

SPECIFICATION TESTING FOR REGRESSION MODELS WITH DEPENDENT DATA

Javier Hidalgo*

London School of Economics

Discussion paper
No. EM/2007/518
May 2007

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6679

* The research was supported by ESRC Grant: Reference number R000239936. Also, I thank the helpful comments of two referees. Of course, all the usual caveats apply.

Abstract

We describe and examine a consistent test for the correct specification of a regression function with dependent data. The test is based on the supremum of the difference between the parametric and nonparametric estimates of the regression model. Rather surprisingly, the behaviour of the test depends on whether the regressors are deterministic or stochastic. In the former situation, the normalization constants necessary to obtain the limiting Gumbel distribution are data dependent and difficult to estimate, so to obtain valid critical values may be difficult, whereas in the latter, the asymptotic distribution may not be even known. Because of that, under very mild regularity conditions we describe a bootstrap analogue for the test, showing its asymptotic validity and finite sample behaviour in a small Monte Carlo experiment.

Key words: Functional specification. Variable selection. Nonparametric kernel regression. Frequency domain bootstrap.

JEL nos.: C14, C22.

1. INTRODUCTION

The main purpose of the present paper is to describe and analyze a consistent test for the correct specification of a regression function with dependent data. In empirical work, model selection (specification testing) and/or the selection of relevant regressors are, with no doubts, important features and they have a long standing as failure to do so can result in improper conclusions. For example, a wrong specification of the regression function, or conditional expectation, leads to wrong conclusions about the marginal effects that a control/regressor variable may induce on the dependent variable. On the other hand, if one of the regressors is a lag value of the variable of interest, then a wrong specification will result in wrong predictions for its future values.

Since Holly (1982) pointed out that standard “parametric” tests lack consistency in directions of possible interest, there has been an interest on tests which incorporate nonparametric features to construct consistent tests. To that end, there are mainly two approaches or methodologies. The first methodology is based on the empirical integrated regression function and the estimated parametric integrated regression function estimated under the null hypothesis. The basic intuition of the approach is based on exploiting orthogonality restrictions between the regressors and the errors of the specified or selected regression model. Although the literature is very broad, we can cite among others, Eubank and Spiegelman (1990), Hong and White (1995), Zheng (1996), Fan and Li (1996), Bierens and Ploberger (1997), Stute (1997), Koul and Stute (1999) and finally Delgado, Hidalgo and Velasco (2005) for the specification of a covariance stationary process.

However, in this paper the methodology that we follow is based on a direct comparison between estimates of the regression model under the null and the nonparametric alternative. More specifically, the test is based on global measures of discrepancy between the parametric and nonparametric estimates. For that purpose, we have chosen the norm of the supremum, e.g. a Kolmogorov-Smirnov type of test. An alternative norm is that given by the L_2 - *norm*, see among others Bickel and Rosenblatt (1973), Härdle and Mammen (1993) or Hidalgo (1999). Because the test is based on the supremum, as a by-product, we will examine uniform bounds for the nonparametric kernel regression estimators.

One of our, rather surprising, results is that the behaviour of the test is different depending on whether the regressors are deterministic or stochastic, and because of that we will explicitly differentiate the two cases or frameworks. More specifically, in the former situation, because the normalization constants to obtain the limiting Gumbel distribution is data dependent and difficult to estimate, to obtain valid critical values may be rather difficult if at all possible. On the other hand, in the latter setup, that is with stochastic regressors, the asymptotic distribution may not be even known. Moreover, the asymptotic distribution depends on the rate of convergence to zero of the bandwidth parameter chosen to estimate nonparametrically the regression model, as well as on the “combined dependence” of the regressor(s) and error term of the regression model. In this way the results are very much in spirit as those obtained for the least squares estimator when we allow that the regressors and error term may exhibit strong dependence. Hence, the previous arguments suggest the necessity to employ bootstrap methods.

The paper is organized as follows. In the next section, we describe the model and the test. Also, we present the regularity conditions and the kernel estimator of the regression function. Section 3 presents the uniform bands for the kernel estimates and the main results of the test when the regressors are deterministic or stochastic. Due to the nonstandard results obtained in Section 3, Section 4 describes and examines a bootstrap algorithm, showing the validity in our context.

