function

9

is

$$f\left(\lambda\right) = \frac{\sigma^2}{2\pi} \left|1 - e^{i\lambda}\right|^{-2d} \left|\frac{\Theta\left(e^{i\lambda}\right)}{\Phi\left(e^{i\lambda}\right)}\right|^2 , \quad -\pi < \lambda \leq \pi, \qquad (3.4)$$

where $0 \leq d < 1/2$, and where $\Theta\left(\lambda\right)$ and $\Phi\left(\lambda\right)$ are the MA and AR polynomials respectively, having no zeroes in or on the unit circle. (3.4) is the familiar fractional autoregressive moving average $(ARFIMA)$ model, see for instance Granger and Joyeux (1980) or Hosking (1981). Another model which exhibits long-range dependence is the fractional Gaussian noise (fgn) process introduced by Mandelbrot and Van Ness (1968), whose spectral density function, see Sinai (1976), is

$$f\left(\lambda\right) = \frac{4\sigma_z^2 \Gamma\left(2d\right)}{\left(2\pi\right)^{3+2d}} \cos\left(\pi d\right) \sin^2\left(\lambda/2\right) \sum_{j=-\infty}^{\infty} \left|j + \frac{\lambda}{2\pi}\right|^{-2-2d} \qquad (3.5)$$

where $\sigma_z^2 = E\left(z_t - E\left(z_t\right)\right)^2 < \infty$ and $\Gamma\left(\cdot\right)$ denotes the gamma function. From (3.4) and (3.5), we observe that the $ARFIMA$ and fgn models do not represent the same process, although their spectral density functions behave as $K\lambda^{-2d}$ as $\lambda \to 0+$, for a generic finite positive constant K. For a review of these models, see Beran's (1994) monograph.

Condition C6, when $\lambda \to 0+$, was assumed elsewhere in Robinson (1995b), so his comments apply here also, while for frequencies λ in any open set outside the origin, the condition is standard in spectral density matrix estimation. The second part of the assumption is not strong, see for instance the comments made after (3.3), once λ^{d_g} is identified as $f_{gg}^{-1/2}$ up to constants. Finally, Condition C7 gives the upper bound on the rate of M to infinity. In particular, M cannot increase faster than $T^{1/2-\delta}$ for any $0 < \delta < 1/2$.

Let $\widehat{f}_{gh}\left(\lambda\right)$ denote the $(g, h)\,th$ element of $\widehat{f}\left(\lambda\right)$ in (3.2).

Proposition 1 *Assuming C1-C7, for $g, h = 1, ..., v$,*

$$E\left(\phi_{gh,p}\right) = O\left(\frac{M \log\left(T\right)}{Tp} \left|f_{gh,p}^*\right|\right) \quad if \quad 1 \leq p \leq M,$$

10

with the convention that $O(a) = o(a) = 0$ if $a = 0$ and where

$$\phi_{gh,p} = \widehat{f}_{gh,p} - f^*_{gh,p} \quad and \quad f^*_{gh}(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^{m} f_{gh}\left(\overline{\lambda}_j + \lambda\right).$$

Proof. The proof is a straightforward application of Proposition 1 of Hidalgo and Yajima (1998) when p satisfies $1 \leq p < [M\delta]$, where $[a]$ denotes the integer part of a, which applies Theorem 2 of Robinson (1995a), and by Proposition 1(c) of Hidalgo and Robinson (1999) when $[M\delta] \leq p \leq M$ for $\delta > 0$. $\qquad\square$

Proposition 2 *Assuming C1-C7, for $a, b, g, h = 1, ..., v$,*

$$M^{-1}T\left|Cov\left(\phi_{gh,p}, \phi_{ab,p}\right)\right| = O\left(f_{gg,p}^{1/2} f_{hh,p}^{1/2} f_{aa,p}^{1/2} f_{bb,p}^{1/2}\right).$$

Proposition 3 *Assuming C1-C7, for $g, h = 1, ..., v$,*

$$\sup_{p=1,...,P}\left|f_{gg,p}^{-1/2} f_{hh,p}^{-1/2} \phi_{gh,p}\right| = O_p\left((PM)^{1/2}/T^{1/2}\right).$$

4.- THE ASYMPTOTIC DISTRIBUTION OF THE HI ESTIMATOR

When analyzing the *HI* estimator in (2.2), and similar to the technical problems encountered in many other non/semi-parametric estimators, since $\widehat{f}_{xx}(0)$ tries to estimate $f_{xx}(0)$ which may be infinity, the *HI* estimator is quite difficult to analyze as it stands. One way to proceed is by trimming the term corresponding to $p = 0$ from (2.2). However, as it will become clear when examining $(A.7)$ in Mathematical Appendix A, the trimming introduces a bias problem that otherwise would not exist if the frequency $\lambda_0 = 0$ was included. Thus, we modify (2.2) to

$$\widehat{c}(j) = \frac{1}{2M} \sum_{p=1}^{2M-1\prime} \widehat{f}_{xx,p}^{-1} \widehat{f}_{yx,p} e^{ij\lambda_p}, \qquad (4.1)$$

where $\sum_{p=1}^{2M-1\prime} a_p e^{ij\lambda_p}$ means $\sum_{p=1}^{2M-1} a_p e^{ij\lambda_p} + a_1$. Intuitively, we have replaced the estimator of C_0 by that of C_1, that is $\widehat{f}_{xx,1}^{-1} \widehat{f}_{yx,1}$.

11

The asymptotic properties of the estimator given in (2.2) were first established by Hannan (1967) for a finite, possibly of unknown order, distributed lag model and Brillinger (1981) for the infinite distributed lag model, when both $f_{xx}(\lambda)$ and $f_{uu}(\lambda)$ are bounded and bounded away from zero. Thus, the aim of this section, due to the possible adverse properties of the LSE, is to show that the estimator defined in (4.1) is root-T consistent and asymptotic normal under the presence of, possibly, long-range dependence. Finally, it is worth noting that due to the cyclical behaviour of $\widehat{c}(j)$, that is, $\widehat{c}(j) = \widehat{c}(j + 2M)$, $c(j)$, $|j| = M + 1, ...$, cannot be estimated, although their contribution is negligible since $\sum_{|j|>M} |c(j)| = O\left(M^{-2+\tau/2}\right) = o\left(T^{-1/2}\right)$ by C9 and C10 respectively given below.

Assume the following:

Condition C8 $\{w_t\} = \{y_t, x_t\}'$ and $\{u_t\}$ are covariance stationary linear processes defined as

$$w_t = \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \|\Phi_j\|^2 < \infty \quad \text{and } u_t = \sum_{j=0}^{\infty} \Phi_j^u \varepsilon_{u,t-j}, \quad \sum_{j=0}^{\infty} |\Phi_j^u|^2 < \infty,$$

where ε_t and $\varepsilon_{u,t}$ satisfy C5, $f_{uu}(\lambda)$ and $f_{ww}(\lambda)$ satisfy C1-C3 and $f_{uu}(\lambda)$ and $\|f_{ww}(\lambda)\|$ are bounded away from zero. Also, $\Phi(\lambda) = \sum_{j=0}^{\infty} \Phi_j e^{ij\lambda}$ and $\Phi^u(\lambda) = \sum_{j=0}^{\infty} \Phi_j^u e^{ij\lambda}$ satisfy C6 and $\{x_t\}$ and $\{u_t\}$ are mutually independent.

Condition C9 $c(|j|) = O\left(|j|^{-3+\tau/2}\right)$ for some $0 < \tau < 1$.

Condition C10 $M^2 T^{-1} + M^{\tau-4} T \to 0$ with τ as in C9.

Condition C8 deals with the requirement of some smoothness of $f_{uu}(\lambda)$ and $f_{ww}(\lambda)$. The requirement of independence between x_t and u_t, as in Robinson and Hidalgo (1997), is necessary for the proof of asymptotic normality. We believe that it might be possible to relax it to some extent, but that will enormously complicate the already technical proof given in Robinson and Hidalgo (1997). This certainly remains an open question. Condition C9 implies that the first derivative of $C(\lambda)$ is Liptchitz continuous with Liptchitz parameter in the interval $(0, 1 - \tau/2)$. Condition C10 strengthens the admissible values of M in C7. Specifically, the rate of increase of M to infinity cannot be slower than $T^{\delta+1/(4-\tau)}$ for arbitrarily small $\delta > 0$.

Theorem 1 *Assuming C8-C10, for any finite collection $j_1, ..., j_q$,*

(i) $T^{1/2}\left(\widehat{c}(j_1) - c(j_1), ..., \widehat{c}(j_q) - c(j_q)\right)' \xrightarrow{d} N\left(0, \Omega = \{\Omega_{j_r j_\ell}\}_{r,\ell=1,..,q}\right)$ *where*

$$\Omega_{j_r j_\ell} = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{xx}^{-1}(\lambda) f_{uu}(\lambda) e^{i(j_r - j_\ell)\lambda} d\lambda$$

which corresponds to the asymptotic covariance between $\widehat{c}(j_r)$ and $\widehat{c}(j_\ell)$.

