

PRELIMINARY VERSION. DO NOT QUOTE.

DISTRIBUTION FREE GOODNESS-OF-FIT TESTS FOR
LINEAR PROCESSES[‡]

MIGUEL A. DELGADO

JAVIER HIDALGO

Universidad Carlos III

London School of Economics

CARLOS VELASCO

Universidad Carlos III

February 15, 2002

Abstract

This article proposes goodness-of-fit tests based on a linear transformation of Barlett's T_p – process, which converges weakly to the standard Brownian motion under the null hypothesis, in the lines suggested by Khmaladze (1981). We show that the proposed test has non trivial power in the direction of contiguous alternatives converging to the null at the parametric rate $n^{-1/2}$. A Monte Carlo study illustrates the performance of the test with moderate sample sizes, showing that it works fairly well, in terms of size accuracy and power. We also show that this testing procedure can be applied to test the specification of general nonstationary integrated processes, as well as lack-of-autocorrelation of the innovations in general dynamic models.

Keywords and Phrases: Goodness-of-fit; linear processes; empirical processes; contiguous alternatives; martingale decomposition.

SHORT TITLE: Goodness-of-fit tests for linear processes.

IMS 2000 Subject Classification: 62G10.

*Research funded by the Spanish Dirección General de Enseñanza Superior (DGES) reference number BEC2001-1270 and by the Economic and Social Research Council (ESRC) reference number: R000238212.

†Corresponding Address: Miguel A. Delgado, Departamento de Estadística y Econometría, University of Carlos III, Madrid 126-128, Getafe, 28903 Madrid, Spain.

1. INTRODUCTION

This paper proposes distribution-free goodness-of-fit tests for linear, possibly nonstationary, processes. The null hypothesis is composite, specifying the functional form of the spectral distribution function, which depends on some unknown parameters. Tests statistics are continuous functionals of a linear transformation of Barlett's T_p - *process* with estimated parameters, which converges under the null hypothesis to the standard Brownian motion on $[0, 1]$. Therefore, for the most common testing functionals, such as the Kolmogorov-Smirnov and Cramer-von Misses criterions, critical values are already available.

Existing goodness-of-fit tests can be roughly classified in two alternative categories: those based on sample autocorrelations and those based on the sample spectral distribution function. Pormanteau-test statistics (e.g. Box and Pierce, 1978) are standardized versions of (weighted) partial sums of squared residual autocorrelations, and enjoy some practical advantages that justify their popularity. These tests were originally designed to test for lack of autocorrelation of finite order, and they are computationally attractive when residuals have a closed form, i.e. when the specification to be tested is an autoregressive process of finite order. Their practical advantages are sensibly diminished when innovations do not have a closed expression in terms of observed data. In some situations, such as specifications exhibiting long memory, e.g. fractional autoregressive moving average (FARIMA) models, Pormanteau tests are still not justified. On the other hand, these tests only employ information on a small number of residual autocorrelations, which may result in a loss of power in the direction of certain alternatives. Indeed, according to Hong (1996), these tests are unable to detect contiguous alternatives converging to the null at the rate $n^{-1/2}$, where n is the sample size. Furthermore, the outcome of the test can be very sensitive to the choice of the number of components in the partial sum. This number plays the role of a smoothing parameter which is chosen by the practitioner, with a large degree of arbitrariness in the absence of a proper theory justifying such a choice in finite samples.

Goodness-of-fit tests have a long standing tradition. Grenander and Rosenblatt (1957) proved the usefulness of the periodogram for testing the white noise simple hypothesis. Barlett (1954, 1966) introduced two processes, the T_p and U_p , for testing the simple hypothesis of a general specification. The U_p - *process* is the integrated periodogram of the observed data suitably standardized. The T_p - *process* can be interpreted as an estimate of the spectral distribution of innovations resulting from a particular linear specification, which is based on the integrated periodogram weighted by

the spectral density function specified under the null hypothesis. Therefore, neither the T_p or U_p processes require to compute the innovations of the specified linear process. The asymptotic weak behavior of these empirical processes has been studied under quite general conditions by Dahlhaus (1985), see also the more recent references of Anderson (1993) and Kokoszka and Mikosh (1997). The weak behavior of spectral U_p - processes indexed by general classes of functions has been developed by Dahlhaus (1988). Tests based on continuous functionals of U_p or T_p processes use information on autocorrelations at all lags, and do not depend on the choice of a smoothing number. However they have a serious disadvantage: they are not asymptotically pivotal when the null hypothesis is composite, depending on unknown parameters, as it is generally the case with usual goodness-of-fit tests for composite hypotheses (eg. Durbin, 1973; or Stute, 1997, for a more recent reference).

There have been several attempts to justify goodness-of-fit testing of composite hypothesis based on spectral distribution estimates. Anderson (1991, 1997) proposed an asymptotically distribution free test for $AR(p)$ processes using the Cramér-von Misses criterion, where the U_p -process is represented by an infinite weighted sum of independent standard normal random variables, via a principal components decomposition of the covariance kernel. The asymptotic distribution of the Cramér-von Misses statistic is tabulated by truncation of the infinite sum and estimation of the weights using numerical methods, like Imhof, - see Durbin, Knott and Taylor (1975) and Stute (1997) for related problems. This procedure seems, however, difficult to generalize to other processes. Velilla (1994) suggested an asymptotically pivotal test for $ARMA(p, q)$ processes based on a smooth version of the U_p - process, which uses an increasing number of standardized residual autocorrelations. Such a number plays the role of the amount of smoothness needed when estimating a nonparametric curve. Assuming a suitable rate of convergence on this smoothing number, the resulting U_p - process converges to the standard Brownian Bridge. Prewitt (1998) and Paparotidis (2000) proposed tests based on smooth estimates of the spectral density. These tests have the disadvantage of depending on the choice of a smoothing number, like Pormanteau tests, and their application to general specifications, including those with long memory, is far from being straightforward. One difficulty comes from the fact that the spectral density estimator is not consistent for frequencies near possible singularities. Another approach, feasible in a wide range of circumstances, consists of approximating the case dependent limiting distribution of test statistics by bootstrap methods in the lines suggested by Chen and Romano (1998), Delgado and Hidalgo (1999) or Hainz and Dahlhaus (2000).

In this paper, we follow a different approach, with the main objective of obtaining tests easy to implement in practice. Our test procedure is based on the T_p - process with estimated parameters,

in order to allow for fairly general specifications, such as long-memory. In order to obtain asymptotically distribution-free tests, we follow ideas put forward by Khmaladze (1981, 1988), who employed a martingale decomposition approach for Gaussian processes for constructing a (linear) transformation with a standard Brownian motion weak limit. The resulting test statistics, continuous functionals of these transformations, are asymptotically pivotal and able to detect contiguous alternatives converging to the null at the parametric rate $n^{-1/2}$. The testing methodology presented in this paper can be straightforwardly applied to other contexts, as testing the specification of non-stationary models and lack of autocorrelation testing of innovations in general dynamic models.

The remainder of the paper has been organized as follows. Section 2 introduces some notation and states the testing problem, whereas Section 3 presents the transformation producing asymptotically pivotal tests. Section 4 discusses the power of the resulting tests, showing that they are able to detect contiguous alternatives converging to the null at the rate $n^{-1/2}$. Section 5 studies the performance of the testing methodology by means of a small Monte Carlo experiment. Section 7 suggests some relevant extensions. Proofs have been placed in Section 6.

2. STATEMENT OF THE PROBLEM

Let $X(t)$ be a covariance stationary time series process observed at times $t \in \mathbb{Z}$, with spectral density function f_x , defined from the relation,

$$\gamma_x(j) := \text{Cov}(X(j), X(0)) = \int_{-\pi}^{\pi} f_x(\lambda) e^{ij\lambda} d\lambda \quad j = 0, 1, \dots$$

Assuming that,

$$\int_{-\pi}^{\pi} \log(f_x(\lambda)) d\lambda > -\infty,$$

f_x admits the factorization

$$f_x(\lambda) = \frac{\sigma^2}{2\pi} |\Delta(e^{i\lambda})|^2, \tag{1}$$

where,

$$\Delta(z) = \sum_{j=0}^{\infty} \delta(j) z^j,$$

for some function $\delta : \mathbb{Z} \rightarrow \mathbb{R}$ such that $\delta(0) = 1$ and $\sum_{j=0}^{\infty} \delta(j)^2 < \infty$. The function $\Delta(e^{i\cdot})$ is known as canonical factorization of f_x . The relation (1) corresponds to the frequency domain version of the Wold's decomposition,

$$X(t) = \mu + \Delta(L) \varepsilon(t), \tag{2}$$

where μ is the mean of the process, L is the lag operator, i.e. $LX(t) = X(t-1)$ and $\varepsilon(t)$ is a zero mean white noise process with variance σ^2 .

Given data, consisting of a record $X(1), \dots, X(n)$, we are interested in testing a composite null hypothesis that completely specifies the pattern of linear serial dependence of $X(t)$ up to a set of unknown parameters. Thus, the specification on the null hypothesis can be expressed in terms of the autocorrelation function or, alternatively, in terms of its Fourier transform, i.e. the spectral density function. Also, it could be expressed in terms of the coefficients δ in the Wold's decomposition.

Formally, given a particular function $h : [0, \pi] \times \Theta \rightarrow \mathbb{R}$, where Θ is a subset of \mathbb{R}^p , we are interesting in testing that,

$$H_0 : |\Delta(e^{i\lambda})|^2 = h(\lambda, \theta_0) \text{ for all } \lambda \in [0, \pi] \text{ and some } \theta_0 \in \Theta,$$

where $h(\lambda, \theta) := |D(e^{i\lambda}, \theta)|^2$ with $D(z, \theta) := \sum_{j=0}^{\infty} d(j, \theta) z^j$ and $d : \mathbb{N} \times \Theta \rightarrow \mathbb{R}$ is the parametric specification of δ , which shares its properties. In particular, we consider specifications such that.

Assumption A0 For each $\theta \in \Theta$, $\sum_{j=0}^{\infty} d(j, \theta)^2 < \infty$, $d(0, \theta) = 1$ and

$$\int_{-\pi}^{\pi} \log h(\lambda, \theta) d\lambda = 0. \quad (3)$$

Condition (3) implies that σ^2 is the variance of the one-step-ahead best linear predictor (see Hannan 1970, pp 157-163), and is functionally independent of θ .

The tests considered in this paper are consistent in the direction of general alternatives, of a nonparametric nature. Therefore, the alternative hypothesis, H_1 say, is the negation of the null. The null hypothesis can be equivalently expressed as

$$H_0 : \frac{F(\lambda, \theta_0)}{F(\pi, \theta_0)} = \frac{\lambda}{\pi} \text{ for all } \lambda \in [0, \pi] \text{ and some } \theta_0 \in \Theta,$$

where,

$$F(\lambda, \theta) := 2 \int_0^{\lambda} \frac{f_x(u)}{h(u, \theta_0)} du$$

is the spectral distribution of the innovations $\{e_{\theta}(t)\}$, which result from applying the filter $D^{-1}(L, \theta)$ to $\{X(t) - \mu\}$, i.e.

$$X(t) = \mu + D(L, \theta) e_{\theta}(t).$$

Therefore, the null hypothesis states that there is a parameter value θ_0 such that the innovations $\{e_{\theta_0}(t)\}$ are not autocorrelated, i.e. they have a constant spectral density.