The bootstrap is performed in the frequency domain and it extends results to the case when the errors are not necessarily weakly dependent. A small Monte Carlo experiment to examine the finite sample performance of the test in small samples is described in Section 5. Section 6 provides the proofs of our main results of Sections 3 and 4 which make use of a series of lemmas in Section 7.

2. THE MODEL AND TEST. REGULARITY CONDITIONS

Given a stretch of data, say $\{(y_t, x_t')'\}_{t=1}^n$, we consider the regression model

$$(2.1) \quad y_t = E(y_t | x_t) + u_t; \quad t = 1, \dots, n,$$

where we assume that the homoscedastic errors $\{u_t\}_{t \in \mathbb{Z}}$ follow a covariance stationary linear process. As we mentioned in the introduction, our main concern is to test the null hypothesis, without loss of generality,

$$(2.2) \quad H_0 : E(y_t | x_t) = \beta_0 + \beta' x_t,$$

where $\beta = (\beta_1, \dots, \beta_p)'$ is a parameter vector, being the alternative the negation of the null. It is worth pointing out that the sole motivation to examine the hypothesis $E(y_t | x_t) = \beta_0 + \beta' x_t$ with homoscedastic errors is basically for notational simplicity as well as to shorten the arguments of the already lengthy technical proofs. If we were interested to test for the somehow “more general” hypothesis $H_0 : E(y_t | x_t) = r(x_t; \theta_1)$ with heteroscedastic errors, or in other words that the data $\{(y_t, x_t')'\}_{t=1}^n$ follows the regression model

$$y_t = r(x_t; \theta_1) + \sigma(x_t; \theta_2) u_t,$$

then the only difference would be that we would obtain the residuals as

$$\hat{u}_t = \left(y_t - r(x_t; \hat{\theta}_1) \right) / \sigma(x_t; \hat{\theta}_2), \quad t = 1, \dots, n,$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of the parameters θ_1 and θ_2 , respectively. We make some extra comments at the beginning of Section 2.1 and in particular after Theorem 3.8 below.

2.1. DETERMINISTIC REGRESSORS.

In this section, we shall discuss the situation where the regressors are deterministic. A classical example of great interest in time series would be a polynomial trend, that is $x_t = (t, t^2, \dots, t^p)$ and/or with regressors of the type “ $\cos t\lambda_0$ ” and/or “ $\sin t\lambda_0$ ”, where $\lambda_0 \neq 0$. The latter type of regressors can be convenient when the practitioner suspects that the data may exhibit some cyclical behaviour. Hence, one possible hypothesis of interest is to know if such a deterministic trend and/or cyclical behaviour is a relevant component of the data. So, let us consider the hypothesis testing (2.2), having in mind that a particular case of great relevance is that $E(y_t | x_t) = \beta_0$.

As we have indicated in the introduction, the test for H_0 in (2.2) will be based on the difference between the parametric and nonparametric estimates of $E(y_t | x_t)$. That is, let

$$(2.3) \quad \hat{\theta} =: \left(\hat{\beta}_0, \hat{\beta}' \right)' = \left(\sum_{t=1}^n \tilde{x}_t \tilde{x}_t' \right)^{-1} \sum_{t=1}^n \tilde{x}_t y_t$$

be the least squares estimator (*LSE*) of $\theta = (\beta_0, \beta)'$ where henceforth we abbreviate $(1, x_t')'$ by \tilde{x}_t . Moreover, in what follows, for notational simplicity, we shall assume that the regressors x_t have been normalized. That is, $x_t' = (t/n, (t/n)^2, \dots, (t/n)^p)$.

On the other hand, our nonparametric estimator of $E(y|x = q/n)$ is the kernel regression estimator given by

$$(2.4) \quad \hat{r}(q/n) = \frac{1}{\tilde{n}} \sum_{t=1}^n y_t K\left(\frac{t-q}{\tilde{n}}\right),$$

where q is such that $\tilde{n} < q < n - \tilde{n}$ and where henceforth $\tilde{n} = [na]$ and $a = a(n)$ denotes the bandwidth parameter such that $a \rightarrow 0$ as n increases to infinity. Here $K(x)$ is a function on $[-1, 1] \rightarrow \mathbb{R}$.