13

(ii) *Let* $\widehat{f}_{uu,p} = \widehat{f}_{yy,p} - \widehat{f}_{xx,p}^{-1}\left|\widehat{f}_{yx,p}\right|^2$. *A consistent estimator of* $\Omega_{j_r j_\ell}$, $r, \ell = 1, ..., q$, *is*

$$\widehat{\Omega}_{j_r j_\ell} = \frac{1}{2M} \sum_{p=1}^{2M-1\prime} \widehat{f}_{xx,p}^{-1}\widehat{f}_{uu,p}e^{i(j_r - j_\ell)\lambda_p}.$$

Theorem 1 indicates that the results, obtained by Hannan (1967) and Brillinger (1981) for short-range dependent data, hold the same under long-range dependence. It is important to observe the following; because $f_{xx}(\lambda) \sim K\lambda^{-2d_x}$ and $f_{uu}(\lambda) \sim K\lambda^{-2d_u}$ as $\lambda \to 0+$ by C8, where $0 \leq d_x, d_u < 1/2$, the asymptotic covariance structure of $\widehat{c}(j)$, when taken as a process indexed by j, behaves as that of a long-range dependent process with $d = d_u - d_x$. In particular, when $f_{xx}(\lambda) = Kf_{uu}(\lambda)$ for all $\lambda \in (-\pi, \pi]$, $\widehat{c}(j)$ is, asymptotically, an $iid\,(0, K^{-1})$ Gaussian process. Generally, $\widehat{c}(j)$ has a spectral density $(2\pi)^{-1}f_{xx}^{-1}(\lambda)f_{uu}(\lambda)$.

5.- A NONPARAMETRIC CAUSALITY TEST

If in (2.3) r was a known finite constant, a Wald test for H_0 could easily be implemented from the results of Theorem 1. However, when there is no known parameterization of the model, e.g. of $c(j)$ in (1.1), in terms of a finite set of parameters, θ, the results given in Theorem 1 cannot be implemented in a straightforward manner. In this framework, one way of testing for Granger causality might be by fitting an $AR(P)$ model to w_t where P increases slowly with T as in Lütkepohl and Poskitt (1996). However, their results depend heavily on the assumption that the data is short-range dependent. Specifically, their Assumptions 1 and 2, and that the sample covariance of the data converges to the population one at the rate $T^{-1/2}$ are not necessarily the case under long-range dependence. Moreover, the asymptotic distribution of the sample covariance may not be a normal random variable, see Taqqu (1975) or Hannan (1976). Thus, we adopt the approach given in (1.3). Given the estimates of

14

$c\left(j\right)$ in (4.1), consider

$$S_T\left(\mu\right) = \mathrm{Re}\left(\frac{1}{M}\sum_{p=1}^{[M\mu]}\left(\sum_{j=-M+1}^{0}\widehat{c}\left(j-1\right)e^{-ij\lambda_p}\right)\right). \tag{5.1}$$

$S_T\left(\mu\right)$ forms the basis for the hypothesis testing in (1.2) as follows. In Corollary 1 below, we show that under H_0, $T^{1/2}S_T\left(\mu\right) \overset{weakly}{\Longrightarrow} \widetilde{B}\left(\mu\right)$, a Gaussian process with covariance structure

$$K\left(\mu_1,\mu_2\right) = \frac{1}{4\pi}\int_0^{\pi\min(\mu_1,\mu_2)} f_{xx}^{-1}\left(\lambda\right)f_{uu}\left(\lambda\right)d\lambda, \tag{5.2}$$

where $"\overset{weakly}{\Longrightarrow}"$ denotes weakly convergence in $\mathcal{D}\left[0,1\right]$ equipped with the Skorohod metric.

Since the function $K\left(\mu,\mu\right)$ is nondecreasing and nonnegative, $\widetilde{B}\left(\mu\right)$ admits the representation $B\left(K\left(\mu,\mu\right)\right)$ in distribution, where $B\left(\mu\right)$ is the standard Brownian motion in $\left[0,1\right]$. This representation, Corollary 1 below and the continuous mapping theorem yield

$$\sup_{\mu\in[0,1]}\left|T^{1/2}S_T\left(\mu\right)\right| \Longrightarrow \sup_{\mu\in[0,K(1,1)]}\left|B\left(\mu\right)\right| = K^{1/2}\left(1,1\right)\sup_{\mu\in[0,1]}\left|B\left(\mu\right)\right| \quad \text{in law.}$$

Let us introduce

Condition C8' C8 is satisfied and in addition $\left|x_t\right|^4$ and $\left|u_t\right|^4$ are uniformly integrable.

Observe that a sufficient condition for the second part of $C8'$ is that $E\left|x_t\right|^{4+\delta} + E\left|u_t\right|^{4+\delta} < \infty$ for some $\delta > 0$.

Corollary 1 *Assuming C8', C9 and C10, under H_0,*

$$T^{1/2}S_T\left(\mu\right) \overset{weakly}{\Longrightarrow} \widetilde{B}\left(\mu\right),$$

in $\mathcal{D}\left[0,1\right]$ with the Skorohod metric and covariance structure $K\left(\mu_1,\mu_2\right)$ given in (5.2).

15

Let $\widehat{K}(\mu, \mu)$ be the consistent estimate of $K(\mu, \mu)$,

$$\widehat{K}(\mu, \mu) = \frac{1}{4M} \sum_{p=1}^{[M\mu]} \widehat{f}_{xx,p}^{-1} \widehat{f}_{uu,p}.$$

Then, Corollary 1 and the comments made above are useful for testing H_0. For example, the Kolmogorov-Smirnov test based on $T^{1/2}S_T(\mu)$ would reject the null if $\sup \left\{ \widehat{K}^{-1/2}(1,1) \left| T^{1/2}S_T(\mu) \right|, \mu \in [0,1] \right\}$ exceeds an appropriate critical value obtained from the boundary crossing probabilities of a Brownian motion, which are readily available on the unit interval. More generally, see Koul and Stute (1999), as

$$\widehat{K}^{-1/2}(1,1) T^{1/2}S_T \left(\left(\widehat{K}(\mu, \mu) \right)^{-1}(t) \right) \stackrel{weakly}{\Longrightarrow} B(\mu)$$

where $\left(\widehat{K}(\mu, \mu) \right)^{-1}(t) = \inf \left\{ \mu \in [0,1], \widehat{K}(\mu, \mu) \geq t \right\}$, the limiting distribution of any continuous functional of $\widehat{K}^{-1/2}(1,1) T^{1/2}S_T \left(\left(\widehat{K}(\mu, \mu) \right)^{-1}(t) \right)$ can be obtained from the distribution of the corresponding functional of $B(\mu)$ on $[0,1]$.

As in other problems involving testing, one key area of interest is to know the properties of the tests under contiguous or Pitman alternatives. To this end, let us introduce

$$H_a : \sum_{j=-\infty}^{0} c(j-1) \cos(\pi j \lambda) = T^{-1/2} h(\pi \lambda),$$

where $h(\vartheta)$ is a continuous function in $[0, \pi]$ such that $0 < |h(\vartheta)|$ in a set $\Delta \subset [0, \pi]$ with positive Lebesgue measure.

Corollary 2 *Assuming C8', C9 and C10, under H_a,*

$$T^{1/2}S_T(\mu) \stackrel{weakly}{\Longrightarrow} \widetilde{B}(\mu) + \int_0^\mu h(\pi \lambda) d\lambda. \tag{5.3}$$

Corollary 2 thus indicates that the test has power against contiguous alternatives that converge to the null at the rate $T^{-1/2}$. It therefore has no zero asymptotic relative efficiency compared to rival parametric tests based on a correct parameterization of

16

(1.1). From Corollary 2, we can easily show that the test is consistent. For that purpose, consider

$$H_1 : \sum_{j=-\infty}^{0} c\,(j-1)\cos\,(\pi j\lambda) = h\,(\pi\lambda)\,.$$

From Corollary 1 we can then immediately prove that under H_1

$$T^{1/2}S_T\,(\mu) - T^{1/2}\int_0^\mu h\,(\pi\lambda)\,d\lambda \overset{weakly}{\Longrightarrow} \widetilde{B}\,(\mu)\,,$$

and thus the consistency of the test.

To finish this section, it is worth noting that as the Kolmogorov-Smirnov test has somewhat poor finite sample properties, we may alternatively attempt to bootstrap the statistic $T^{1/2}S_T\,(\mu)$, in a similar way to Hidalgo and Kreiss (1999), using a combination of the Wild and Moving Block Boostraps. Specifically, let $\ell = \ell\,(T)$ be a number such that $\ell^{-1} + \ell T^{-1} \to 0$. Consider $L = T - \ell + 1$ groups, such that the qth group is formed from the observations $(y_q, x_q)'\,, ..., (y_{\ell+q-1}, x_{\ell+q-1})'$. Then the bootstrap analogue of $T^{1/2}S_T\,(\mu)$ is

$$S_T^*\,(\mu) = L^{-1/2}\sum_{q=1}^{L}\mathrm{Re}\left(\frac{1}{M}\sum_{p=1}^{[M\mu]}\left(\sum_{j=-M+1}^{0}\ell^{1/2}\,(\widehat{c}_q\,(j-1) - \widehat{c}\,(j-1))\,e^{-ij\lambda_p}\right)\right)\vartheta_q$$

where ϑ_q are $iid\,(0,1)$ random variables and $\widehat{c}_q\,(j-1)$ is identical to (4.1) but using only the observations from the qth group.

6.- IMPLEMENTATION OF THE TEST AND ITS EXTENSION TO MULTIVARIATE DATA

In empirical studies to implement the test and estimate the parameters $c\,(j)$, we face the problem of how to choose M or m (recall that $M = [T/4m]$). Based on the approach adopted to estimate the parameters $c\,(j)$, the choice of M can be regarded as that of the number of leads/lags in the model

$$y_t = \sum_{j=-M=-[T/4m]}^{M=[T/4m]} c\,(j)\,x_{t-j} + \widetilde{u}_t,\ \ t = 1, ..., T. \tag{6.1}$$

17

One standard criterion for choosing M, and thus m, in (6.1) is the minimization of the Akaike's (1974) AIC criterion

$$\log \widehat{\sigma}^2_M + 2M/T \tag{6.2}$$

where, using Kolmogorov's formula,

$$\widehat{\sigma}^2_M = \exp\left(\frac{1}{4\pi M} \sum_{j=1}^{2M-1} \log\left(\widehat{f}_{uu}(\lambda_j)\right)\right),$$

that is, the estimator of the one-step-prediction error based on the residuals $y_t - \sum_{j=-M=-[T/4m]}^{M=[T/4m]} \widehat{c}(j) x_{t-j}$.