A natural estimator of $F(\lambda, \theta)$ is,

$$F_n(\lambda, \theta) := \frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\tilde{n}\lambda/\pi]} \frac{I_{nx}(\lambda_j)}{h(\lambda_j, \theta)},$$

where $\lambda_j = (2\pi j)/n$, for integer $j = 1, \dots, \tilde{n}$, with $\tilde{n} = [n/2]$, $[\cdot]$ denotes nearest integer, and

$$I_{nx}(\lambda) := \frac{1}{2\pi n} \left| \sum_{t=1}^n X(t) e^{it\lambda} \right|^2$$

is the periodogram of $X(t)$. The periodogram I_{nx} is location invariant, because of the omission of the zero frequency in the sum. Thus, there is not need to center the data, and we can take $\mu = 0$ without loss of generality. Barlett (1954) proposed a standardized version of $F_n(\cdot, \theta_0)$, the T_p - process, for goodness-of-fit testing of the simple hypothesis (θ_0 given). The T_p - process is defined as,

$$T_n(\omega, \theta) := \sqrt{\tilde{n}} \left[\frac{F_n(\pi\omega, \theta)}{F_n(\pi, \theta)} - \omega \right].$$

Under H_0 , $T_n(\cdot, \theta_0) = T_n^0(\cdot)$, where,

$$T_n^0(\omega) := \sqrt{\tilde{n}} \left[\frac{F_n^0(\pi\omega)}{F_n^0(\pi)} - \omega \right],$$

and $F_n^0(\lambda) = F_n(\lambda, \theta_0)$.

Barlett (1954) also proposed an alternative to the T_p - process T_n^0 , the U_p - process,

$$U_n^0(\omega) = \frac{1}{\sqrt{\tilde{n}}} \frac{1}{F_n^0(\pi)} \sum_{j=1}^{[\tilde{n}\omega]} \left[I_{nx}(\lambda_j) - F_n^0(\pi) |\Delta(e^{i\lambda_j})|^2 \right].$$

The empirical process $\sqrt{\tilde{n}}F_n^0(\lambda)$, and hence $T_n^0(\omega)$ and $U_n^0(\omega)$, are random elements of the cadlag functional space $\mathbb{D}[0, 1]$, which is endowed with the Skorohod's metric (see Billingsley, 1968).

Under suitable regularity conditions (e.g. Grenander and Rosenblatt 1957, Ibragimov 1963 and, more recently, Dahlhaus 1985 and Anderson 1993), U_n^0 converges weakly to a Gaussian process with covariance function depending on the fourth cumulant, but results in Hannan (1973) suggest that one necessary assumption for the weak convergence is the squared integrability of f_x , which avoids some interesting specifications that we are taking into consideration in this paper, such as process where f_x has singularities, as it is the case in long-memory specifications. However, T_n^0 converges weakly to the standard Brownian bridge on $[0, 1]$ under fairly weak regularity conditions (A1 and A2 below), see Proposition 1 below.

We allow for a much wider class of linear processes described in the following assumption.

Assumption A1 The function $|\Delta(e^{i\cdot})|^2$ is continuously differentiable in any open set outside a neighbourhood of the frequency $v \in [0, \pi]$, such that

$$\left| \frac{\partial}{\partial \lambda} \ln |\Delta(e^{i\lambda})|^2 \right| = O\left(\frac{1}{|\lambda - v|}\right) \text{ as } \lambda \rightarrow v.$$

Condition A1 allows $h(\cdot, \theta_0)$ to possess a singularity at the frequency v , that is, we allow for processes $X(t)$ exhibiting long memory; e.g. when $v = 0$, are the *FARIMA* model introduced by Adenstedt (1974) and examined in Granger and Joyeux (1980) and Hosking (1981) and when $|v| > 0$ are the *GARMA* models (see e.g. Ansel, 1980; Gray et al., 1989). This condition is equivalent to Assumption A6 of Giraitis et al. (2001) and extends Assumption A2' of Robinson (1995b) and similar conditions in Giraitis and Surgailis (1990) to allow for $v \neq 0$. Although the results below allow for spectral density functions with more than one singularity, we do not pursue this for the sake of expositional simplicity.

Assumption A2 The white noise zero mean innovation process $\varepsilon(t)$ in (2) satisfies that:

$$E\left[\varepsilon(t)^k \middle| \mathcal{F}_{t-1}\right] = \mu_k, \quad k = 1, \dots, 4, \quad \text{and} \quad E|\varepsilon(t)|^k = \mu_k, \quad k = 1, \dots, 8 \text{ with } \mu_8 < \infty,$$

where μ_k , $k = 1, \dots, 8$ are constants and \mathcal{F}_{t-1} is the sigma algebra generated by the events $\{\varepsilon(s), s < t\}$.

To obtain the asymptotic distribution of parameter estimates, Robinson (1995b) uses only four constant conditional moments in his Assumption A3', as Velasco and Robinson (2000), while Giraitis et al. (2001) further require $\sup_{t \geq 1} E|\varepsilon(t)|^{4+\rho} < \infty$ for some $\rho > 0$. This will suffice also for the convergence of the finite dimensional distributions (fidis) of T_n^0 , as stated in next Proposition. The extra moment assumptions are needed for tightness. This could be relaxed at the cost of further conditions on $\varepsilon(t)$ (as *idd*, see Kokoszka and Mikosch, 2000), but is not much stronger than the assumptions of Dahlhaus (1985), which assumes only constant conditional moments of order three, allowing for more general specifications in our Assumption A1.

Proposition 1 *If A1 and A2 are satisfied, then,*

$$T_n^0 \text{ converges in distribution to } B \text{ in the Skorohod's metric space } \mathbb{D}[0, 1],$$

where B is distributed as the standard Brownian Bridge; i.e. $B(\omega) \stackrel{d}{=} \bar{B}(\omega) - \omega \bar{B}(1)$, with \bar{B} distributed as a standard Brownian Motion on $[0, 1]$, a Gaussian process centered at zero with covariance function $E[\bar{B}(\omega_1) \bar{B}(\omega_2)] = \min(\omega_1, \omega_2)$.

This proposition and the rest of results are proved in the last section. The convergence of the finite dimensional distributions follows straightforwardly from results in Robinson (1995a) and Giraitis et al. (2001), and tightness is proved using standard arguments.

Since the null hypothesis is composite, and θ_0 is unknown, T_n^0 is unsuitable in practice for testing H_0 . A natural estimator of θ_0 is the Whittle's estimator,

$$\theta_n = \arg \min_{\theta \in \Theta} F_n(\pi, \theta),$$

which admits a first order asymptotic linearization under certain assumptions on $h(\cdot; \cdot)$. In particular, we need the following assumption.

Assumption A3 For a generic finite constant K ,

A3.1

$$\sup_{\{\lambda \in [0, \pi] \setminus \{v\}, \theta: \|\theta - \theta_0\| \leq Kn^{-1/2}\}} \left| \frac{1}{\log |\lambda - v| \cdot \|\theta - \theta_0\|^2} \left[\frac{h(\lambda, \theta)}{h(\lambda, \theta_0)} - 1 + \phi(\lambda, \theta_0)'(\theta - \theta_0) \right] \right| \leq K,$$

where $\phi : [0, \pi] \times \Theta \rightarrow \mathbb{R}$ is such that,

$$\sup_{\{\lambda \in [0, \pi] \setminus \{v\}, \theta: \|\theta - \theta_0\| \leq Kn^{-1/2}\}} \frac{1}{|\log |\lambda - v||} \|\phi(\lambda, \theta)\| \leq K,$$

$$\sup_{\{\lambda \in [0, \pi] \setminus \{v\}; |\mu| \leq |\lambda - v|/2; \theta: \|\theta - \theta_0\| \leq Kn^{-1/2}\}} \frac{|\lambda - v|}{|\mu|} \|\phi(\lambda, \theta) - \phi(\lambda + \mu, \theta)\| \leq K.$$

A3.2 $\phi(\cdot, \theta_0)$ is squared integrable, with

$$A_0 := \int_0^1 \phi(\pi u, \theta_0) \phi(\pi u, \theta_0)' du$$

nonsingular.

A3.3 Also $R(1) = 0$, $\sup_{\omega \in [0, 1]} |R(\omega)| < \infty$, where,

$$R(\omega) := \int_0^\omega \phi(\pi u, \theta_0) du.$$

Part A3.2 and A3.3 are standard in Whittle estimation. Part A3.2 is an identification condition, whereas part A3.3 matches (3) in A0. Part A3.1 needs more explanation; however, it is satisfied for most models used in practice, such as ARMA, Bloomfield (1973), and ARFIMA. A sufficient condition is that the spectral density function, i.e. $h(\cdot, \theta)$, is twice continuously differentiable in θ for each $\lambda \in [0, \pi] \setminus \{v\}$ and for $\lambda \rightarrow v$, the derivative of the logarithm of the spectral density is $O(\log |\lambda - v|)$. Similar conditions in a different format were given in Giraitis et al. (2001) Assumption A1.

Assumption A4 Under H_0 ,

$$\theta_n = \theta_0 - A_0^{-1} \frac{2\pi}{\sigma^2 \tilde{n}} \sum_{j=1}^{\tilde{n}} \phi(\lambda_j, \theta_0) \frac{I_{nx}(\lambda_j)}{h(\lambda_j, \theta_0)} + o_p(n^{-1/2}).$$

Though θ_n is most widely employed, there are other estimators based on robust proposals (e.g. Taniguchi, 1981), which may yield different asymptotic linearizations. Sufficient conditions for A4, assuming also that A1 to A3 hold, can be found in Fox and Taqqu (1986), Hosoya (1997) and Giraitis et al. (2001).

Thus, a feasible version of $T_n^0(\omega)$ under H_0 is,

$$\hat{T}_n(\omega) := \sqrt{\tilde{n}} \left[\frac{\hat{F}_n(\omega\pi)}{\hat{F}_n(\pi)} - \omega \right],$$

with $\hat{F}_n(\lambda) = F_n(\lambda, \theta_n)$, and an asymptotic expansion of T_n^0 is an immediate consequence of Proposition 1 and Assumption A3, as stated in the following proposition.