Then, we have that the test for (2.2) is based on

$$(2.5) \quad \mathcal{T} = \sup_{\tilde{n} < q < n - \tilde{n}} \left| \hat{\theta} \tilde{x}_q - \hat{r}(q/n) \right|,$$

Let us introduce the following regularity conditions.

C1: $\{u_t\}_{t \in \mathbb{Z}}$ is a stationary linear process defined as

$$u_t = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}; \quad \sum_{j=0}^{\infty} \vartheta_j^2 < \infty, \text{ with } \vartheta_0 = 1,$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a zero mean iid sequence with $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(|\varepsilon_t|^\ell) = \mu_\ell < \infty$ for some $\ell > 4$. Also, for $j \geq 0$,

$$\vartheta_j = \sum_{k=0}^j \bar{\vartheta}_k b_{j-k}, \quad \bar{\vartheta}_0 = b_0 = 1,$$

and where, for $k \geq 1$, $\bar{\vartheta}_k = \ell(k) k^{d-1}$, $d \in [0, 1/2)$; $|\ell(k) - \ell(k+1)| < \ell'(k) k^{-1}$ with $\ell'(k) > 0$ and $\sum_{k=0}^{\infty} k^2 |b_k| < \infty$.

The case $d = 0$ refers to weak dependence, whereas the case $0 < d < 1/2$ refers to strong dependence. It is worth mentioning that it is also plausible to allow for $-1/2 < d < 0$. However to shorten, the otherwise technical mathematical arguments, we have preferred not to proceed with the latter case. Nevertheless, it appears that the former situation is the most relevant with real data.

Condition C1 is similar to that in Marinucci and Robinson (2003). It implies that

$$(2.6) \quad u_t = \sum_{j=0}^{\infty} \bar{\vartheta}_j \varepsilon'_{t-j}; \quad \varepsilon'_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}.$$

One model satisfying (2.6) is the *FARIMA*(p, d, q) process

$$(1-L)^d \Phi_p(L) u_t = \Theta_q(L) \varepsilon_t,$$

where $(1-L)^{-d} = \sum_{k=0}^{\infty} \bar{\vartheta}_k L^k$ with $\bar{\vartheta}_k = \Gamma(k+d) / (\Gamma(d) \Gamma(k+1))$, where $\Gamma(\cdot)$ denotes the gamma function such that $\Gamma(c) = \infty$ for $c = 0$ and $\Gamma(0) / \Gamma(0) = 1$, and $\Phi_p(L)$ and $\Theta_q(L)$ are the autoregressive and moving average polynomials with no common roots and outside the unit circle. The latter implies that $\Phi_p^{-1}(L) \Theta_q(L) = \sum_{j=0}^{\infty} b_j L^j$ with $b_j = O(j^{-c})$ for any $c > 0$. Denoting

$$(2.7) \quad C(\lambda) = \sum_{j=0}^{\infty} \bar{\vartheta}_j e^{-ij\lambda}; \quad B(\lambda) = \sum_{j=0}^{\infty} b_j e^{-ij\lambda}; \quad \lambda \in (-\pi, \pi],$$

the decomposition given in (2.6) implies that the spectral density function of $\{u_t\}_{t \in \mathbb{Z}}$, denoted $f(\lambda)$, can be factorized as

$$(2.8) \quad f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} g(\lambda) h(\lambda),$$

where $g(\lambda) = |C(\lambda)|^2$ and $h(\lambda) = |B(\lambda)|^2$. So, the condition $\sum_{k=0}^{\infty} k^2 |b_k| < \infty$ implies that $h(\lambda)$ is twice continuously differentiable for all $\lambda \in [0, \pi]$. An example of the function $\ell(k)$ is $\ell(k) = \log^{\iota} k$ for any $\iota > 0$ or any slowly varying function. Moreover, because by *C1*, the sequence $\{\bar{\vartheta}_k\}_{k=0}^{\infty}$ is of bounded variation, that is $\sum_{k=0}^{\infty} |\bar{\vartheta}_k - \bar{\vartheta}_{k+1}| < \infty$, and that because $\{\bar{\vartheta}_k\}_{k=0}^{\infty}$ is also a quasi-monotonic sequence, see Yong (1974, p.2) for a definition, then we have that $g(\lambda) \sim D\lambda^{-2d}$ as $\lambda \rightarrow 0+$, $0 < D < \infty$, and continuous differentiable outside any open set around the zero frequency, see Yong's (1974) Theorems III-11 and 12, and finally we have that