The motivation of the AIC criterion is twofold. First, it gives an approximately unbiased measure for the prediction error and secondly, it provides an asymptotic unbiased "estimate" of the Kullback-Leibler information (using its Whittle's approximation),

$$KL(f, g) = 2T \log(2\pi) + \frac{T}{2\pi} \int_{-\pi}^{\pi} \left(\log g(\lambda) + \frac{f(\lambda)}{g(\lambda)}\right) d\lambda,$$

where f and g are the true and candidate spectral density functions for a given data set.

An alternative criterion to the AIC given in (6.2), which can be used, is based on the choice of $m = [T/4M]$ which minimizes the Cross-Validation criterion

$$\mathcal{Q}_m = \sum_{j=1}^{[T/2m]} \left\{\log \widehat{f}^m_{uu,(j)}(\lambda_j) + \frac{I_{\widehat{u}\widehat{u},m}(\lambda_j)}{\widehat{f}^m_{uu,(j)}(\lambda_j)}\right\}, \tag{6.3}$$

where $I_{\widehat{u}\widehat{u},m}(\lambda_\ell)$ is the periodogram of the residuals in (6.1) and $\widehat{f}^m_{uu,(j)}(\lambda_j) = \widehat{f}^m_{yy,(j)}(\lambda_j) - \left|\widehat{C}(\lambda_j)\right|^2 \widehat{f}^m_{xx,(j)}(\lambda_j)$ is the one-leave-out estimate of the spectral density function of u_t given in part (ii) of Theorem 1: that is

$$\widehat{f}^m_{uu,(j)}(\lambda_j) = \frac{1}{2m+1} \sum_{\ell=-m,\ell\neq 0}^{m} \left(I_{yy}\left(\overline{\lambda}_\ell + \lambda_j\right) - \left|\widehat{C}(\lambda_j)\right|^2 I_{xx}\left(\overline{\lambda}_\ell + \lambda_j\right)\right).$$

The motivation of the above criteria $\mathcal{Q}_m$ in (6.3) is that as shown by, for example, Hurvich and Beltrão (1990), the expected value of $KL\left(f_{uu}, \widehat{f}_{uu}\right)$, where $\widehat{f}_{uu}$ is a nonparametric estimate of f_{uu}, is asymptotically equivalent to the Cross-Validation criterion (apart from the constant $2T \log (2\pi)$) for the estimate $\widehat{f}_{uu}$. Moreover, it is worth mentioning that the two criteria (6.2) and (6.3) are asymptotically equivalent, see for instance Stone (1977) or Kavalieris (1989).

Now, let us describe how the results of Sections 4 and 5 are extended to a $p = p_1 + p_2$ dimensional covariance stationary vector $w_t = (y'_t, x'_t)'$. Assume that w_t admits the $AR(\infty)$ representation

$$A\left(L\right) w_t = \sum_{j=0}^{\infty} A_j w_{t-j} = \varepsilon_t, \ t = 1, 2..., T, \tag{6.4}$$

where ε_t is a $p-$dimensional martingale difference sequence and A_0 is the identity matrix. As was argued in Section 1, the null hypothesis, H_0, that $y_t \nRightarrow x_t$, is equivalent to testing whether the $p_1 \times p_2$ matrices $c\left(j\right)$ are zero for $j < 0$ in the infinite distributed lag representation of (6.4)

$$y_t = \sum_{j=-\infty}^{\infty} c\left(j\right) x_{t-j} + u_t,$$

where, by construction, $E\left[u_t \left| x_s, -\infty < s < \infty\right.\right] = 0$.

In this case, the HI estimator of $c\left(j\right)$ is defined as

$$\widehat{c}\left(j\right) = \frac{1}{2M} \sum_{p=1}^{2M-1\prime} \widehat{f}_{yx,p} \widehat{f}_{xx,p}^{-1} e^{ij\lambda_p},$$

whereas the test for H_0 is given by

$$S_T\left(\mu\right) = \mathrm{Re}\left(\frac{1}{M} \sum_{p=1}^{[M\mu]} vec\left(\sum_{j=-M+1}^{0} \widehat{c}\left(j-1\right) e^{-ij\lambda_p}\right)\right).$$

We then achieve the following results.

Theorem 2 *Assuming C9, C10 and an obvious extension of C8 to multivariate data, for any finite collection $j_1, ..., j_q$,*

(i) $T^{1/2} \left(vec \left(\widehat{c} \left(j_1 \right) - c \left(j_1 \right) \right), ..., vec \left(\widehat{c} \left(j_q \right) - c \left(j_q \right) \right) \right)' \overset{d}{\to} N \left(0, \Omega = \{ \Omega_{j_r j_\ell} \}_{r, \ell = 1, .., q} \right)$

where

$$\Omega_{j_r j_\ell} = (2\pi)^{-1} \int_{-\pi}^{\pi} \left(f_{xx}^{-1} \left(-\lambda \right) \otimes f_{uu} \left(\lambda \right) \right) e^{i(j_r - j_\ell)\lambda} d\lambda$$

which corresponds to the asymptotic covariance matrix between $vec \left(\widehat{c} \left(j_r \right) \right)$ and $vec \left(\widehat{c} \left(j_\ell \right) \right)$.

(ii) Let $\widehat{f}_{uu,p} = \widehat{f}_{yy,p} - \widehat{f}_{yx,p} \widehat{f}_{xx,p}^{-1} \widehat{f}_{xy,p}$. A consistent estimator of $\Omega_{j_r j_\ell}$, $r, \ell = 1, ..., q$, is

$$\widehat{\Omega}_{j_r j_\ell} = \frac{1}{2M} \sum_{p=1}^{2M-1'} \left(\widehat{f}_{xx,-p}^{-1} \otimes \widehat{f}_{uu,p} \right) e^{i(j_r - j_\ell)\lambda_p}.$$

Proof. The proof of this theorem or any other result in this section follows by routine extension to those of Sections 4 and 5 and whose details can be seen in Hidalgo (1998).

Corollary 3 *Assuming C9, C10 and an obvious extension of C8' to multivariate data, under H_0,*

$$T^{1/2} S_T \left(\mu \right) \overset{weakly}{\Longrightarrow} vec \left(\widetilde{B} \left(\mu \right) \right)$$

in $\mathcal{D}^{p_1 \times p_2} [0, 1]$ with the Skorohod metric, and where $vec \left(\widetilde{B} \left(\mu \right) \right)$ is a $p_1 \times p_2$-Gaussian process with covariance structure

$$K \left(\mu_1, \mu_2 \right) = \frac{1}{4\pi} \int_0^{\pi \min(\mu_1, \mu_2)} \left(f_{xx}^{-1} \left(-\lambda \right) \otimes f_{uu} \left(\lambda \right) \right) d\lambda.$$

From Corollary 3, the test for H_0 can be implemented in a similar fashion to that employed in Section 5. Finally, to examine the properties of the tests under contiguous alternatives, introduce

$$H_a : \sum_{j=-\infty}^{0} c \left(j - 1 \right) \cos \left(\pi j \lambda \right) = T^{-1/2} h \left(\pi \lambda \right),$$

where $h \left(\vartheta \right)$ is a continuous function in $[0, \pi]$ such that $0 < \| h \left(\vartheta \right) \|$ in a set $\Delta \subset [0, \pi]$ with positive Lebesgue measure.

Corollary 4 *Assuming the same conditions of Corollary 3, under H_a,*

$$T^{1/2} S_T(\mu) \overset{weakly}{\Longrightarrow} vec\left(\widetilde{B}(\mu) + \int_0^\mu h(\pi\lambda)\, d\lambda\right).$$

7.- CONCLUSIONS

In this paper we have first shown that the asymptotic properties of Hannan's (1967) HI estimator hold the same under long-range dependence, making the estimation procedure a very desirable one in view of the possible adverse properties of the LSE, as shown in Robinson (1994). Secondly, we have proposed a nonparametric Granger-causality test when the data possibly exhibits long-range dependence. By means of spectral analysis, we have shown the asymptotic properties of the test and examined its implementation. More importantly, we have shown that the test has power against contiguous alternatives converging to the null at the rate $T^{-1/2}$. That is, although the test is nonparametric, it has an asymptotic relative efficiency greater than zero when compared to parametric tests based on a correct specification of the model.

REFERENCES

[1] Akaike, H. (1974), A new look at the statistical model identification. IEEE Trans. Autom. Control, 19, 716-723.

[2] Beran, J. (1994), Statistics for long-memory processes. Chapman and Hall. London.

[3] Billingsley, P. (1968), Convergence of Probability Measures. John Wiley, New York.

[4] Bloomfield, P. (1973), An exponential model for the spectrum of a scalar time series. Biometrika, 60, pp. 217-226.

[5] Brillinger, D.R. (1981), Time Series, Data Analysis and Theory. San Francisco: Holden-Day.

[6] Geweke, J. (1982), Measurement for linear dependence and feedback between multiple time series. Journal of the American Statistical Association, 77, pp. 303-323.

[7] Geweke, J. (1986), The superneutrality of money in the USA. An interpretation of its evidence. Econometrica, 54, pp. 1-21.

[8] Granger, C.W.J. (1969), Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37, pp. 424-438.

[9] Granger, C.W.J. and R. Joyeux (1980), An introduction to long memory time series and fractional differencing. Journal of Time Series Analysis, 1, pp. 15-30.

[10] Hannan, E.J. (1963), Regression for time series. In M. Rosenblatt (ed.), Time Series Analysis, pp. 17-37. New York: Wiley.

[11] Hannan, E.J. (1967), The estimation of a lagged regression relation. Biometrika, 54, pp. 409-418.