Proposition 2 *Assuming A0-A3, under H_0 ,*

$$\sup_{\omega \in [0,1]} \left| \hat{T}_n(\omega) - \mathcal{L}T_n^0(\omega) \right| = o_p(1),$$

where, for any generic function $g : [0, 1] \rightarrow \mathbb{R}^+$,

$$\mathcal{L}g(\omega) = g(\omega) - R(\omega)' A_0^{-1} \int_0^1 \phi(\pi v, \theta_0) g(dv),$$

with $R(\cdot) := \int_0^\cdot \phi(\pi v, \theta_0) dv$.

Therefore, \hat{T}_n and $\mathcal{L}T_n^0$ share the same limiting distribution. Next Corollary is a straightforward consequence of Propositions 1 and 2.

Corollary 1 *Assuming A0-A3, under H_0 ,*

$$\hat{T}_n \text{ converges in distribution to } \mathcal{L}B \text{ in the space } \mathbb{D}[0, 1].$$

Tests based on continuous functionals of \hat{T}_n are not asymptotically pivotal, since the transformation \mathcal{L} depends on the unknown parameter θ_0 . Next Section introduces a linear transformation of \hat{T}_n which converges weakly, under H_0 , to the standard Brownian Motion and, hence, the corresponding critical values can be easily tabulated.

3. DISTRIBUTION FREE TESTS

We are interested in a linear transformation, $\bar{\mathcal{L}}$ say, such that $\bar{\mathcal{L}}\hat{T}_n$ converges weakly to \bar{B} . Fidis convergence of $\bar{\mathcal{L}}\hat{T}_n$ to those of $\bar{\mathcal{L}} \circ \mathcal{L}B$ is a consequence of Corollary 1. Notice that,

$$\mathcal{L}B(\omega) \stackrel{d}{=} \bar{B}(\omega) - G(\omega)' \begin{bmatrix} 1 & 0 \\ 0 & A_0^{-1} \end{bmatrix} \int_0^1 \begin{pmatrix} 1 \\ \phi(\pi v, \theta_0) \end{pmatrix} \bar{B}(dv),$$

with $G(\cdot)' := (\cdot, R(\cdot))$. Therefore, the linear transformation $\bar{\mathcal{L}}$ satisfies

$$\bar{\mathcal{L}} \circ \mathcal{L}B(\omega) \stackrel{d}{=} \bar{B}(\omega),$$

if

$$\sup_{\omega \in [0,1]} \|\bar{\mathcal{L}}G(\omega)\| = 0, \quad (4)$$

and

$$\bar{\mathcal{L}}\bar{B}(\omega) \stackrel{d}{=} \bar{B}(\omega). \quad (5)$$

Reasoning as in Khmaladze (1981), we suggest the transformation,

$$\begin{aligned} \bar{\mathcal{L}}g(\omega) &:= g(\omega) - \int_0^\omega \frac{\partial G'(u)}{\partial u} \left[\int_u^1 \frac{\partial G(v)}{\partial v} \frac{\partial G'(v)}{\partial v} dv \right]^{-1} \int_u^1 \frac{\partial G(v)}{\partial v} g(dv) du \\ &= g(\omega) - \int_0^\omega \begin{pmatrix} 1 \\ \phi(\pi u, \theta_0) \end{pmatrix}' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1} \int_u^1 \begin{pmatrix} 1 \\ \phi(\pi v, \theta_0) \end{pmatrix} g(dv) du, \end{aligned}$$

where $A_0(\omega) := \int_\omega^1 \phi(\pi u, \theta_0) \phi(\pi u, \theta_0)' du$. The transformation is applied, on one hand, to functions of bounded variation and, on the other hand, to the Brownian motion \bar{B} .

Next assumption is needed for implementing the transformation.

Assumption A4. There exists a $\omega_0 \leq 1$ such that $A_0(\omega_0)$ is nonsingular.

For many specifications, e.g. ARFIMA models, we can take $\omega_0 = 1$.

Next theorem states that $\bar{\mathcal{L}}\hat{T}_n$ converges weakly to \bar{B} , and will be proved in Section 6 using the following arguments. Notice that (4) is trivially satisfied, and a necessary and sufficient condition for (5) is that $E[(\bar{\mathcal{L}}\bar{B}(\omega_1))(\bar{\mathcal{L}}\bar{B}(\omega_2))] = \min(\omega_1, \omega_2)$, which can be proved reasoning as in Stute et al. (1998) Lemma 3.1. Also, it can be shown that, if A1-A4 hold, $\sup_\omega \left| \bar{\mathcal{L}}(\hat{T}_n(\omega) - T_n^0(\omega)) \right| = o_p(1)$. The convergence of the finite dimensional distributions of $\bar{\mathcal{L}}T_n^0$ to those of \bar{B} is an immediate consequence of Proposition 1, (4) and (5); tightness of $\bar{\mathcal{L}}T_n^0$ is established applying standard arguments.

Theorem 1 Assume A0-A4. Then, under H_0 ,

$$\overline{\mathcal{L}}\hat{T}_n \text{ converges in distribution to } \bar{B} \text{ on the space } \mathbb{D}[0, \omega_0].$$

The transformation $\overline{\mathcal{L}}$ is infeasible, as it depends on the unknown parameter θ_0 , but in practice we can take the sample counterpart,

$$\overline{\mathcal{L}}_n g(\omega) := g(\omega) - \int_0^\omega \begin{pmatrix} 1 \\ \hat{\phi}_n(\pi u) \end{pmatrix}' \begin{bmatrix} 1-u & -\hat{R}_n(u) \\ -\hat{R}_n(u) & \hat{A}_n(u) \end{bmatrix}^{-1} \int_u^1 \begin{pmatrix} 1 \\ \hat{\phi}_n(\pi v) \end{pmatrix} g(dv) du,$$

where $\hat{R}_n(\omega) := (1/\tilde{n}) \sum_{j=1}^{[\tilde{n}\omega]} \hat{\phi}_n(\lambda_j)$ and $\hat{A}_n(\omega) := (1/\tilde{n}) \sum_{j=[\tilde{n}\omega]}^{\tilde{n}} \hat{\phi}_n(\lambda_j) \hat{\phi}_n(\lambda_j)'$ with $\hat{\phi}_n(\cdot) := \phi(\cdot; \hat{\theta}_n)$.

Theorem 2 Assuming A.0 – A.4, under H_0 ,

$$\sup_{\omega \in [0, \omega_0]} |(\overline{\mathcal{L}}_n - \overline{\mathcal{L}}) \hat{T}_n(\omega)| = o_p(1),$$

and consequently,

$$\overline{\mathcal{L}}\hat{T}_n \text{ converges in distribution to } \bar{B} \text{ in the Skorohod's metric space } \mathbb{D}[0, \omega_0].$$

This theorem justifies asymptotic consistent tests based on continuous functionals of $\overline{\mathcal{L}}_n \hat{T}_n$, as stated in the following Corollary.

Corollary 2 For any continuous functional $\varphi : \mathbb{D}[0, \omega_0] \mapsto \mathbb{R}^+$, under H_0 and assuming A0-A4,

$$\varphi(\overline{\mathcal{L}}\hat{T}_n) \text{ converges in distribution to } \varphi(\bar{B}).$$

The distribution of $\varphi(\bar{B})$ can be tabulated by Monte Carlo. For the main goodness-of-fit proposals, Kolmogorov-Smirnov and Cramér-von Misses, $\varphi(\bar{B})$ is already tabulated. Notice that, when $\omega_0 \in (0, 1]$, for any constant $c > 0$,

$$\begin{aligned} \Pr \left(\sup_{\omega \in [0, \omega_0]} |\bar{B}(\omega)| > c\sqrt{\omega_0} \right) &= \Pr \left(\sup_{\omega \in [0, 1]} |\bar{B}(\omega)| > c \right), \\ \Pr \left(\int_0^{\omega_0} \bar{B}(\omega)^2 d\omega > c\omega_0^2 \right) &= \Pr \left(\int_0^1 \bar{B}(\omega)^2 d\omega > c \right), \end{aligned}$$

and the values of the right hand side of the above expressions, which correspond to the critical values with $\omega_0 = 1$, can be found, for instance, in Shorack and Wellner (1980) pp. 34 and 748.

4. POWER CONSIDERATIONS: LOCAL ALTERNATIVES AND CONSISTENCY

Consider contiguous alternatives of the form,

$$H_{an} : |\Delta(e^{i\lambda})|^2 = h(\lambda, \theta_0) \left(1 + n^{-1/2}l(\lambda)\right) \text{ for some } \theta_0 \in \Theta \quad (6)$$

where $l : [0, \pi] \rightarrow (-n^{1/2}, \infty)$ for all $n \geq 1$. Under the set of assumptions in the last section, the proposed test does not have trivial power, as stated in the following theorem.

Theorem 3 *Assuming A0-A4, under H_{an} ,*

$$\overline{\mathcal{L}}\hat{T}_n \text{ converges in distribution to } \bar{B} + \overline{\mathcal{L}}S \text{ on the space } \mathbb{D}[0, \omega_0],$$

where,

$$S(\omega) := L(\omega) - \omega L(1) - R(\omega) A_0^{-1} \int_0^\pi \phi(\lambda, \theta) l(\lambda) d\lambda,$$

$L(\omega) := \int_0^\omega l(u\pi) du$, and S is different than zero uniformly in ω in a set with Lebesgue's measure greater than zero.

Therefore, for a suitable functional $\varphi : \mathbb{D}[0, \omega_0] \mapsto \mathbb{R}^+$, e.g. the Crámer-von Misses or the Kolmogorov-Smirnov criterion, $\Pr[\varphi(\bar{B} + S) > \varphi(\bar{B})] = 1$, and the test will detect contiguous departures from the null of the form H_{an} .

The test is consistent in the direction of general fixed alternatives. A formal justification of this statement under precise regularity conditions is beyond the scope of this paper, we just sketch the argument. Assuming certain regularity conditions, A3 and A4 are satisfied under the alternative, H_1 , where now, $\theta_0 := \arg \min_{\theta \in \Theta} F(\pi, \theta)$ is a pseudo-parameter that does not satisfy the null hypothesis, see Dahlhaus and Wefelmeyer (1996) for an study of the Whittle's estimator under misspecification. Reasoning as in Dahlhaus and Wefelmeyer (1996), we shall have that, for each $\omega \in [0, 1]$,

$$\hat{T}_n(\omega) = \mathcal{L}T_n(\omega, \theta_0) + o_p(1),$$

The consistency of the test is justified taking into account that,

$$T_n(\omega, \theta_0) = \frac{1}{\sqrt{\tilde{n}}} \sum_{j=1}^{[\tilde{n}\omega]} \left[\frac{2\pi I_{nx}(\lambda_j)}{F_n(\pi, \theta_0) |\Delta(e^{i\lambda_j})|^2} - 1 \right] \frac{|\Delta(e^{i\lambda_j})|^2}{h(\lambda_j, \theta_0)} + \sqrt{\tilde{n}} \left[\frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\tilde{n}\omega]} \frac{|\Delta(e^{i\lambda_j})|^2}{F_n(\pi, \theta_0) h(\lambda_j, \theta_0)} - \omega \right],$$

where the first term is an $O_p(1)$ under suitable regularity conditions, while the second term is an $O_p(n^{1/2})$.