$$(2.9) \quad \begin{aligned} \gamma_j &= \int_{-\pi}^{\pi} g(\lambda) \cos(j\lambda) d\lambda \\ &\sim \zeta(d) j^{2d-1} \quad \text{as } j \rightarrow \infty, d \in (0, 1/2), \\ &= 2\pi \mathcal{I}(j=0), \quad d=0, \end{aligned}$$

where $\mathcal{I}(A)$ is the indicator function and $\zeta(d) = 2\Gamma(1-2d) \cos(\pi(\frac{1}{2}-d))$. It is worth mentioning that when examining the asymptotic properties of the supremum of $\hat{r}(q/n)$ we could allow for more than one singularity. However, this condition eases the notation and some of the arguments for the proof of the validity of the bootstrap described in Section 4 below.

C2: $K : [-1, 1] \rightarrow \mathbb{R}$ is **(a)** symmetric, **(b)** a continuous differentiable function in $(-1, 1)$ and **(c)** it integrates 1.

Recalling that $\tilde{n} = [na]$, the bandwidth parameter a satisfies:

C3: As $n \rightarrow \infty$, $(i) \tilde{n}^{-1} \rightarrow 0$ and $(ii) \tilde{n}^{\frac{1}{2}-d} a^2 \leq D < \infty$.

Part (ii) differs from the analogue assumed by Robinson (1997). We, contrary to the latter work, do not need to assume that $\tilde{n}^{\frac{1}{2}-d} a^2 \rightarrow 0$ as $n \rightarrow \infty$. This allows us to be able to choose the optimal bandwidth parameter a , in the sense of being the value a which minimizes the *MSE* of the nonparametric regression estimate. More precisely, suppose that $d=0$. Then, it is known that the optimal choice of a satisfies $a = Dn^{-1/5}$ for some finite positive constant D , which corresponds to the choice of the bandwidth parameter by, say, cross-validation.

2.2. STOCHASTIC REGRESSORS.

As we have suggested in the introduction, we shall consider the, anyway, relevant null hypothesis

$$(2.10) \quad H_0 : E(y_t | x_t) = \mu.$$

Observe that one leading situation of interest in model specification is to check the relevance of a regressor, see Delgado and Gonzalez-Manteiga (2001) or Hidalgo (1997) or Hidalgo (2000) in a distributed lag model to test for causality. That is, the null hypothesis in (2.10) against the alternative that $E(y_t | x_t)$ depends on x_t . The reason to consider the latter null hypothesis in (2.10) is to simplify some of the arguments, but at the same time it gives the main ingredients on the type of differences that we encounter when we compare the results with those obtained with deterministic regressors. However, after Theorem 3.8, we will discuss the main issues when the null is a nonlinear regression model with heteroscedastic errors.

As indicated in the introduction, the test will be based on the difference between the parametric and nonparametric estimates of $E(y_t | x_t)$. That is, let

$$\hat{\mu}_p = \frac{1}{n} \sum_{t=1}^n y_t$$

be the estimator of μ in $y_t = \mu + u_t$.

On the other hand, our nonparametric estimator of μ is the kernel regression estimator

$$(2.11) \quad \widehat{\mu}_{np}(x) = \frac{\sum_{t=1}^n y_t K_t(x)}{\sum_{t=1}^n K_t(x)},$$

where here $K(x)$ is a function $\mathbb{R} \rightarrow \mathbb{R}$, and we abbreviate $K((x_t - x)/a)$ by $K_t(x)$. Then, we have that the test for the null hypothesis in (2.10) is based on

$$(2.12) \quad \mathcal{T}_s = \sup_{x \in \mathcal{X}} |\widehat{\mu}_p - \widehat{\mu}_{np}(x)|,$$

where \mathcal{X} is a discrete set dense in the compact support of x_t such that for all x^1 and $x^2 \in \mathcal{X}$, we have that $\min |x^1 - x^2| > a \log n$.

Let us introduce the following regularity conditions.