[12] Hannan, E.J. (1976), The asymptotic distribution of serial covariances. Annals of Statistics, 4, pp. 396-399.

[13] Hannan, E.J. and R.D. Terrell (1973), Multiple equation systems with stationary errors. Econometrica, 41, pp. 299-320.

[14] Hidalgo, J. (1998), Estimation of distributed lag models with long-range dependence. Mimeo, London School of Economics.

[15] Hidalgo, J. and J.-P. Kreiss (1999), Boostrap specification tests for covariance stationary processes. Preprint.

[16] Hidalgo, J. and P.M. Robinson (1999), Adapting to unknown error autocorrelation in regression with long memory. Preprint.

[17] Hidalgo, J. and Y. Yajima (1998), Semiparametric estimation of the long-range parameter. Preprint.

[18] Hosking, J. (1981), Fractional differencing. Biometrika, 68, pp. 165-176.

[19] Hosoya, Y. (1977), On the Granger condition for non-causality. Econometrica, 45, pp. 1735-1736.

[20] Hosoya, Y. (1991), The decomposition and measurement of the interdependency between second-order stationary processes. Probability Theory and Related Fields, 88, pp. 429-444.

[21] Hosoya, Y. (1995), Causal relations between possibly nonstationary time-series and related statistical inference. Invited paper of the $7th$ World Congress of the Econometric Society.

[22] Hosoya, Y. and M. Taniguchi (1982), A central limit theorem for stationary processes and the parameter estimation of linear processes. Annals of Statistics, 10, pp. 132-153.

[23] Hurvich, C.M. and K.I. Beltrão (1990), Cross-validatory choice of a spectrum estimate and its connections with AIC. Journal of Time Series Analysis, 11, 121-137.

[24] Kavalieris, L. (1989), The estimation of the order of an Autoregression using residuals and cross-validation. Journal of Time Series Analysis, 10, pp. 271-281.

[25] Koul, H.L. and W. Stute (1999), Nonparametric model checks for time series. Annals of Statistics, 27, pp.204-236.

[26] Lütkepohl, H. and D.S. Poskitt (1996), Testing for causation using infinite order vector autoregressive processes. Econometric Theory, 12, pp. 61-87.

[27] Mandelbrot, B.B. and J.W. Van Ness (1968), Fractional Brownian motions, fractional noises and applications. SIAM Review, 10, pp. 422-437.

[28] Robinson, P.M. (1979), Distributed lag approximation to linear time-invariant systems. Annals of Statistics, 7, pp. 507-515.

[29] Robinson, P.M. (1991), Automatic frequency domain inference on semiparametric and nonparametric models. Econometrica, 59, pp.1329-1363.

[30] Robinson, P.M. (1994), Time series with strong dependence. In C.A. Sims, ed., Advances in Econometrics: Sixth World Congress, Vol. 1, pp. 47-95. Cambridge University Press, Cambridge.

[31] Robinson, P.M. (1995a), Log-periodogram regression for time series with long range dependence. Annals of Statistics, 23, pp. 1048-1072.

[32] Robinson, P.M. (1995b), Gaussian semiparametric estimation of long-range dependence. Annals of Statistics, 23, pp. 1630-1661.

[33] Robinson, P.M. and J. Hidalgo (1997), Time series regression with long range dependence. Annals of Statistics, 25, pp. 77-104.

[34] Sims, C.A. (1972), Money, income and causality. American Economic Review, 62, pp. 540-552.

[35] Sims, C.A. (1974), Distributed lags. In Frontiers of Quantitative Economics, Vol. II (M.D. Intrilligator and D.A. Kendrick, eds.). North Holland: Amsterdam.

[36] Sims, C.A., J.H. Stock, and M.W. Watson (1990), Inference in linear time series models with some unit roots. Econometrica, 58, pp. 113-144.

[37] Sinai, Y.G. (1976), Self-similar probability distribution. Theo. Probability and its Applications, 21, pp. 64-80.

[38] Stone, M. (1977), An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. Journal of the Royal Statis. Socie., Ser. B, 39, 44-47.

[39] Taqqu, M.S. (1975), Weak convergence to fractional Brownian motion and the Rosenblatt process. Zeitschrift für Wahrsch. und Verwandte Gebiete, 31, pp. 287-302.

[40] Toda, H.Y. and P.C.B. Phillips (1993), Vector autoregression and causality. Econometrica, 61, pp. 1367-1393.

MATHEMATICAL APPENDIX A

Henceforth, $\sum_j$ and $\sum_{j,k}$ denote $\sum_{j=-m}^m$ and $\sum_{j,k=-m}^m$, respectively, and K denotes a finite positive constant.

Proof of Proposition 2

The $Cov\left(\phi_{gh,p}, \phi_{ab,p}\right)$ is

$$\frac{1}{\left(2m+1\right)^2 4\pi^2 T^2} \sum_{j,k} \sum_{t,r,s,u=1}^T \left\{ \gamma_{ga}\left(t-s\right) \gamma_{hb}\left(r-u\right) + \gamma_{gb}\left(t-u\right) \gamma_{ha}\left(r-s\right) \right.$$
$$\left. + cum\left(z_{tg}, z_{rh}, z_{sa}, z_{ub}\right) \right\} e^{i(t-r)\overline{\lambda}_{2mp+j} - i(s-u)\overline{\lambda}_{2mp+k}}. \tag{A.1}$$

We deal with the contribution of the first and third term in braces, the second being similarly handled to the first. The contribution of the first term of $(A.1)$ is by (3.1)

$$\frac{1}{\left(2m+1\right)^2 4\pi^2 T^2} \sum_{j,k} \left\{ \int_{-\pi}^\pi H\left(\theta_1\right) H\left(\overline{\lambda}_{k-j} - \theta_1\right) f_{ga}\left(\theta_1 + \overline{\lambda}_{2mp+j}\right) d\theta_1 \right.$$
$$\left. \times \int_{-\pi}^\pi H\left(-\theta_2\right) H\left(\theta_2 + \overline{\lambda}_{j-k}\right) f_{hb}\left(\theta_2 - \overline{\lambda}_{2mp+k}\right) d\theta_2 \right\}, \tag{A.2}$$

where $H\left(\omega\right) = \sum_{t=1}^T e^{-it\omega}$. We study the first factor inside the braces of $(A.2)$, the second being identical. Adding and subtracting $f_{ga}\left(\overline{\lambda}_{2mp+j}\right) \int_{-\pi}^\pi H\left(\theta_1\right) H\left(\overline{\lambda}_{k-j} - \theta_1\right) d\theta_1$, that factor is

$$\int_{-\pi}^\pi H\left(\theta_1\right) H\left(\overline{\lambda}_{k-j} - \theta_1\right) \left(f_{ga}\left(\theta_1 + \overline{\lambda}_{2mp+j}\right) - f_{ga}\left(\overline{\lambda}_{2mp+j}\right) \right) d\theta_1$$
$$+ f_{ga}\left(\overline{\lambda}_{2mp+j}\right) \int_{-\pi}^\pi H\left(\theta_1\right) H\left(\overline{\lambda}_{k-j} - \theta_1\right) d\theta_1,$$

whose first term is $O\left(\Delta_{ga,j}\left(2mp+j\right)^{-1} T \log\left(2mp+k\right)\right)$ by Theorem 2(c,d) of Robinson $(1995a)$, where $\Delta_{ga,j} = f_{gg}^{1/2}\left(\overline{\lambda}_{2mp+j}\right) f_{aa}^{1/2}\left(\overline{\lambda}_{2mp+j}\right)$, whereas the second term is 0 if $j \neq k$ or $2\pi T f_{ga}\left(\overline{\lambda}_{2mp+j}\right)$ if $j = k$.

26

Thus, as $T \to \infty$, the modulus of $(A.2)$ is bounded by

$$\frac{K}{(2m+1)^2} \sum_{j,k} \Delta_{ga,j} \Delta_{hb,k} \frac{1}{2mp+j} \frac{\log^2 (2mp+k)}{2mp+k}$$

$$+ \frac{K}{(2m+1)^2} \sum_j \Delta_{ga,j} \Delta_{hb,j} \frac{\log (2mp+j)}{2mp+j}$$

$$+ \frac{1}{(2m+1)^2} \sum_j \left| f_{ga} \left(\overline{\lambda}_{2mp+j} \right) \right| \left| f_{hb} \left(\overline{\lambda}_{2mp+j} \right) \right| . \tag{A.3}$$

The first term of $(A.3)$ is $O\left((mp)^{-2} \log^2 (mp) \Upsilon_p\right)$ where $\Upsilon_p = f_{aa,p}^{1/2} f_{gg,p}^{1/2} f_{bb,p}^{1/2} f_{hh,p}^{1/2}$, because, by C1-C3, $\Delta_{ga,j}$ and $\Delta_{hb,j}$ satisfy the conditions of Lemma 1 (see Appendix B). Similarly, by Lemma 1, the second term of $(A.3)$ is $O\left(m^{-2} p^{-1} \log (mp) \Upsilon_p\right)$, whereas the third term is

$$\frac{\Upsilon_p |R_{ga,p}| |R_{hb,p}|}{2m+1} \left(1 + o(1)\right),$$

since $\overline{\lambda}_{2mp} = \lambda_p$ and $|f_{ga,p}| |f_{hb,p}| = \Upsilon_p |R_{ga,p}| |R_{hb,p}|$ satisfies the conditions of Lemma 1. Thus, the first term of $(A.1)$ is

$$\frac{\Upsilon_p |R_{ga,p}| |R_{hb,p}|}{2m+1} \left(1 + o(1)\right) + O\left(\Upsilon_p \left\{ \frac{\log (mp)}{m^2 p} + \frac{\log^2 (mp)}{(mp)^2} \right\}\right).$$