5. MONTE-CARLO EXPERIMENT

A small Monte-Carlo study was carried out to investigate the finite sample performance of the test, as well as its relative performance compared to Box-Pierce tests. To that end, we have considered two models. Specifically, an AR(1) with a sole parameter $\theta_0 = \rho_0$ and a FARIMA(0, d_0 ,0) with parameter $\theta_0 = d_0$, given by

$$(1 - \rho_0 L) X(t) = \varepsilon(t) \tag{7}$$

and

$$(1 - L)^{d_0} X(t) = \varepsilon(t) \tag{8}$$

respectively, where the innovations $\varepsilon(t)$ are *iid* $N(0, 1)$, for sample sizes $n = 200$ and 500 and different values of the parameter. For model (7), we have considered $\rho_0 = -.8, -.5, 0.0, .5$ and $.8$, whereas for model (8), $d_0 = 0.1, 0.2, 0.3, 0.4$. The FARIMA model is simulated exactly by the algorithm of Hosking (1984).

For both models (7) and (8), and the different values of θ_0 , we have computed the proportion of rejections in 50,000 generated samples for both samples sizes. We only report the results for 5% tests using the Crámer-von Misses criterion, i.e. the test statistic C_n ,

$$C_n := \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \left(\mathcal{L}_n \hat{T}_n \left(\frac{j}{\tilde{n}} \right) \right)^2,$$

(the results with the Kolmogorov-Smirnov criterion were very similar). We also report, as a benchmark, the proportion of rejections using the infeasible test statistic

$$C_n^0 := \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} T_n^0 \left(\frac{j}{\tilde{n}} \right)^2,$$

suitable for testing simple hypotheses, which converges in distribution to $\int_0^1 B(\omega)^2 d\omega$ under H_0 (for a tabulation see e.g. Shorack and Wellner, 1986, pp. 147). Also, for the sake of comparison, we provide results for the popular Pormanteau test statistic proposed by Ljung and Box (1978)

$$Q_m = n(n+2) \sum_{j=1}^m \frac{\hat{\rho}_{\hat{\varepsilon}}(j)^2}{n-j},$$

for several values of m , where

$$\hat{\rho}_{\hat{\varepsilon}}(j) := \left(\sum_{t=1}^n \hat{\varepsilon}(t)^2 \right)^{-1} \sum_{t=j+1}^n \hat{\varepsilon}(t) \hat{\varepsilon}(t-j), j \geq 1,$$

are the sample autocorrelations of residuals $\varepsilon(t)$. In the AR(1) case,

$$\hat{\varepsilon}(t) = (1 - \theta_n L) X(t),$$

with $X(t) = 0$ for $t \leq 0$, and in the FARIMA(0, θ_0 , 0) case

$$\hat{\varepsilon}(t) = \sum_{j=0}^t \psi(j, \theta_n) X_{t-j},$$

where $\psi(j, \theta_n)$ are the coefficients in the formal expansion $(1 - L)^{\theta_n} = \sum_{j=0}^{\infty} \psi(j, \theta_n) L^j$, and equal to

$$\psi(j, \theta_n) = \frac{\Gamma(j - \theta_n)}{\Gamma(-\theta_n) \Gamma(\theta_n + 1)}; \quad \Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx.$$

The values of Q_m are compared against the critical value of the χ_{m-1}^2 distribution.

In our experiment, both tests showed sizes close to nominal, in many cases improving with respect to those based on the infeasible T_p -process T_n^0 . The asymptotic approximation improves noticeably when the sample size increases from $n = 200$ to $n = 500$, being this improvement uniform for all the models. For the choices of m implemented, Q_m over-rejected H_0 . We have also tried non-Gaussian innovations. In this case, although there is some loss of accuracy, as expected, this is not very noticeable (negligible). These results are not reported here.

Next, we study the empirical power of the tests. For the previous set up, we analyse the power of the 5% tests for the null of AR(1) against the alternative of fractional integration and viceversa. These two models describe parsimoniously low frequency behaviour and are often competitive in applied modeling of time series. The empirical power is reported in Tables 2 and 3.

Tests based on $\bar{\mathcal{L}}_n \hat{T}_n$ showed reasonable power for series on length 200 for most values of the parameters, while for $n = 500$ they distinguish without trouble both classes of linear models. Here, we found that the Crámer von Mises test was uniformly more powerful than the Kolmogorov-Smirnov, surely because the low order parameterizations under the null causes a bad fit in a wider interval around zero frequency. Here Q_m is often superior to distinguish autorregressive data with large θ_0 from fractional models, though both type of tests have problems when moving closer to a unit root ($\theta_0 = 0.8$). However for $n = 500$ C_n test has more power than Q_m when $\theta_0 = 0.2$ for all m , and for larger θ_0 , C_n has power comparable to Q_m for large choices of m . For fractional data, CvM tests are more powerful than Q_m for all d and most choices of m .

6. SOME EXTENSIONS

In this section we discuss two extensions of the proposed methodology. The first deals with the specification of possibly nonstationary processes when the analyst has no prior knowledge on the required number of differences to achieve stationarity, as it is the case when one applies the Box and Jenkins (1976) methodology. The second application investigates the consistent specification of dynamic regression models.

Nonstationary Processes

The parameterization of $h(\lambda; \theta)$ can also describe models for nonstationary or integrated models without a priori assumptions about the possible stationarity of the time series. We focus on difference stationary series by means of a generalized spectral density. Thus, if the increments $Y(t) = (1 - L)X(t)$, $t \in \mathbb{Z}$, are covariance stationary with zero mean and spectral density f_y such that

$$\lim_{\lambda \rightarrow 0} |\lambda|^{2(d-1)} f_y(\lambda) = G > 0,$$

for some $d \in [0.5, 1.5)$, we can define the pseudo spectrum of the observed $X(t)$ as

$$f_x(\lambda) := |1 - e^{i\lambda}|^{-2} f_y(\lambda). \tag{9}$$

If $X(t)$ is stationary, then f_x is simply the usual spectral density and when

$$\lim_{\lambda \rightarrow 0} |\lambda|^{2d} f_x(\lambda) = G > 0,$$

for some $d \in (-0.5, 0.5)$, as happens for many long memory time series, the parameter d describes the persistence properties of the original series in a continuous way. This fact allows consistent estimation of an appropriate parameterization of f_x without previous specification of whether $d \geq 0.5$ or $d < 0.5$ (see Velasco, 1999a, b and Velasco and Robinson, 2000).

The pseudo spectral density f_x can be interpreted as the limit of the expected sample periodogram as suggested by Solo (1992) and Hurvich and Ray (1995). However, for highly nonstationary series, $d \geq 1$, tapering is an essential tool to improve the properties of the periodogram I_{nx} and to construct parameter estimates with uniform properties in $d \in (-0.5, 1.5)$. The tapered discrete Fourier transform (DFT) of $X(t)$ for a taper sequence $p(t)$ is defined as

$$w_x^p(\lambda) := \frac{\sum_{t=1}^n p(t) X(t) e^{i\lambda t}}{\left(2\pi \sum_{t=1}^n p(t)^2\right)^{1/2}}$$

and the tapered periodogram as $I_x^p(\cdot) := |w_x^p(\cdot)|^2$. The usual DFT has $p(t) \equiv 1$ all t . Typically p downweights the observations at both extremes of the sequence, leaving almost unchanged the central part of the data. We use throughout the full cosine bell taper,

$$p(t) = \frac{1}{2} \left(1 - \cos \frac{2\pi t}{n} \right),$$

because of its desirable properties and because in this case the asymptotic analysis is only slightly complicated. The basic effects of tapering are a reduction of periodogram bias, even for nonstationary data, and an increase of variance and correlation among periodogram ordinates, which will lead to a different standardization of $\hat{T}_n(\omega)$ and related statistics. There are other alternatives, but similar results may be expected for many data tapers (see Velasco, 1999a).

With the cosine tapering, assuming that the singularity is at frequency $\nu = 0$ in assumption A1, we exclude the first Fourier frequency, λ_1 , in all statistics because of its different asymptotic properties for nonstationary data, so in this case,

$$F_n(\lambda, \theta) = \frac{2\pi}{\tilde{n}} \sum_{j=2}^{[\tilde{n}\lambda/\pi]} \frac{I_{nx}^p(\lambda_j)}{h(\lambda_j; \theta)}.$$

The Whittle's estimate, θ_n , minimize this $F_n(\pi, \theta)$ with the tapered periodogram, and a similar linearization than in A3, substituting I_{nx} by I_{nx}^p will also hold in this case, which with A1 and A2 are sufficient conditions for Proposition 2 and 3, as well as, Corollary 1 and 2, where the T_p -process $T_n(\cdot, \cdot)$ should be standardized by $\sqrt{\tilde{n}/P_4^2}$, rather than by $\sqrt{\tilde{n}}$, where,

$$P_4^2 = \lim_{n \rightarrow \infty} \frac{n \sum_{t=1}^n p(t)^4}{\left(\sum_{t=1}^n p(t)^2 \right)^2} = \frac{35}{18}.$$

Dynamic Models

Other direct extension of the linear model considered in Section 2 is the *ARX* parameterization when the vector of explanatory variables $Z(t) = (Z_1(t), \dots, Z_p(t))'$ are linear in an autoregressive model,

$$X(t) = \mu_0 + \theta_{01}X(t-1) + \dots + \theta_{0p}X(t-p) + \beta_{01}Z_1(t) + \dots + \beta_{0q}Z_q(t) + \varepsilon(t), \quad (10)$$

where the $p+q+1$ vector of parameters $\gamma'_0 = (\mu_0, \theta_0, \beta_0)$ is identified as the solution of the $p+q+1$ moment conditions,

$$E[W(t)(X(t) - \gamma'W(t))] = 0, \quad (11)$$

where $W(t)' = (1, X(t-1), \dots, X(t-p), Z(t))$, assuming that $E[W(t)W(t)']$ is non singular. Moving average terms could be considered, but the previous form is interesting enough, since it permits to illustrate an alternative construction of our distribution-free tests when residuals can be easily calculated for a given estimate γ_n , typically obtained by ordinary least squares.

The null hypothesis is again that the model is correctly specified, in the sense that innovations are not autocorrelated. That is,

$$H_0 : \gamma_\varepsilon(j; \gamma_0) = 0, \text{ for all } j \geq 1 \text{ and for some } \gamma_0 \in \Gamma \subset \mathbb{R}^{p+q} \text{ satisfying (11),}$$

with

$$\gamma_\varepsilon(j; \theta) = E[(X(t) - \theta'W(t))(X(t+|j|) - \theta'W(t+|j|))].$$

The alternative is the negation of the null. The set $\Gamma = \Theta \times \mathbb{R}^{q+1}$ is chosen such that for $\theta \in \Theta$, $\Phi(z, \theta) = 1 - \theta_1 z - \dots - \theta_p z^p$ has no roots inside or on the unit circle. Note that least squares estimation is inconsistent if H_0 fails.

To obtain feasible tests of H_0 we assume that estimates $\gamma'_n = (\mu_n, \theta_n, \beta_n)$ are available such that the following condition, analog to A3, holds,

Assumption A3' :

$$\theta_n = \theta_0 - A_0^{-1} \frac{2\pi}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \phi(\lambda_j, \theta_0) I_{n\varepsilon}(\lambda_j) + o_p(n^{-1/2}),$$

where $\phi(\lambda, \theta) = (\partial/\partial\theta) \log |\Phi(e^{i\lambda}, \theta)|^{-2}$, with

$$A_0 = \int_0^1 \phi(\pi u, \theta_0) \phi(\pi u, \theta_0)' du > 0,$$

$$\beta_n = \beta_0 + O_p(n^{-1/2}) \text{ and } \mu_n = \mu_0 + O_p(n^{-1/2}).$$

Given the particular form of the model (10) we can calculate residuals

$$\hat{\varepsilon}(t) := X(t) - \gamma'_n W(t),$$

and residual periodograms $I_{n\hat{\varepsilon}}(\lambda_j)$, and set directly

$$F_n(\lambda, \gamma_n) = \frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\tilde{n}\lambda/\pi]} I_{n\hat{\varepsilon}}(\lambda_j),$$

while \hat{T}_n is defined as before in terms of $F_n(\lambda, \gamma_n)$. In this set up, minimization of $F_n(\pi, \gamma)$ is alternative to least squares estimation if initial observations $X(1-p), \dots, X(-1)$ are available.

The cross spectral density of $(\varepsilon(t), z(t))$, $f_{\varepsilon z}$ is equal to zero on $\lambda \in [0, \pi]$, and under suitable regularity conditions,

$$\sup_{\omega \in [0, 1]} \left| \frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\omega \tilde{n}]} I_{n, \varepsilon z}(\lambda_j) \right| = o_p(1), \quad (12)$$

where, for two processes $V_1(t)$ and $V_2(t)$,

$$I_{n, v_1 v_2}(\lambda) := \frac{1}{2\pi n} \left(\sum_{t=1}^n V_1(t) e^{it\lambda} \right) \left(\sum_{t=1}^n V_2(t) e^{-it\lambda} \right)$$

is the cross-periodogram of $V_1(t)$ and $V_2(t)$. This fact, together with the rate of convergence of $\beta_n - \beta_0$, is enough to show that the inclusion of exogenous regressors in the linear model does not affect the asymptotic distribution of \hat{T}_n . Define, $I_{n, \varepsilon x(1, p)}(\lambda) = (I_{n, \varepsilon x_{-1}}(\lambda), \dots, I_{n, \varepsilon x_{-p}}(\lambda))'$, where

$$I_{n, \varepsilon x_{-j}}(\lambda) := \frac{1}{2\pi n} \left(\sum_{t=1}^n \varepsilon(t) e^{it\lambda} \right) \left(\sum_{t=1}^n X(t-j) e^{-it\lambda} \right)$$

Assuming $A1$, $A2$ and $A3'$, under H_0 , uniformly in ω ,

$$\begin{aligned} \hat{T}_n(\omega) &= T_n(\omega; \theta_0) - 2 \frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\omega \tilde{n}]} I_{n, \varepsilon x(1, p)}(\lambda_j) (\theta_n - \theta_0) \\ &\quad - 2 \frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\omega \tilde{n}]} I_{n, \varepsilon z}(\lambda_j) (\beta_n - \beta_0) + o_p(n^{-1/2}) \\ &= T_n(\omega; \theta_0) - 2 \frac{2\pi}{\tilde{n}} \sum_{j=1}^{[\omega \tilde{n}]} I_{n, \varepsilon x(1, p)}(\lambda_j) (\theta_n - \theta_0) + o_p(n^{-1/2}), \end{aligned}$$

and reasoning as in Proposition 2,

$$\sup_{\omega} \left| \hat{T}_n(\omega) - \mathcal{L}T_n(\omega, \theta_0) \right| = o_p(1),$$

The transformation $\bar{\mathcal{L}}$ is identical to that suggested in the last section, and using similar arguments to those in the proof of Theorem 1, it is established that $\bar{\mathcal{L}}\hat{T}_n$ converges weakly to \bar{B} .

7. PROOFS

Henceforth, $\xi(u)' := (1, \phi(\pi u, \theta_0))'$, $u \in [0, 1]$.

Proposition 3 *Assume $A1$, $A2$ and $A3$ hold, then,*

$\left\{ \int_0^\omega \xi(u) T_n^0(du) : \omega \in [0, 1] \right\}$ converges in distribution to $\left\{ \int_0^\omega \xi(u) B(du) : \omega \in [0, 1] \right\}$ in the space $\mathbb{D}_{p+1}[0, 1]$.

Proof: For simplicity we take $v = 0$ in A1 and A3. Reasoning as Robinson (1995b, proof of Theorem 2), for each $\omega \in [0, 1]$,

$$\int_0^\omega \xi(u) [T_n^0(du) - \bar{T}_n^0(du)] = o_p(1), \quad (13)$$

where,

$$\bar{T}_n^0(\omega) := \sqrt{\tilde{n}} \left[\frac{\bar{F}_n^0(\omega\pi)}{\bar{F}_n^0(\pi)} - \omega \right] \text{ with } \bar{F}_n^0(\cdot) := \frac{2\pi}{\tilde{n}} \sum_{i=1}^{[\tilde{n}\cdot/\pi]} I_{n\varepsilon}(\lambda_j),$$

and $I_{n\varepsilon}$ is the periodogram of $\varepsilon(t)$. Also, it follows, from Robinson (1995b) the weak convergence of $\{\int_0^\omega \xi(u) T_n^0(du) : \omega \in J\}$ for each finite set $J \in [0, 1]$. Therefore, fdis convergence of $\int_0^\omega \xi(u) T_n^0(du)$ is a fairly straightforward consequence of results in Robinson (1995b). In fact, A2 is sufficient for Corollary 3.7 in Dahlhaus (1985) to hold, that is,

$$\bar{T}_n^0 \text{ converges in distribution to } B \text{ in the space } \mathbb{D}[0, 1].$$

Then, it suffices to prove tightness. The process T_n^0 is a random element of $\mathbb{D}[0, 1]$, but the process $T'(\omega) = T_n^0(\omega) + \frac{\tilde{n}\omega - [\tilde{n}\omega]}{\sqrt{\tilde{n}}} \left[\frac{I_n(\lambda_{[\tilde{n}\omega]+1})}{\bar{F}_n^0(\pi)} - [\tilde{n}\omega] - 1 \right]$, constructed by interpolation is an element of the functional space of continuous functions in $[0, 1]$, $\mathbb{C}[0, 1]$. Therefore it suffices to check Chernoff's criterion (see Billingsley 1968, Theorem 12.3), that is, for $0 \leq \omega_1 < \omega_2 \leq 1$ and some $\delta > 1$,

$$E \left\| \sqrt{\tilde{n}} \int_{\omega_1}^{\omega_2} \xi(u) F_n^0(\pi du) \right\|^4 \leq K (\omega_2 - \omega_1)^\delta, \quad (14)$$

where K is a generic constant not depending on n or ω_2, ω_1 . Write,

$$\begin{aligned} \int_0^\omega \xi(u) F_n^0(\pi du) &= \int_0^\omega \xi(u) E[F_n^0(\pi du)] + \int_0^\omega \xi(u) (F_n^0(\pi du) - E[F_n^0(\pi du)]) \\ &= H_n(\omega) + W_n(\omega). \end{aligned}$$

Then, applying triangular inequality,

$$E \left\| \sqrt{\tilde{n}} \int_{\omega_1}^{\omega_2} \xi(u) F_n^0(\pi du) \right\|^4 \leq K \left[\|H_n(\omega_1) - H_n(\omega_2)\|^4 + E \|W_n(\omega_1) - W_n(\omega_2)\|^4 \right],$$

and sufficient conditions for (14) are that, there exists a $\delta > 1$ such that,

$$\|H_n(\omega_1) - H_n(\omega_2)\|^4 \leq K (\omega_2 - \omega_1)^\delta, \quad (15)$$

$$E \|W_n(\omega_1) - W_n(\omega_2)\|^4 \leq K (\omega_2 - \omega_1)^\delta. \quad (16)$$

In order to simplify notation, let consider $\xi(x)$ as a scalar, and $\tilde{n}^{-1} \leq \omega_2 - \omega_1$ because otherwise the result is immediate. Henceforth, η will denote an arbitrarily small positive constant. To prove

(15), notice that,

$$\begin{aligned}
|H_n(\omega_2) - H_n(\omega_1)| &\leq \frac{K}{\sqrt{\tilde{n}}} \sum_{s=[\tilde{n}\omega_1]+1}^{[\tilde{n}\omega_2]} \left| \xi\left(\frac{s}{\tilde{n}}\right) \left| \frac{E[I_{Xn}(\lambda_s)]}{2\pi h(\lambda_s, \theta_0)} - \sigma^2 \right| + Kn^{1/2} \left| \frac{1}{\tilde{n}} \sum_{s=[\tilde{n}\omega_1]+1}^{[\tilde{n}\omega_2]} \xi\left(\frac{s}{\tilde{n}}\right) - \int_{\omega_1}^{\omega_2} \xi(x) dx \right| \right| \\
&\leq \frac{K}{\sqrt{\tilde{n}}} \sum_{s=[\tilde{n}\omega_1]+1}^{[\tilde{n}\omega_2]} \left| \xi\left(\frac{s}{\tilde{n}}\right) \left| \frac{\log s}{s} + \frac{K}{\sqrt{\tilde{n}}} \sum_{s=[\tilde{n}\omega_1]+1}^{[\tilde{n}\omega_2]} \left| \int_{s/\tilde{n}}^{(s+1)/\tilde{n}} \left\{ \xi\left(\frac{s}{\tilde{n}}\right) - \xi(x) \right\} dx \right| \right| \right| \\
&\leq \frac{K}{\sqrt{\tilde{n}}} \log^3 n + \frac{K}{\sqrt{\tilde{n}}} \sum_{s=[\tilde{n}\omega_1]+1}^{[\tilde{n}\omega_2]} \frac{1}{s} \\
&\leq K |\omega_2 - \omega_1|^{1/2-\eta},
\end{aligned}$$

using Theorem 2 of Robinson (1995a), A3, and taking into account that $\tilde{n}^{-1} \leq \omega_2 - \omega_1$. We prove (16) reasoning as in the proof of Theorem 3.3. in Dahlhaus (1985). The ℓ -th cumulant, $\ell = 2, 4$, of the difference $W_n(\omega_2) - W_n(\omega_1)$, is