Next, proceeding as above, the second term of $(A.1)$ is

$$O\left(\Upsilon_p \frac{\log^2 (mp)}{(mp)^2}\right).$$

Thus, we are left with the third term of $(A.1)$. Writing $\widetilde{\eta}(\lambda) = \eta(\lambda) \Xi^{1/2}$ and after straightforward calculations, we deduce that it is bounded, in absolute value, by

$$\frac{K}{T^2 (2m+1)^2} \sum_{s_\ell} \left| \sum_{j,k} \int_{\Pi^3} \widetilde{\eta}_{gs_1} \left(\overline{\lambda}_{2mp+j} \right) \widetilde{\eta}_{as_2} \left(-\overline{\lambda}_{2mp+j} \right) \widetilde{\eta}_{hs_3} \left(-\overline{\lambda}_{2mp+k} \right) \widetilde{\eta}_{bs_4} \left(\overline{\lambda}_{2mp+k} \right) \right.$$

$$\left. \times H\left(\overline{\lambda}_{2mp+j} - \lambda \right) H\left(\lambda - \mu - \overline{\lambda}_{2mp+j} \right) H\left(\mu - \xi - \overline{\lambda}_{2mp+k} \right) H\left(\overline{\lambda}_{2mp+k} + \xi \right) d\lambda d\mu d\xi \right| \tag{A.4}$$

27

where $\widetilde{\eta}_{ab}(\lambda)$ is the indicated element of $\widetilde{\eta}(\lambda)$, the first sum is over $s_\ell = 1, ..., v$ and $\ell = 1, ..., 4$, and $\prod = [-\pi, \pi]$. By Cauchy-Schwarz inequality, a typical term in $(A.4)$ is bounded by $KT^{-2}(2m+1)^{-2}$ times

$$\left\{ \int_{\prod^2} \left| \sum_j \widetilde{\eta}_{gs_1}\left(\overline{\lambda}_{2mp+j}\right) \widetilde{\eta}_{as_2}\left(-\overline{\lambda}_{2mp+j}\right) H\left(\overline{\lambda}_{2mp+j} - \lambda\right) H\left(\lambda - \mu - \overline{\lambda}_{2mp+j}\right) \right|^2 d\lambda d\mu \right. \quad (A.5)$$

$$\left. \times \int_{\prod^2} \left| \sum_k \widetilde{\eta}_{hs_3}\left(-\overline{\lambda}_{2mp+k}\right) \widetilde{\eta}_{bs_4}\left(\overline{\lambda}_{2mp+k}\right) H\left(\mu - \xi - \overline{\lambda}_{2mp+k}\right) H\left(\overline{\lambda}_{2mp+k} + \xi\right) \right|^2 d\mu d\xi \right\}^{1/2}.$$

The first integral in $(A.5)$ is

$$\int_{\prod^2} \left(\sum_{j_1} \widetilde{\eta}_{gs_1}\left(\overline{\lambda}_{2mp+j_1}\right) \widetilde{\eta}_{as_2}\left(-\overline{\lambda}_{2mp+j_1}\right) H\left(\overline{\lambda}_{2mp+j_1} - \lambda\right) H\left(\lambda - \mu - \overline{\lambda}_{2mp+j_1}\right) \right.$$

$$\left. \times \sum_{j_2} \widetilde{\eta}_{gs_1}\left(-\overline{\lambda}_{2mp+j_2}\right) \widetilde{\eta}_{as_2}\left(\overline{\lambda}_{2mp+j_2}\right) H\left(\lambda - \overline{\lambda}_{2mp+j_2}\right) H\left(\mu - \lambda + \overline{\lambda}_{2mp+j_2}\right) \right) d\lambda d\mu.$$

Because

$$\int_{\prod} H\left(\lambda - \mu - \overline{\lambda}_{2mp+j_1}\right) H\left(\mu - \lambda + \overline{\lambda}_{2mp+j_2}\right) d\mu = 2\pi H\left(\overline{\lambda}_{j_2 - j_1}\right)$$

and

$$\int_{\prod} H\left(\overline{\lambda}_{2mp+j_1} - \lambda\right) H\left(\lambda - \overline{\lambda}_{2mp+j_2}\right) d\lambda = 2\pi H\left(\overline{\lambda}_{j_1 - j_2}\right),$$

it follows, by C6, that

$$(A.5) = O\left(\left(T^2 \sum_{j,k} \Delta_{ga,j} \Delta_{ag,k} K\left(\overline{\lambda}_{j-k}\right) \sum_{j,k} \Delta_{hb,j} \Delta_{bh,k} K\left(\overline{\lambda}_{j-k}\right) \right)^{1/2} \right) = O\left(T^2 m \Upsilon_p\right)$$

where $K(\lambda)$ is the Fèjer kernel, because

$$(2m+1)^{-2} \sum_{j,k} \Delta_{ga,j} \Delta_{ga,k} K\left(\overline{\lambda}_{j-k}\right) = (T/m) f_{aa,p} f_{gg,p} \left(1 + o(1)\right).$$

So $(A.4) = O\left(\Upsilon_p M/T\right)$, which concludes that $\left| Cov\left(\phi_{gh,p}, \phi_{ab,p}\right) \right| = O\left(\Upsilon_p M/T\right)$. $\quad \square$

Proof of Proposition 3

Write $\delta_{gh,p} = f_{gg,p}^{-1/2} f_{hh,p}^{-1/2}$. By elementary inequalities and because $\sup_\ell |a_\ell| \le \sum_\ell |a_\ell|$,

$$\sup_{p=1,\ldots,P} \left| \delta_{gh,p} \phi_{gh,p} \right| \le \sum_{p=1}^{P} \delta_{gh,p} \left| E\left(\phi_{gh,p}\right) \right| + \sup_{p=1,\ldots,P} \delta_{gh,p} \left| \phi_{gh,p} - E\left(\phi_{gh,p}\right) \right|.$$

Since $\sup_\ell |a_\ell| \le \left(\sum_\ell a_\ell^2\right)^{1/2}$, the second term on the right is $O_p\left((PM)^{1/2}/T^{1/2}\right)$ by Proposition 2 whereas, by Proposition 1, the first term on the right is bounded by

$$K \frac{M \log T}{T} \sum_{p=1}^{P} \frac{1}{p} \delta_{gh,p} \left| f_{gh,p}^* \right|.$$

But $f_{gg}^{1/2}\left(\lambda\right) f_{hh}^{1/2}\left(\lambda\right) \left| R_{gh}\left(\lambda\right) \right|$ satisfies the conditions of Lemma 1, so by straightforward arguments, we conclude the proof of the Proposition. $\square$

Proof of Theorem 1

We establish only (i), omitting for the sake of brevity the much easier proof of (ii). By Wold device, it suffices to show that for any finite $q > 0$,

$$T^{1/2} \sum_{\ell=1}^{q} \Psi_\ell \left(\widehat{c}\left(j_\ell\right) - c\left(j_\ell\right) \right) \xrightarrow{d} N\left(0, \left(2\pi\right)^{-1} \sum_{\ell_1,\ell_2=1}^{q} \Psi_{\ell_1} \int_{-\pi}^{\pi} f_{xx}^{-1}\left(\lambda\right) f_{uu}\left(\lambda\right) e^{i\left(j_{\ell_1} - j_{\ell_2}\right)\lambda} d\lambda \; \Psi_{\ell_2} \right)$$

where Ψ_ℓ are constants such that $\sum_{\ell=1}^{q} \Psi_\ell^2 = 1$. To that end, introduce $\alpha_p = \widehat{f}_{yx,p} - E\widehat{f}_{yx,p}$, $A_p = E\widehat{f}_{yx,p}$, $\beta_p = \widehat{f}_{xx,p} - E\widehat{f}_{xx,p}$ and $B_p = E\widehat{f}_{xx,p}$. A typical element on the left of the last displayed expression, say, $\widehat{c}\left(j\right) - c\left(j\right)$ is

$$\frac{1}{2M} \sum_{p=1}^{2M-1\prime} \left(B_p^{-1}\alpha_p - B_p^{-2} A_p \beta_p \right) e^{ij\lambda_p} \tag{A.6}$$

$$+\frac{1}{2M} \sum_{p=1}^{2M-1\prime} B_p^{-1} A_p e^{ij\lambda_p} - c\left(j\right) \tag{A.7}$$

$$-\frac{1}{2M} \sum_{p=1}^{2M-1\prime} \left(B_p^{-2}\alpha_p\beta_p - \left(\beta_p + B_p\right)^{-1} \left(\alpha_p + A_p\right) B_p^{-2} \beta_p^2 \right) e^{ij\lambda_p}. \tag{A.8}$$

29

We show below that

$$(A.6) + (A.7) + (A.8) = \frac{1}{2M} \sum_{p=1}^{2M-1} f_{xx,p}^{-1} \widehat{f}_{ux,p} e^{ij\lambda_p} + L(j) \qquad (A.9)$$

where $L(j) = O_p \left((TM)^{-1/2} \log M + M^{-2+\tau/2} + T^{-3/2} M^2 + T^{-1} M \right)$ uniformly in j. Thus, by C10,

$$T^{1/2} \sum_{\ell=1}^{q} \Psi_\ell \left(\widehat{c}(j_\ell) - c(j_\ell) \right) = \frac{T^{1/2}}{2M} \sum_{p=1}^{2M-1} f_{xx,p}^{-1} \widehat{f}_{ux,p} \left(\sum_{\ell=1}^{q} \Psi_\ell e^{ij_\ell \lambda_p} \right) + o_p(1).$$

But, by similar arguments to those of Theorem 1 of Robinson and Hidalgo (1997), the first term on the right of the last displayed equation converges in distribution to

$$N \left(0, (2\pi)^{-1} \sum_{\ell_1, \ell_2=1}^{q} \Psi_{\ell_1} \int_{-\pi}^{\pi} f_{xx}^{-1}(\lambda) f_{uu}(\lambda) e^{i(j_{\ell_1} - j_{\ell_2})\lambda} d\lambda \; \Psi_{\ell_2} \right).$$

So, to complete the proof we need to show our claim in $(A.9)$.

We begin examining $(A.8)$. This term is

$$-\frac{1}{2M} \sum_{p=1}^{2M-1\prime} B_p^{-2} \alpha_p \beta_p e^{ij\lambda_p} + \frac{1}{2M} \sum_{p=1}^{2M-1\prime} C_p^* B_p^{-2} \beta_p^2 e^{ij\lambda_p}$$

$$+\frac{1}{2M} \sum_{p=1}^{2M-1\prime} \left((\beta_p + B_p)^{-1} (\alpha_p + A_p) - C_p^* \right) B_p^{-2} \beta_p^2 e^{ij\lambda_p}, \qquad (A.10)$$

where $C_p^* = f_{xx,p}^{*-1} f_{yx,p}^*$ with $f_{v_1 v_2, p}^* = (2m+1)^{-1} \sum_j f_{v_1 v_2} \left(\lambda_p + \overline{\lambda}_j \right)$.

The first term of $(A.10)$ is $O_p(M/T)$ by Cauchy-Schwarz and Markov inequalities because by Proposition 2, $f_{xx,p}^{-1/2} f_{yy,p}^{-1/2} \alpha_p$ and $f_{xx,p}^{-1} \beta_p$ are $O_p \left(M^{1/2} T^{-1/2} \right)$ with $z_{th} = x_t$ and $z_{tg} = y_t$, and $z_{tg} = z_{th} = x_t$ there, $\left(B_p - f_{xx,p}^* \right) = o\left(\left| f_{xx,p}^* \right| \right) = O(f_{xx,p})$ by Proposition 1 and Lemma 1 respectively and $|C_p| < K$ by C9. By similar arguments and since $\left| C_p^* \right| < K$ by Lemma 2 (see Appendix B), the second term is also $O_p(M/T)$. The third term of $(A.10)$ is $O_p \left(T^{-3/2} M^2 \right)$ uniformly in j, because by standard linearization arguments and Propositions 1 to 3,

$$\sup_{p=1,\dots,M} \left| \widehat{f}_{xx,p}^{-1} \widehat{f}_{yx,p} - C_p^* \right| = O_p \left(MT^{-1/2} \right).$$

30

Thus, since $\sum_{p=1}^{2M-1\prime} a_p e^{ij\lambda_p} = \sum_{p=1}^{2M-1} a_p e^{ij\lambda_p} + a_1$, we conclude that

$$(A.8) = O_p\left(\frac{M^2}{T^{3/2}} + \frac{M}{T}\right).$$

Next the bias term $(A.7)$. Because $\lambda_{2M} = 2\pi$ and $\sum_{p=0}^{2M-1} e^{ij\lambda_p} = \sum_{p=1}^{2M} e^{ij\lambda_p}$,

$$
\begin{aligned}
c(j) &= \frac{1}{2M} \sum_{p=1}^{2M} \left(\sum_{\ell=-M}^{M} c(\ell) e^{-i\ell\lambda_p} \right) e^{ij\lambda_p} \\
&= \frac{1}{2M} \sum_{p=1}^{2M-1} \left(\sum_{\ell=-M}^{M} c(\ell) e^{-i\ell\lambda_p} \right) e^{ij\lambda_p} + \frac{1}{2M} \sum_{\ell=-M}^{M} c(\ell),
\end{aligned}
$$

so that $(A.7)$ is by C9

$$\frac{1}{2M}\left\{ B_1^{-1} A_1 - \sum_{\ell=-\infty}^{\infty} c(\ell) \right\} + \frac{1}{2M} \sum_{p=1}^{2M-1} \left(B_p^{-1} A_p - \sum_{\ell=-\infty}^{\infty} c(\ell) e^{-i\ell\lambda_p} \right) e^{ij\lambda_p} + O\left(\frac{1}{M^{2-\tau/2}}\right).$$
$$(A.11)$$

Next,

$$
\begin{aligned}
B_p^{-1} A_p &= f_{xx,p}^{*-1} f_{yx,p}^* + \left(B_p^{-1} A_p - f_{xx,p}^{*-1} f_{yx,p}^* \right) \\
&= f_{xx,p}^{*-1} f_{yx,p}^* + f_{xx,p}^{*-1}\left(A_p - f_{yx,p}^* \right) - C_p^* f_{xx,p}^{*-1}\left(B_p - f_{xx,p}^* \right) + O\left(\frac{M^2 \log^2 T}{T^2 p^2}\right)
\end{aligned}
$$

by Proposition 1, standard linearization arguments and because $|C(\lambda)| < K$ by C9.
Thus, replacing $\sum_{\ell=-\infty}^{\infty} c(\ell) e^{-i\ell\lambda_p}$ by its definition $f_{xx,p}^{-1} f_{yx,p}$, $(A.11)$ is

$$\frac{1}{2M}\left\{ f_{xx,1}^{*-1} f_{yx,1}^* - \sum_{\ell=-\infty}^{\infty} c(\ell) \right\} + \frac{1}{2M}\left(f_{xx,1}^{*-1}\left(A_1 - f_{yx,1}^* \right) - C_1^* f_{xx,1}^{*-1}\left(B_1 - f_{xx,1}^* \right) \right)$$

$$+\frac{1}{2M} \sum_{p=1}^{2M-1} \left(f_{xx,p}^{*-1} f_{yx,p}^* - f_{xx,p}^{-1} f_{yx,p} \right) e^{ij\lambda_p} \qquad (A.12)$$

$$+\frac{1}{2M} \sum_{p=1}^{2M-1} \left(f_{xx,p}^{*-1}\left(A_p - f_{yx,p}^* \right) - C_p^* f_{xx,p}^{*-1}\left(B_p - f_{xx,p}^* \right) \right) e^{ij\lambda_p} + O\left(\frac{1}{M^{2-\tau/2}} + \frac{M\log^2 T}{T^2}\right).$$

By Lemma 2 the third term of $(A.12)$ is bounded, in absolute value, by

$$\frac{K}{M^2} \sum_{p=1}^{2M-1} \frac{1}{p} \widetilde{C}_p + O\left(\frac{1}{M^{2-\tau/2}}\right) = O\left(\frac{1}{M^{2-\tau/2}}\right)$$

31

since $\left| \widetilde{C}_p = \sum_{\ell=-\infty}^{\infty} \ell c(\ell) e^{-i\ell \lambda_p} \right| < K$ by C9. The fourth term of $(A.12)$ is $O\left(T^{-1} \log T\right)$ by Proposition 1 and since by Lemma 2, $\left| C_p^* \right| < K$, so is the second term of $(A.12)$ by the same argument.

The first term of $(A.12)$ is

$$\frac{1}{2M} \left(f_{xx}^{*-1}(\lambda_1) f_{yx}^*(\lambda_1) - \sum_{\ell=-\infty}^{\infty} c(\ell) e^{-i\ell \lambda_1} \right) + \frac{1}{2M} \sum_{\ell=-\infty}^{\infty} c(\ell) \left(e^{-i\ell \lambda_1} - 1 \right) = O\left(M^{-2}\right),$$

because $\left| e^{-i\ell \lambda_1} - 1 \right| \leq K \ell M^{-1}$ as $\lambda_1 = \pi/M$ and $\sum_{\ell=-\infty}^{\infty} |\ell c(\ell)| < K$ by C9, so the second term on the left of the last displayed expression is $O\left(M^{-2}\right)$, whereas the first term is $O\left(M^{-2}\right)$ by Lemma 2. Thus, we conclude that

$$(A.7) = O\left(\frac{1}{M^{2-\tau/2}} + \frac{\log T}{T} \right).$$

To finish the proof we need to examine $(A.6)$. Replacing B_p and $A_p B_p^{-1}$ by $f_{xx,p}^*$ and C_p^*, respectively,

$$(A.6) = \frac{1}{2M} \sum_{p=1}^{2M-1} f_{xx,p}^{*-1} \left(\alpha_p - C_p^* \beta_p \right) e^{ij\lambda_p} + O_p\left(\frac{M^{1/2}}{T^{1/2}} \frac{\log T}{T} \right)$$

after standard calculations, because by Proposition 2 $f_{yy,p}^{-1/2} f_{xx,p}^{-1/2} \alpha_p$ and $f_{xx,p}^{-1} \beta_p$ are $O_p\left(M^{1/2} T^{-1/2}\right)$, by Proposition 1 $B_p - f_{xx,p}^* = O\left(\left| f_{xx,p}^* \right| M \log(T) / (Tp)\right)$, for example, and by Cauchy inequality.

Next, since $\left(f_{xx,p}^{-1} f_{xx,p}^* - 1 \right) = O\left(p^{-1}\right)$ by Lemma 1, as $f_{xx}(\lambda)$ satisfies the conditions there,

$$(A.6) = \frac{1}{2M} \sum_{p=1}^{2M-1} f_{xx,p}^{-1} \left(\alpha_p - C_p \beta_p - \widehat{f}_{ux,p} \right) e^{ij\lambda_p} + \frac{1}{2M} \sum_{p=1}^{2M-1} f_{xx,p}^{-1} \widehat{f}_{ux,p} e^{ij\lambda_p}$$

$$+ \frac{1}{2M} \sum_{p=1}^{2M-1} f_{xx,p}^{-1} \left(C_p - C_p^* \right) \beta_p e^{ij\lambda_p} + O_p\left(\frac{\log M}{T^{1/2} M^{1/2}} \right) \tag{A.13}$$

by Proposition 2. The third term on the right of $(A.13)$ is $O_p\left(M^{-1+\tau/2} T^{-1/2}\right)$ by Lemma 2 and because $\sup_{p=1,\ldots,M} \left| f_{xx,p}^{-1} \beta_p \right| = O_p\left(T^{-1/2} M\right)$ by Proposition 3.