$$\text{cum}_\ell [W_n(\omega_2) - W_n(\omega_1)] = \frac{1}{n^{\ell/2}} \sum_{[\tilde{n}\omega_1] < s_1, \dots, s_\ell \leq [\tilde{n}\omega_2]} \prod_{j=1}^{\ell} \xi\left(\frac{s_j}{\tilde{n}}\right) \sum_{m=1}^{\ell} M_{s_1, \dots, s_\ell}(m)$$

where

$$\begin{aligned}
M_{s_1, \dots, s_\ell}(m) &= \sum_{i.p.(m)} \int_{\Pi^{2\ell-m}} \prod_{k=1}^m f_{|P_k|}(\beta_{\bar{P}_k}) \prod_{j=1}^{\ell} |\Delta(\lambda_{s_j})|^{-2} \\
&\quad \times \frac{1}{n^\ell} \prod_{j=1}^{\ell} D(\lambda_{s_j} - \beta_{a_j}) D(-\lambda_{s_j} - \beta_{b_j}) \lambda^{2\ell-m}(d\beta),
\end{aligned}$$

and the summation in $i.p.(m)$ runs for all indecomposable partitions $\{P_1, \dots, P_m\}$ of the $\ell \times 2$ table $\{a_j, b_j\}_{j=1}^{\ell}$, and $D(\lambda) = \sum_{t=1}^n e^{it\lambda}$. Writing $\Delta(\beta) := \Delta(e^{i\beta})$, the $|P_k|$ -th order spectra $f_{|P_k|}$ under A1 is equal to, up to a constant,

$$\prod_{r=1}^{|P_k|} \Delta(\beta_{c_r}) = \Delta\left(-\sum_{r=1}^{|P_k|} \beta_{c_r}\right) \prod_{r=1}^{|P_k|} \Delta(\beta_{c_r}),$$

for $P_k = \{c_1, \dots, c_{|P_k|}\}$ and $\bar{P}_k = \{c_1, \dots, c_{|P_k|-1}\}$, where $\beta_{c_{|P_k|}} = -\sum_{r=1}^{|\bar{P}_k|} \beta_{c_r}$ and $c_r \in \{a_j, b_j\}_{j=1}^{\ell}$. Then $\beta = (\beta_{\bar{P}_1}, \dots, \beta_{\bar{P}_m})$, with $|P_k| \geq 2$ because we can take $Ex_t = 0$ as $s_j \neq 0 \pmod n$.

Notice that, neither $f_{|P_k|}(\beta_{\bar{P}_k})$ nor $f_{|P_k|}(\beta_{\bar{P}_k}) \prod_{j=1}^{\ell} |\Delta(\lambda_{s_j})|^{-2}$ are uniformly bounded (in β), since we have not assumed any integrability conditions on $\Delta(\lambda)$, we can not adapt the rest of the proof of Theorem 3.3. of Dahlhaus (1985). Instead we use the results of Robinson (1995a, b).

For $\ell = 2$, and $m = 2$ all partitions have $|P_k| = 2$, $k = 1, 2$, and there are two type of combinations in $M_{s_1, s_2}(2)$, depending on whether there is a restriction on the frequencies s_1 and s_2 (variance

terms) or not (covariance), so for a generic positive constant K ,

$$|M_{s_1, s_2}(2)| \leq K 1_{\{s_1 = s_2\}} + K \frac{\log(\max(s_1, s_2))^2}{s_1 s_2},$$

where we have used Theorem 1 of Robinson (1995a), part (a) when $s_1 = s_2$ and $P_1 = \{a_1, b_2\}$, $P_2 = \{a_2, b_1\}$; and parts (c) and (d) for $s_1 \neq s_2$ and for any other partition, where retracing his proof we can check for the discrete Fourier transform $w_x(\lambda_{s_1})$ that

$$E \left[\frac{w_x(\lambda_{s_1}) w_x(-\lambda_{s_2})}{\Delta(\lambda_{s_1}) \Delta(-\lambda_{s_2})} \right] \leq K \frac{\log(\max(s_1, s_2))}{(s_1 s_2)^{1/2}}, \quad s_1, s_2 = 1, \dots, \tilde{n}. \quad (17)$$

When $m = 1$ the unique partition has a single component with $|P_1| = 4$ involving f_4 , and using that

$$\frac{1}{(2\pi)^3 n} \int_{\Pi^3} \prod_{j=1}^2 D(\lambda_{s_j} - \beta_{a_j}) D(-\lambda_{s_j} - \beta_{b_j}) \boldsymbol{\lambda}^3(d\beta) = 1,$$

noting the restriction $\sum_{j=1}^2 (\beta_{a_j} + \beta_{b_j}) = 0$ on β , we obtain that

$$|M_{s_1, s_2}(1)| \leq \frac{K}{n} + \frac{K}{n^2} \left| \int_{\Pi^3} \left\{ \prod_{r=1}^4 \Delta(\beta_{c_r}) \prod_{j=1}^2 \frac{1}{|\Delta(\lambda_{s_j})|^2} - 1 \right\} \prod_{j=1}^2 D(\lambda_{s_j} - \beta_{a_j}) D(-\lambda_{s_j} - \beta_{b_j}) \boldsymbol{\lambda}^3(d\beta) \right|. \quad (18)$$

Now (18) is bounded using the methods of Robinson (1995b, pp. 1649-51), factorizing the term in curly brackets in combinations of products of $[\Delta(\beta_{c_r}) \Delta(\pm \lambda_{s_j})^{-1} - 1]$, obtaining that

$$|M_{s_1, s_2}(1)| \leq \frac{K}{n} + K \sum_{a_i} \prod_{i=1}^3 \frac{1}{s_{a_i}^{1/2}},$$

uniformly in s_1, s_2 , where for each 3-tuple (a_1, a_2, a_3) , $a_i \in \{1, 2\}$, a particular $s_{a_i} \in \{s_1, s_2\}$ appears at most twice -and at least once.

Therefore, collecting terms for $m = 1, 2$, for $\eta > 0$ arbitrarily small,

$$\begin{aligned} & |\text{cum}_2 [W_n(\omega_2) - W_n(\omega_1)]| \\ & \leq \frac{K}{\tilde{n}} \sum_{[\tilde{n}\omega_1] < s_1, s_2 \leq [\tilde{n}\omega_2]} \left| \xi\left(\frac{s_1}{\tilde{n}}\right) \xi\left(\frac{s_2}{\tilde{n}}\right) \right| \left\{ 1_{\{s_1 = s_2\}} + \frac{\log(\max(s_1, s_2))^2}{s_1 s_2} + \frac{1}{\tilde{n}} + \sum_{a_i} \prod_{i=1}^3 \frac{1}{s_{a_i}^{1/2}} \right\} \\ & \leq \frac{K}{\tilde{n}} \sum_{[\tilde{n}\omega_1] < s \leq [\tilde{n}\omega_2]} \xi\left(\frac{s}{\tilde{n}}\right)^2 + \frac{K}{\tilde{n}} \log^6 \tilde{n} + \frac{K}{\tilde{n}^{1/2}} \log^2 \tilde{n} \left(\frac{1}{\tilde{n}} \sum_{s_1} \left(\frac{\tilde{n}}{s_1}\right)^{1/2} \right) \left(\sum_{s_2} \frac{1}{s_2} \right) \\ & \leq K |\omega_2 - \omega_1|^{1-\eta} \end{aligned}$$

using A3, $|\xi(s/\tilde{n})| \leq K \log n$, $\tilde{n}^{-1} \leq \omega_2 - \omega_1$, and that

$$\left| \frac{1}{\tilde{n}} \sum_{[\tilde{n}\omega_1] < s_1 \leq [\tilde{n}\omega_2]} \left(\frac{\tilde{n}}{s_1}\right)^{1/2} - \int_{\omega_1}^{\omega_2} \frac{1}{u^{1/2}} du \right| \leq \frac{K}{\tilde{n}^{1/2}} \leq K |\omega_2 - \omega_1|^{1/2}$$

so for $g(u) = 2u^{1/2}$ it holds that

$$\left| \int_{\omega_1}^{\omega_2} \frac{1}{u^{1/2}} du \right| \leq |g(\omega_2) - g(\omega_1)| \leq K |\omega_2 - \omega_1|^{1/2}.$$

For $\ell = 4$ and $m = 1$ there is a single partition in f_8 , such that

$$|M_{s_1, \dots, s_4}(1)| \leq K \sum_{a_i} \prod_{i=1}^7 \frac{1}{s_{a_i}^{1/2}},$$

generalizing the treatment for (18) in Robinson (1995b) to higher dimensions. Note that the contribution of this term for certain linear restrictions among the s_i , such that $\sum_j s_j = 0 \pmod n$, is $O(n^{-3})$, so it is included in the bound given. Similarly for $\ell = 4$ and $m = 2$ the partitions $\{P_1, P_2\}$ involve pairs of cumulant spectra like $f_2 f_6, f_3 f_5, (f_4)^2$. Each of the $m = 2$ integrals in $|\bar{P}_k| = |P_k| - 1$ dimensions is similar to expression (18), using that

$$\frac{1}{(2\pi)^{|\bar{P}_k|} n} \int_{\Pi^{|\bar{P}_k|}} \prod_{a_j, b_j \in P_k} D(\lambda_{s_j} - \beta_{a_j}) D(-\lambda_{s_j} - \beta_{b_j}) \boldsymbol{\lambda}^{i-1}(d\beta),$$

is either 1 if there is linear restriction among the s_j , i.e. $\sum_{a_j, b_j \in P_k} s_j = 0 \pmod n$, or either 0, otherwise. Thus for $|\bar{P}_k| = 1$ the integral corresponding to f_2 involve a typical term like

$$\frac{1}{2\pi n} \int_{\Pi} \left\{ \frac{|\Delta(\beta)|^2}{\Delta(\pm\lambda_{s_i})\Delta(\pm\lambda_{s_j})} - 1 \right\} D(\pm\lambda_{s_i} - \beta) D(\pm\lambda_{s_j} + \beta) d\beta, \quad (19)$$

whose contribution is bounded using (17), and in higher $d = |\bar{P}_k| \geq 2$ dimensions the integral is, uniformly in s_j , less than $K n^{(d+1)/2} \prod_{i=1}^d s_{a_i}^{-1/2}$ using the same technique as for the integral involving f_4 in (18). Considering all these cases we obtain that

$$|M_{s_1, \dots, s_4}(2)| \leq K \sum_{a_i} \prod_{i=1}^6 \frac{1}{s_{a_i}^{1/2}},$$

where any factor $s_{a_i} \in \{s_1, \dots, s_4\}$ in the summations in a_i appears at most twice in each summand.