To complete the proof, we are left with the first term on the right of $(A.13)$. Because by C8 $E\widehat{f}_{ux,p} = 0$,

$$\alpha_p - C_p\beta_p - \widehat{f}_{ux,p} = \left(\widehat{f}_{yx,p} - C_p\widehat{f}_{xx,p} - \widehat{f}_{ux,p}\right) - E\left(\widehat{f}_{yx,p} - C_p\widehat{f}_{xx,p} - \widehat{f}_{ux,p}\right), \quad (A.14)$$

whose first term on the right, by definition of $\widehat{f}_{yx,p}$ and $\widehat{f}_{xx,p}$ and (1.1), is

$$\frac{1}{(2m+1)\,2\pi T}\sum_j\left[\sum_{\ell=-\infty}^{\infty}\left(c_\ell\sum_{t=1}^{T}x_{t-\ell}e^{-it\overline{\lambda}_{2mp+j}}\right)\left(\sum_{s=1}^{T}x_s e^{is\overline{\lambda}_{2mp+j}}\right)\right] \quad (A.15)$$

$$-\left(\sum_{\ell=-\infty}^{\infty}c_\ell e^{-i\ell\lambda_p}\right)\frac{1}{(2m+1)\,2\pi T}\sum_j\left[\left(\sum_{t=1}^{T}x_t e^{-it\overline{\lambda}_{2mp+j}}\right)\left(\sum_{s=1}^{T}x_s e^{is\overline{\lambda}_{2mp+j}}\right)\right].$$

Observe that

$$\frac{1}{2\pi T}\left(\sum_{t=1}^{T}x_{t-\ell}e^{-it\lambda}\right)\left(\sum_{s=1}^{T}x_s e^{is\lambda}\right)$$

is the cross-periodogram of $\{x_t\}$ and $\{x_{t-\ell}\}$, whose cross-spectrum is $f_{xx}(\lambda)\,e^{-i\ell\lambda}$.

Thus, using $(A.15)$ and by C9, the right side of $(A.14)$ is

$$\sum_{|\ell|\leq M}c_\ell e^{-i\ell\lambda_p}\frac{1}{(2m+1)}\sum_j\left(\left(e^{-i\ell\overline{\lambda}_j}-1\right)\left(I_{xx}\left(\overline{\lambda}_{2mp+j}\right)-EI_{xx}\left(\overline{\lambda}_{2mp+j}\right)\right)\right)+O_p\left(M^{-2+\tau/2}\right).$$

Because $\left|1-e^{-i\ell\overline{\lambda}_j}\right|\leq K\ell M^{-1}\,(j/m)$ and by routine extension of Proposition 2,

$$\frac{1}{2m+1}\sum_j\left(\frac{j}{m}\right)\left(I_{xx}\left(\overline{\lambda}_{2mp+j}\right)-EI_{xx}\left(\overline{\lambda}_{2mp+j}\right)\right)=O_p\left(M^{1/2}T^{-1/2}f_{xx,p}\right),$$

then

$$(A.6)=\frac{1}{2M}\sum_{p=1}^{2M-1}f_{xx,p}^{-1}\widehat{f}_{ux,p}e^{ij\lambda_p}+O\left(\frac{1}{M^{2-\tau/2}}+\frac{\log M}{T^{1/2}M^{1/2}}\right).$$

This concludes the proof of part (i) and the theorem. $\qquad\square$

Proof of Corollary 1

Because under H_0 $c(j)=0$ for all $j<0$, by Theorem 1 $T^{1/2}\widehat{c}(j)=T^{1/2}L(j)+a_j$ where

$$a_j=\frac{T^{1/2}}{2M}\sum_{p=1}^{2M-1}f_{xx,p}^{-1}\widehat{f}_{ux,p}e^{ij\lambda_p}$$

33

suppressing reference to T in a_j or $L(j)$. Thus, with the change of subindex $-j$ by j,

$$T^{1/2}S_T(\mu) = \mathrm{Re}\left(\sum_{j=0}^{M-1}\left(a_{-j-1}+T^{1/2}L(-j-1)\right)\frac{1}{M}\sum_{p=1}^{[M\mu]}e^{ij\lambda_p}\right)$$

$$= \left(a_{-1}+T^{1/2}L(-1)\right)\mu + \sum_{j=1}^{M-1}\left(a_{-j-1}+T^{1/2}L(-j-1)\right)\frac{\sin(\pi j\mu)}{\pi j}(1+o(1))$$

since $\mathrm{Re}\left(M^{-1}\sum_{p=1}^{[M\mu]}e^{ij\lambda_p}\right)\underset{M\to\infty}{\to}(\pi j)^{-1}\sin(\pi j\mu)$ uniformly in $\mu\in[0,1]$. Now,

$$\sup_{\mu}\left|\sum_{j=1}^{M-1}T^{1/2}L(-j-1)\frac{\sin(\pi j\mu)}{j}\right| = o_p(1)$$

since by Theorem 1 $T^{1/2}L(-j-1)=O_p\left(M^{-1/2}\log M\right)$ and $\sum_{j=1}^{M-1}j^{-1}=O(\log M)$.

So the behaviour of (5.1) is governed by

$$a_{-1}\mu + \sum_{j=1}^{k-1}a_{-j-1}\frac{\sin(\pi j\mu)}{\pi j} + \sum_{j=k}^{M-1}a_{-j-1}\frac{\sin(\pi j\mu)}{\pi j},$$

where k is a fixed but large constant. The proof is thus completed if

$$G_k(\mu) = a_{-1}\mu + \sum_{j=1}^{k-1}a_{-j-1}\frac{\sin(\pi j\mu)}{\pi j} \text{ converges to a Gaussian process indexed by } \mu,$$

$$\hspace{13cm}(\text{A.16})$$

$$\left|EG_k^2(\mu)-\frac{1}{4\pi}\int_0^{\pi\mu}f_{xx}^{-1}(\lambda)f_{uu}(\lambda)d\lambda\right| < \varepsilon \hspace{3cm}(\text{A.17})$$

for any arbitrary $\varepsilon > 0$ and k large enough, and

$$\sum_{j=k}^{M-1}a_{-j-1}\frac{\sin(\pi j\mu)}{\pi j} \text{ is small uniformly in } M \text{ and } \mu. \hspace{2cm}(\text{A.18})$$

We begin with assertion (A.18). For $0 < s_1 < s_2 < M$,

$$\left|\sum_{j=s_1}^{s_2}a_{-j-1}\frac{\sin(\pi j\mu)}{\pi j}\right|^2 \leq \left|\sum_{j=s_1}^{s_2}a_{-j-1}\frac{e^{\pi j\mu}}{\pi j}\right|^2 \leq \frac{2}{\pi^2}\sum_{v=1}^{s_2-s_1}\left|\sum_{j=s_1}^{s_2-v}\frac{a_{-j-1}a_{-j-1-v}}{j(j+v)}\right|. \hspace{0.5cm}(\text{A.19})$$

34

Let $\omega_{|j-\ell|}$ denote the $(j, \ell)\, th$ element of Ω. Then,

$$
E\left|\sum_{j=s_1}^{s_2-v} \frac{a_{-j-1} a_{-j-1-v}}{j(j+v)}\right|^2 = \sum_{j_1,j_2=s_1}^{s_2-v} E\left(\frac{a_{-j_1-1} a_{-j_1-1-v} a_{-j_2-1} a_{-j_2-1-v}}{j_1(j_1+v) j_2(j_2+v)}\right)
$$

$$
\leq \sum_{j_1,j_2=s_1}^{s_2-v} \left|\frac{\omega_v^2 + \omega_{|j_1-j_2|}^2 + \omega_{|j_1-j_2-v|}\omega_{|j_1-j_2+v|}}{j_1(j_1+v) j_2(j_2+v)}\right| (1+o(1)),
$$

because $(a_{-j})_{j=1,\ldots,k} \xrightarrow{d} N(0, \Omega)$ by Theorem 1, and by the uniform integrability of a_{-j}^4 by C8', $E\left|a_{-j_1} a_{-j_2} a_{-j_3} a_{-j_4}\right|$ converges to the corresponding expectation of the limit distribution of $\{a_{-j}\}$. The latter follows because if $X_t \xrightarrow{d} X$ and $|X_t|^4$ is uniformly integrable then $E|X_t|^4 \to E|X|^4$. Now choosing $s_1 = 2^\ell$ and $s_2 = 2^{\ell+1}$, by Cauchy inequality, the expectation of the left side of $(A.19)$ is bounded by

$$
K \sum_{v=1}^{2^\ell} \left|\sum_{j_1,j_2=2^\ell}^{2^{\ell+1}-v} \frac{\omega_v^2 + \omega_{|j_1-j_2|}^2 + \omega_{|j_1-j_2-v|}\omega_{|j_1-j_2+v|}}{j_1(j_1+v) j_2(j_2+v)}\right|^{1/2}
$$

$$
\leq K \sum_{v=1}^{2^\ell} \left|\sum_{j_1=2^\ell}^{2^{\ell+1}-v} \frac{1}{j_1^2(j_1+v)^2}\right|^{1/2} + K \sum_{v=1}^{2^\ell} \left|\sum_{j_1=2^\ell, \, j_1<j_2}^{2^{\ell+1}-v} \frac{\omega_v^2 + \omega_{|j_1-j_2|}^2 + \omega_{|j_1-j_2-v|}\omega_{|j_1-j_2+v|}}{j_1(j_1+v) j_2(j_2+v)}\right|^{1/2},
$$

by triangle inequality. But $\omega_v \sim K v^{-1+\alpha}$, where $\alpha = d_x - d_u < 1$ (recall the comments after Theorem 1), so the right side of the last displayed inequality, after standard calculations, is bounded by

$$
K \sum_{v=1}^{2^\ell} \left|\sum_{j_1=2^\ell}^{2^{\ell+1}-v} \frac{1}{j_1^2(j_1+v)^2}\right|^{1/2} + K \sum_{v=1}^{2^\ell} v^{-1+\alpha} \left|\sum_{j_1=2^\ell}^{2^{\ell+1}-v} \frac{1}{j_1(j_1+v)}\right| \leq K\left(2^{-\ell/2} + 2^{-\ell(1-\alpha)}\right).
$$

Let $T_\ell = \max_{0 \leq \mu \leq 1} \left|\sum_{j=2^\ell}^{2^{\ell+1}} a_{-j-1} \frac{\sin(\pi j \mu)}{\pi j}\right|$. Then,

$$
E\left(T_\ell^2\right) \leq K\left(2^{-\ell/2} + 2^{-\ell(1-\alpha)}\right),
$$

and thus, with probability greater than $1 - K\left(2^{-\ell/2} + 2^{-\ell(1-\alpha)}\right)$, $T_\ell \leq K\left(2^{-\ell/4} + 2^{-\ell(1-\alpha)/2}\right)$.

Now choose $k = 2^n$, so that $(A.18)$ is, in absolute value, bounded by

$$
\sum_{\ell=n}^{\lceil \log_2(M) \rceil + 1} T_\ell \leq K\left(2^{-n(1-\alpha)/2} + 2^{-n/4}\right),
$$

and choosing n large enough, we conclude the proof of $(A.18)$.

Assertion $(A.16)$. From Theorem 1, the finite dimensional distributions of $G_k(\mu)$ converge to a normal random variable. To finish, we need to prove the tightness condition which, by Billingsley's (1968) Theorem 12.3, it suffices to check

$$E\left(G_k(\mu) - G_k(\mu + v)\right)^4 \leq K\left|v\right|^2. \qquad (A.20)$$

But that follows because for all j and $u, v \in [0, \pi]$,

$$\left(\frac{\sin\left(ju\right)}{j} - \frac{\sin\left(j\left(u+v\right)\right)}{j}\right)^2$$

$$= \left(\frac{\sin\left(ju\right)\cos\left(ju+\xi jv\right)}{j} - \frac{\sin\left(j\left(u+v\right)\right)\cos\left(j\left(u+v\right)+\xi ju\right)}{j}\right)v$$

where $\xi \in (0, 1)$, which implies that

$$E\left(a_{-j-1}\left(\frac{\sin\left(2\pi j\mu\right)}{j} - \frac{\sin\left(2\pi j\left(\mu+v\right)\right)}{j}\right)\right)^4 \leq K\frac{\left|v\right|^2}{j^2},$$

and thus $(A.20)$. That concludes the proof of $(A.16)$.

Finally $(A.17)$ follows since by $(A.16)$, with the convention that $\sin\left(ax\right)/x = a$ for $x = 0$,

$$E\left(G_k^2\left(\mu\right)\right) = \sum_{j_1, j_2=0}^{k-1} \omega_{|j_1-j_2|} \frac{\sin\left(\pi j_1 \mu\right)}{\pi j_1} \frac{\sin\left(\pi j_2 \mu\right)}{\pi j_2} \underset{k\to\infty}{\to} \frac{1}{4\pi} \int_0^{\pi\mu} f_{xx}^{-1}\left(\lambda\right) f_{uu}\left(\lambda\right) d\lambda, \quad \text{and}$$

$$E\left(G_k\left(\mu_1\right) G_k\left(\mu_2\right)\right) \underset{k\to\infty}{\to} \frac{1}{4\pi} \int_0^{\pi\min(\mu_1,\mu_2)} f_{xx}^{-1}\left(\lambda\right) f_{uu}\left(\lambda\right) d\lambda. \qquad \square$$

Proof of Corollary 2

First, the left side of (5.3) is

$$\mathrm{Re}\left(\frac{T^{1/2}}{M} \sum_{p=1}^{[M\mu]} \left(\sum_{j=-M+1}^{0} \left(\widehat{c}\left(j-1\right) - c\left(j-1\right)\right) e^{-ij\lambda_p}\right)\right.$$

$$\left. +\frac{1}{M} \sum_{p=1}^{[M\mu]} \left(T^{1/2} \sum_{j=-M+1}^{0} c\left(j-1\right) e^{-ij\lambda_p}\right)\right).$$

36

But, from Corollary 1, the first term converges weakly to $\widetilde{B}(\mu)$, whereas the second term, under H_a, is $M^{-1}\sum_{p=1}^{[M\mu]} h(\lambda_p) + o(1)$ by C9 and C10. From here the conclusion of the Corollary is immediate. $\qquad\square$

MATHEMATICAL APPENDIX B

Let $g(\lambda)$ be a function which satisfies

A1.- $g(\lambda) = K\lambda^{-\alpha}(1 + O(\lambda^{\eta}))$ as $\lambda \to 0+$, where $\alpha \le 2$ and $\eta \in (0, 2]$.

A2.- $g(\lambda)$ is twice continuously differentiable in any open set outside the origin.

A3.- $\frac{\partial^j}{\partial \lambda^j}\log(g(\lambda)) = O(\lambda^{-j})$ as $\lambda \to 0+$ for $j = 1, 2$.

Lemma 1 *Assuming A1-A3,*

$$\left| \frac{1}{2m+1}\sum_j \left(g\left(\overline{\lambda}_{j+2mp}\right) - g\left(\overline{\lambda}_{2mp}\right)\right) \right| = O\left(p^{-1}g\left(\overline{\lambda}_{2mp}\right)\right). \qquad (B.1)$$

Proof. By A2, A3 and the mean value theorem

$$g\left(\overline{\lambda}_{j+2mp}\right) = g\left(\overline{\lambda}_{2mp}\right) + \left(\frac{2\pi j}{T}\right)g'\left(\overline{\lambda}_{\xi j+2mp}\right),$$

where $\xi = \xi(j) \in (0, 1)$, and by A1-A3

$$g^{-1}\left(\overline{\lambda}_{2mp}\right)g'\left(\overline{\lambda}_{\xi j+2mp}\right) = O\left(\left(\frac{T}{\xi j + 2mp}\right)^{1+\alpha}\left(\frac{mp}{T}\right)^{\alpha}\right).$$

So the left side of $(B.1)$ is bounded by

$$\frac{Kg\left(\overline{\lambda}_{2mp}\right)}{2m+1}\sum_j \left|\frac{j}{T}\right|\left|\frac{T}{\xi j + 2mp}\right|^{1+\alpha}\left|\frac{mp}{T}\right|^{\alpha} = O\left(p^{-1}g\left(\overline{\lambda}_{2mp}\right)\right)$$

since $p \ge 1$ and $K^{-1} < |mp/(\xi j + 2mp)| < K$. $\square$

Lemma 2 *Assuming C8 and C9,*

$$f^{*-1}_{xx,p}f^{*}_{yx,p} - f^{-1}_{xx,p}f_{yx,p} = \frac{4\pi}{Mp}\widetilde{C}_p + O\left(\frac{1}{M^{2-\tau/2}}\right). \qquad (B.2)$$

Proof. The left side of $(B.2)$ is

$$\frac{1}{2m+1}\sum_j \left(C\left(\lambda_p + \overline{\lambda}_j\right) - C\left(\lambda_p\right)\right)f_{xx}\left(\lambda_p + \overline{\lambda}_j\right)\left(\frac{1}{2m+1}\sum_j f_{xx}\left(\lambda_p + \overline{\lambda}_j\right)\right)^{-1}$$

$$= \frac{1}{2m+1}\sum_j \left(C\left(\lambda_p + \overline{\lambda}_j\right) - C\left(\lambda_p\right)\right)\nu_p^{-1} \qquad (B.3)$$

$$+ \frac{1}{2m+1}\sum_j \left(C\left(\lambda_p + \overline{\lambda}_j\right) - C\left(\lambda_p\right)\right)\left(f^{-1}_{xx}\left(\lambda_p\right)f_{xx}\left(\lambda_p + \overline{\lambda}_j\right) - 1\right)\nu_p^{-1},$$

where $\nu_p = f_{xx}^{-1}(\lambda_p)(2m+1)^{-1}\sum_j f_{xx}(\lambda_p + \overline{\lambda}_j)$. But, by Taylor expansion and C8,

$$\left| f_{xx}^{-1}(\lambda_p) f_{xx}(\lambda_p + \overline{\lambda}_j) - 1 \right| = \frac{2M}{T}\frac{j}{p}\left(1 + O\left(\frac{M}{T}\frac{j}{p}\right)\right), \tag{B.4}$$

whereas, by C9, $C\left(\lambda_p + \overline{\lambda}_j\right) - C(\lambda_p)$ is

$$(2\pi)\frac{j}{T}\left(\sum_{\ell=-\infty}^{\infty} \ell c(\ell) e^{-i\ell\lambda_p} + O\left(\left|\frac{j}{T}\right|^{1-\tau/2}\right)\right) = (2\pi)\frac{j}{T}\left(\widetilde{C}_p + O\left(\left|\frac{j}{T}\right|^{1-\tau/2}\right)\right), \tag{B.5}$$

where $\widetilde{C}_p = \sum_{\ell=-\infty}^{\infty} \ell c(\ell) e^{-i\ell\lambda_p}$. Thus, adding and subtracting $C\left(\lambda_p + \overline{\lambda}_j\right) - C(\lambda_p)$ into the first term on the right of $(B.3)$ and using $(B.4)$, $(B.5)$ and $\sum_{j=-m}^{m} j = 0$, the right side of $(B.3)$ is $4\pi (Mp)^{-1}\widetilde{C}_p + O\left(M^{-2+\tau/2}\right)$. $\qquad\square$