For $m = 3$, reasoning in the same way, with typical contributions $(f_2)^2 f_4$ and $(f_3)^2 f_2$, we obtain

$$|M_{s_1, \dots, s_4}(3)| \leq K \sum_{a_i, b_i, c_i} \prod_{i=1}^2 \frac{\log(\max(s_{a_i}, s_{b_i}))}{(s_{a_i} s_{b_i} s_{c_i})^{1/2}},$$

so each $s_i \in \{s_1, \dots, s_4\}$ appears at most twice in each combination a_i, b_i, c_i . Finally for $m = 4$, the typical partition with $(f_2)^4$ has contribution

$$|M_{s_1, \dots, s_4}(4)| \leq K \sum_{a_i, b_i} \prod_{i=1}^4 \frac{\log(\max(s_{a_i}, s_{b_i}))}{(s_{a_i} s_{b_i})^{1/2}},$$

since the corresponding integrals are bounded if there is a restriction in the frequencies $a_i, b_i \in P_k$, while otherwise, the integrations are similar to (19) and bounded using (17).

Therefore, collecting the bounds for $|M_{s_1, \dots, s_4}(m)|$, $m = 1, \dots, 4$, and for some $\eta > 0$ arbitrarily small,

$$\begin{aligned} |\text{cum}_4 [W_n(\omega_2) - W_n(\omega_1)]| &\leq \frac{K}{n^2} \sum_{[\tilde{n}\omega_1] < s_1, \dots, s_4 \leq [\tilde{n}\omega_2]} \prod_{j=1}^4 \left| \xi \left(\frac{s_j}{\tilde{n}} \right) \right| \sum_{a_i, b_i} \prod_{i=1}^3 \left(\frac{\log(\max(s_{a_i}, s_{b_i}))}{s_{a_i} s_{b_i}} \right) \\ &\quad + \frac{K}{n} \log^7 n \left(\sum_{s_2} \frac{1}{s_2} \right)^2 \left(\frac{1}{n} \sum_{s_1} \left(\frac{n}{s_1} \right)^{1/2} \right)^2 \\ &\quad + \frac{K}{n} \log^7 n \left(\sum_{s_2} \frac{1}{s_2} \right)^3 \left(n^{-1} \sum_{s_1} 1 \right) \\ &\leq K \frac{1}{n} \log^{10} n |\omega_2 - \omega_1| \leq K |\omega_2 - \omega_1|^{2-\eta}, \end{aligned}$$

using that $|\xi(s/\tilde{n})| \leq K |\log \lambda_s| \leq K \log n$ uniformly in s .

Then the theorem follows using the product theorem of cumulants, i.e.

$$E |W_n(\omega_1) - W_n(\omega_2)|^4 = 3\text{cum}_2 [W_n(\omega_2) - W_n(\omega_1)]^2 + \text{cum}_4 [W_n(\omega_2) - W_n(\omega_1)],$$

because $E[W_n(\omega_1) - W_n(\omega_2)] = 0$. \square

PROOF OF PROPOSITION 1: Immediate from the last proposition.

PROOF OF PROPOSITION 2: A4 implies that, uniformly in $\omega \in [0, 1]$,

$$\begin{aligned} \frac{1}{\sqrt{\tilde{n}}} \sum_{j=1}^{[\tilde{n}\omega]} \frac{I_{nx}(\lambda_j)}{h(\lambda_j, \theta_0)} \left[\frac{h(\lambda_j, \theta_0)}{h(\lambda_j, \theta_n)} - 1 - \phi(\lambda_j, \theta_0) (\theta_n - \theta_0) \right] &\leq K \frac{\|\theta_n - \theta_0\|^2}{\sqrt{\tilde{n}}} \sum_{j=1}^{[\tilde{n}\omega]} \log \lambda_j \frac{I_{nx}(\lambda_j)}{h(\lambda_j, \theta_0)} \\ &= O_p \left(\frac{1}{\sqrt{\tilde{n}}} \right), \end{aligned} \quad (20)$$

by A3. This last inequality and A3 guarantee that, uniformly in $\omega \in [0, 1]$,

$$\begin{aligned} \hat{T}_n(\omega) &= \sqrt{\tilde{n}} \frac{F_n^0(\pi)}{\hat{F}_n(\pi)} \left\{ \left[\frac{\hat{F}_n(\pi\omega)}{F_n^0(\pi)} - \omega \right] - \omega \left[\frac{\hat{F}_n(\pi)}{F_n^0(\pi)} - 1 \right] \right\} \\ &= T_n^0(\omega) - R_n(\omega)' A_0^{-1} \sqrt{\tilde{n}} R_n(1) + o_p(1), \end{aligned}$$

where

$$R_n(\omega) := \frac{2\pi}{F_n^0(\pi) \tilde{n}} \sum_{j=1}^{[\tilde{n}\omega]} \phi(\lambda_j, \theta_0) \frac{I_{nx}(\lambda_j)}{h(\lambda_j, \theta_0)}.$$

After taking into account that, by Proposition 3, $\sup_{\omega \in [0, 1]} \|R_n(\omega) - R(\omega)\| = O_p(n^{-1/2})$, and using the fact that $1/\tilde{n} \sum_{j=1}^{\tilde{n}} \phi(\lambda_j, \theta_0) = \int_0^1 \phi(\pi\lambda, \theta_0) d\lambda + O(n^{-1} \log n)$, where $\int_0^1 \phi(\pi\lambda, \theta_0) d\lambda = 0$ by A0, and $F_n^0(\pi) = \sigma^2 + o_p(1)$ by Robinson (1995b) proof of Theorem 2.

PROOF OF COROLLARY 1: Proposition 1 and squared integrability of $\phi(\cdot, \theta_0)$ ensures fidis convergence of $\mathcal{L}T_n^0$. Tightness of $\mathcal{L}T_n^0$ follows from tightness of T_n^0 and continuity of R .

PROOF OF THEOREM 1: Since it is trivially satisfied that $\sup_\omega \|\overline{\mathcal{L}}R(\omega)\| = \sup_\omega |\overline{\mathcal{L}}\omega| = 0$, the Theorem is a consequence of the following facts,

$$E [\overline{\mathcal{L}}\bar{B}(r) \overline{\mathcal{L}}\bar{B}(s)] = \min(r, s), \quad (21)$$

$$\sup_{\omega \in [0, \omega_0]} \left| \overline{\mathcal{L}} \left(\hat{T}_n(\omega) - \mathcal{L}T_n^0(\omega) \right) \right| = o_p(1), \quad (22)$$

$$\overline{\mathcal{L}}T_n^0 \text{ converges in distribution to } \bar{B} \text{ in the space } \mathbb{D}[0, \omega_0]. \quad (23)$$

We prove (21) reasoning as Stute et al. (1998, Lemma 3.1). Define,

$$N(r) = \int_0^r \xi(u)' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1} \int_u^1 \xi(v) \bar{B}(dv) du.$$

Then,

$$E [\overline{\mathcal{L}}\bar{B}(r) \overline{\mathcal{L}}\bar{B}(s)] = E [\bar{B}(r) \bar{B}(s)] - E [\bar{B}(r) N(s)] - E [\bar{B}(s) N(r)] + E [N(r) N(s)].$$

Now, $E [\bar{B}(r) \bar{B}(s)] = \min(r, s)$,

$$\begin{aligned} E [\bar{B}(r) N(s)] &= E [\bar{B}(s) N(r)] \\ &= \int_0^{\min(r,s)} \xi(u)' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1} \int_u^1 \xi(v) dv du, \end{aligned}$$

and

$$\begin{aligned} E [N(r) N(s)] &= \int_0^r \int_0^s \left\{ \xi(u)' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1} \begin{bmatrix} 1-\max(u,v) & -R(\max(u,v)) \\ -R(\max(u,v))' & A_0(\max(u,v)) \end{bmatrix} \right. \\ &\quad \left. \times \begin{bmatrix} 1-v & -R(v) \\ -R(v)' & A_0(v) \end{bmatrix}^{-1} \xi(v) \right\} dv du \\ &= 2 \int_0^{\min(r,s)} \xi(u)' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1} \int_u^1 \xi(v) dv du, \end{aligned}$$

because by Fubini's theorem, $\int_0^r \int_0^s g(u, v) dudv = \int_0^r \int_0^u g(u, v) dv du + \int_0^s \int_0^v g(u, v) dudv$.

Now, we prove (22),

$$\begin{aligned} \bar{\mathcal{L}}\left(\hat{T}_n(\omega) - \mathcal{L}T_n^0(\omega)\right) &= \left[\hat{T}_n(\omega) - \mathcal{L}T_n^0(\omega)\right] - \left\{ \int_0^\omega \xi(u)' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1} \right. \\ &\quad \left. \times \int_u^1 \xi(v) \left(\hat{T}_n(dv) - \mathcal{L}T_n^0(dv)\right) du \right\}. \end{aligned} \quad (25)$$

Proposition 2 implies that the first term of (25) is $o_p(1)$ uniformly in $\omega \in [0, 1]$. Thus, it suffices to prove the uniform convergence of the second term, which follows by,

$$\begin{aligned} &\sup_{\omega \in [0,1]} \left| \int_\omega^1 \xi(v) \left[\hat{T}_n(dv) - \mathcal{L}T_n^0(dv)\right] \right| \\ &= \sup_{\omega \in [0,1]} \left| \frac{2\pi}{\sqrt{\tilde{n}}} \sum_{j=1+[\tilde{n}\omega]}^{\tilde{n}} \xi(\lambda_j) \left[\frac{I_{nx}(\lambda_j)}{\hat{F}_n(\pi)h(\lambda_j, \theta_n)} - \frac{I_{nx}(\lambda_j)}{F_n^0(\pi)h(\lambda_j, \theta_0)} \right. \right. \\ &\quad \left. \left. + \phi(\lambda_j, \theta_0) A_0^{-1} \frac{2\pi}{\tilde{n}F_n^0(\pi)} \sum_{j=1}^{\tilde{n}} \phi(\lambda_j, \theta_0) \frac{I_{nx}(\lambda_j)}{h(\lambda_j, \theta_0)} \right] \right| \\ &= o_p(1) \end{aligned}$$

applying A3 and Proposition 1 and (20). The proof is completed by proving (23). Fidis convergence follows by Proposition 1, (21) and (22). Then, it suffices to prove tightness of $\bar{\mathcal{L}}T_n^0$. Since T_n is tight, we only have to prove tightness of the term,

$$K_n(\omega) := \int_0^\omega H(u) \frac{2\pi}{\sqrt{\tilde{n}}} \sum_{j=1+[\tilde{n}u]}^{\tilde{n}} \left[\frac{I_{nx}(\lambda_j)}{\sigma^2 h(\lambda_j, \theta_0)} - 1 \right] \xi(\lambda_j) du,$$

where,

$$H(u) := \xi(u)' \begin{bmatrix} 1-u & -R(u) \\ -R(u)' & A_0(u) \end{bmatrix}^{-1}.$$

Now, for $r > s$, reasoning as in the proof of Proposition 3,

$$\begin{aligned} E |K_n(r) - K_n(s)|^2 &= \int_s^r \int_s^r H(u_1) H(u_2) \sum_{j=1+[\tilde{n}u_1]}^{\tilde{n}} \sum_{k=1+[\tilde{n}u_2]}^{\tilde{n}} E \left\{ \left[\frac{I_{nx}(\lambda_j)}{\sigma^2 h(\lambda_j, \theta_0)} - 1 \right] \left[\frac{I_{nx}(\lambda_k)}{\sigma^2 h(\lambda_k, \theta_0)} - 1 \right] \right\} \\ &\quad \times \frac{4\pi^2}{\tilde{n}} \xi(\lambda_j) \xi(\lambda_k) du_1 du_2 \\ &\leq K \int_s^r \int_s^r \|H(u_1)\| \|H(u_2)\| du_1 du_2 \\ &= K |L(r) - L(s)|^2, \end{aligned}$$

where $L(\cdot) = \int_0^\cdot \|H(u)\| du$ is a monotonic continuous function.

□.

REFERENCES

- [1] ADENSTEDT, I. (1974): "On large-sample estimation for the mean of a stationary random sequence," *Annals of Mathematical Statistics*, 2, 1095-1107.
- [2] ANDEL, J. (1986): "Long memory time series models," *Kybernetika*, 22, 105-123.
- [3] ANDERSON, T.W. (1991): "Goodness of fit tests for spectral distributions," Technical Report 27, Department of Statistics, Stanford University.
- [4] ANDERSON, T.W. (1993): "Goodness of fit tests for spectral distributions," *Annals of Statistics*, 21, 830-847.
- [5] ANDERSON, T.W. (1997): "Goodness-of-fit for autoregressive processes," *Journal of Time Series Analysis*, 18, 321-339.
- [6] BARLETT, M.S. (1954): "Problèmes de l'analyse spectrale des séries temporelles stationnaires," *Publications Institut de Statistique Université de Paris*, III, 119-134.
- [7] BARLETT, M.S. (1966): *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, 2nd ed. Cambridge University Press, London.
- [8] BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. Wiley, New York.
- [9] BLOOMFIELD, P. (1973), "An exponential model for the spectrum of a scalar time series," *Biometrika*, 60, 217-226.
- [10] BOX, G.E.P. AND D.A. PIERCE (1970): "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American Statistical Association*, 65, 1509-1526.
- [11] BLOOMFIELD, P. (1973): "An exponential model for the spectrum of a scalar time series," *Biometrika*, 60, 217-226.
- [12] BOX, G.E.P. AND G.M. JENKINS (1976): *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- [13] CHEN, H. AND J.P. ROMANO (1999): "Bootstrap-assisted goodness-of-fit tests in the frequency domain," *Journal of Time Series Analysis*, 20, 619-654.

- [14] DAHLHAUS, R. (1985): "On the asymptotic distribution of Barlett's U_p -statistic," *Journal of Time Series Analysis*, 6, 213-227.
- [15] DAHLHAUS, R. (1988): "Empirical spectral processes and their application to time series analysis," *Stochastic Processes and its Applications*, 30, 69-83.
- [16] DAHLHAUS, R. AND W. WEFELMEYER (1996): "Asymptotically optimal estimation in misspecified time series models," *Annals of Statistics*, 24, 952-974.
- [17] DELGADO, M.A. AND J. HIDALGO (2000): "Bootstrap goodness-of-fit test for linear processes," Preprint.
- [18] DURBIN, J. (1973): "Weak convergence of the sample distribution function when parameters are estimated," *Annals of Statistics*, 1, 274-290.
- [19] DURBIN, J., M. KNOTT AND C.C. TAYLOR (1975): "Components of Cramer-von Mises Statistics. II," *Journal of the Royal Statistical Society, Series B*, 37, 216-237.
- [20] FOX, R. AND M.S. TAQQU (1986): "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Annals of Statistics*, 14, 517-532.
- [21] GIRAITIS, L. AND D. SURGAILIS (1990): "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate," *Probability Theory and Related Fields*, 86, 87-104.
- [22] GIRAITIS, L., J. HIDALGO AND P.M. ROBINSON (2001): "Gaussian estimation of parametric spectral density with unknown pole," Forthcoming in *Annals of Statistics*.
- [23] GRANGER, C.W. AND R. JOYEUX. (1980): "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, 1, 15-29.
- [24] GRAY, H.L, N. ZHANG, AND W.A. WOODWARD (1989): "On generalized fractional processes," *Journal of Time Series Analysis*, 86, 87-104.
- [25] GRENANDER, U. AND M. ROSENBLATT (1957): *Statistical Analysis of Stationary Time Series*. John Wiley, New York.
- [26] HAINZ, G. AND DAHLHAUS, R. (2000): "Spectral domain bootstrap tests for stationary time series", Preprint.

- [27] HANNAN, E.J. (1970): *Multiple Time Series*. John Wiley, New York.
- [28] HANNAN, E.J. (1973): "The asymptotic theory of linear time series models," *Journal of Applied Probability*, 10, 130-145.
- [29] HONG, Y. (1996): "Consistent testing for autocorrelation of unknown form," *Econometrica*, 64, 837-864.
- [30] HOSKING, J.R.M. (1981): "Fractional differencing," *Biometrika*, 68, 165-176.
- [31] HOSKING, J.R.M. (1984): "Modeling persistence in hydrological time series using fractional differencing," *Water Resources Research*, 20, 1898-1908.
- [32] HOSOYA, Y. (1997): "A limit theory for long-range dependence and statistical inference on related models," *Annals of Statistics*, 25, 105-137.
- [33] HURVICH, C. M. AND B. K. RAY (1995), "Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes," *Journal of Time Series Analysis*, 16, 17-42.
- [34] IBRAGIMOV, I.A. (1963): "On estimation of the spectral function of a stationary Gaussian process," *Theory of Probability and its Applications*, 8, 366-401.
- [35] KHMALADZE, E.V. (1981): "Martingale approach in the theory of Goodness-of-fit tests," *Theory of Probability and its Applications*, 26, 240-257.
- [36] KHMALADZE, E.V. (1988): "An innovation approach to goodness-of-fit tests in \mathbb{R}^m ," *Annals of Statistics*, 16, 1503-1516.
- [37] KOKOSZKA, P. AND T. MIKOSCH (1997): "The integrated periodogram for long-memory processes with finite or infinite variance," *Stochastic Processes and their Applications*, 66, 55-78.
- [38] KOKOSZKA, P. AND T. MIKOSCH (2000): "The periodogram at the Fourier frequencies," *Stochastic Processes and their Applications*, 86, 49-79.
- [39] LJUNG, G.M. AND G.E.P. BOX (1978): "On a measure of lack of fit in time series models," *Biometrika*, 65, 297-303.
- [40] PAPANODITIS, E. (2000): "Spectral density based goodness-of-fit tests for time series models," *Scandinavian Journal of Statistics*, 27, 143-176.

- [41] PREWITT, K. (1993): "Goodness-of-fit test in parametric time series models differenced," *Journal of Time Series Analysis*, 19, 549-574.
- [42] ROBINSON, P.M. (1995a): "Log-periodogram regression for time series with long range dependence," *Annals of Statistics*, 23, 1048-1072.
- [43] ROBINSON, P.M. (1995b): "Gaussian semiparametric estimation of long-range dependence," *Annals of Statistics*, 23, 1630-1661.
- [44] SHORACK, G.R. AND J.A. WELLNER (1986): *Empirical Processes with Applications to Statistics*. John Wiley, New York.
- [45] SOLO, V. (1992), "Intrinsic random functions and the paradox of 1/f noise", *SIAM Journal of Applied Mathematics*, 52, 270-291.
- [46] STUTE, W. (1997): "Nonparametric model checks for regression," *Annals of Statistics*, 25, 613-641.
- [47] STUTE, W., S. THIES AND L.-X. ZHU (1998): "Model checks for regression: An innovation process approach," *Annals of Statistics*, 26, 1916-1934.
- [48] TANIGUCHI, M. (1981): "An estimation procedure of parameters of a certain spectral density model," *Journal of the Royal Statistical Society, Series B*, 43, 34-40.
- [49] VELASCO, C. (1999a), "Non-stationary log-periodogram regression," *Journal of Econometrics*, 91, 325-371, 1999.
- [50] VELASCO, C. (1999b), "Gaussian semiparametric estimation for non-stationary time series," *Journal of Time Series Analysis*, 20, 87-127.
- [51] VELASCO, C. AND P.M. ROBINSON (2000): "Whittle pseudo-maximum likelihood estimates for non-stationary time series," *Journal of the American Statistical Association*, 95, 1229-1243.
- [52] VELILLA, S. (1994): "A goodness-of-fit test for autoregressive-moving-average models based on the standardized sample distribution of the residuals," *Journal of Time Series Analysis*, 15, 637-648.

Table 1. Size of 5% Tests.

	$n = 200$				$n = 500$			
	C_n	C_n^0	Q_{10}	Q_{20}	C_n	C_n^0	Q_{15}	Q_{35}
$\phi, \text{AR}(1)$								
-0.8	4.91	5.07	5.59	6.18	5.09	5.05	5.47	6.13
-0.5	4.82	5.10	5.27	5.92	5.15	5.02	5.23	6.00
0.0	4.39	5.16	5.13	5.66	5.14	4.97	5.24	5.93
0.5	3.85	5.14	5.22	5.58	4.81	4.93	5.29	5.95
0.8	3.37	5.11	5.33	5.79	4.49	4.98	5.41	6.02
$d, \text{I}(d)$								
0.0	3.74	5.16	5.29	5.79	4.70	4.97	4.71	4.31
0.1	3.74	5.15	5.27	5.77	4.70	4.97	5.37	5.97
0.2	3.76	5.21	5.25	5.75	4.70	4.98	5.36	5.97
0.3	3.75	5.34	5.25	5.72	4.73	5.09	5.38	5.99
0.4	3.74	5.52	5.39	5.73	4.73	5.28	5.40	6.03

Table 2. Power of 5% Tests. $H_0 : \text{I}(d)$. $H_1 : \text{AR}(1), \phi$.

ϕ	$n = 200$			$n = 500$		
	C_n	Q_{10}	Q_{20}	C_n	Q_{15}	Q_{35}
0.2	11.01	14.64	12.78	34.85	26.81	20.07
0.5	26.23	41.60	32.02	74.75	83.11	59.50
0.8	9.45	26.66	21.62	32.53	65.20	46.12

Table 3. Power of 5% Tests. $H_0 : \text{AR}(1)$. $H_1 : \text{I}(d)$.

d	$n = 200$			$n = 500$		
	C_n	Q_{10}	Q_{20}	C_n	Q_{15}	Q_{35}
0.1	8.42	7.69	7.60	17.46	14.26	12.33
0.2	20.35	18.51	16.04	53.10	50.26	40.53
0.3	36.66	35.28	29.95	83.03	83.04	73.47
0.4	49.41	48.21	41.87	94.64	94.98	89.15