S1: $C1$ is satisfied and $\{x_t\}_{t \in \mathbb{Z}}$ is a nonlinear transformation of a $\{\xi_t\}_{t \in \mathbb{Z}}$ sequence, that is $x_t = G(\xi_t)$, satisfying

$$\xi_t = \sum_{j=0}^{\infty} v_j \varsigma_{t-j}; \quad \sum_{j=0}^{\infty} v_j^2 < \infty, \text{ with } v_0 = 1,$$

where $\{\varsigma_t\}_{t \in \mathbb{Z}}$ is a Gaussian zero mean iid sequence with $E(\varsigma_t^2) = \sigma_\varsigma^2$. Also, for $j \geq 0$,

$$v_j = \sum_{k=0}^j \overline{\vartheta}_k^\xi b_{j-k}^\xi, \quad \overline{\vartheta}_0^\xi = b_0^\xi = 1,$$

and where, for $k \geq 1$, $\overline{\vartheta}_k^\xi = \ell(k) k^{d_\xi - 1}$, $d_\xi \in [0, 1/2)$; $|\ell(k) - \ell(k+1)| < \ell'(k) k^{-1}$ with $\ell'(k) > 0$ and $\sum_{k=0}^{\infty} |b_k^\xi| < \infty$.

Condition $S1$ might be relaxed to allow the sequence $\{\xi_t\}_{t \in \mathbb{Z}}$ to be a covariance stationary linear process with $\{\varsigma_t\}_{t \in \mathbb{Z}}$ be an *iid* sequence. However, since we wish to indicate the main results with stochastic regressors and the main differences with the previous situation of deterministic regressors, then to simplify the arguments we have preferred to keep $S1$ as it stands. The sole main difference is that in the latter case, we will employ Apple polynomials instead of Hermite ones for the expansion of $K_t(x)$, see for instance Hidalgo (1997) for some details.

S2: $K : \mathbb{R} \rightarrow \mathbb{R}$ is **(a)** symmetric, **(b)** a continuous differentiable function in \mathbb{R} and **(c)** it integrates 1.

3. RESULTS

3.1. DETERMINISTIC REGRESSORS.

We begin this section presenting the behaviour of the *LSE*, $\widehat{\theta}$, given in (2.3) and the kernel regression estimator in (2.4). To that end, let $G = \left((i+j-1)^{-1} \right)_{i,j=1,\dots,p+1}$. Then, we have the following result:

Proposition 3.1. *Let $\widehat{\theta}$ be the LSE in (2.3). Assuming $C1$, as $n \rightarrow \infty$,*

$$n^{1/2-d} \left(\widehat{\theta} - \theta_0 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{\sigma_\varepsilon^2}{2\pi} G^{-1} V G^{-1} \right),$$

where $V = (V_{i,j})_{i,j=1,\dots,p+1}$ and

$$V_{i,j} = \begin{cases} h(0) \zeta(d) \int_0^1 \int_0^1 v^{i-1} w^{j-1} |w-v|^{2d-1} dw dv & \text{if } d > 0 \\ h(0) \zeta(0) & \text{if } d = 0, \end{cases}$$

and where $h(0)$ and $\zeta(d)$ are defined in (2.8) and (2.9) respectively. Moreover, $E \left(\widehat{\beta}_k - \beta_k \right)^{2a} = O \left(n^{a(2d-1)} \right)$ for $k = 0, 1, \dots, p$, and integer $a = 1, 2$.

Proof. The proof follows immediately from Wang et al. (2003) and then standard arguments, so it is omitted. \square

Before we give the results for the uniform bound for the kernel regression estimate in (2.4), and then those for the test in (2.5), we first examine the covariance of $\widehat{r}(q) =: \widehat{r}(q/n)$ at two different points $q_1 \leq q_2$. Define $b(q_1, q_2) = (q_2 - q_1)/\check{n}$.

Proposition 3.2. *Assuming C1 – C3, for any $\check{n} < q_1 \leq q_2 < n - \check{n}$, as $n \rightarrow \infty$,*

$$(3.1) \quad \check{n}^{1-2d} \text{Cov}(\widehat{r}(q_1), \widehat{r}(q_2)) \rightarrow \rho(b, d),$$

where $b := \lim_{n \rightarrow \infty} b(q_1, q_2)$ is finite and

(a) if $0 < d < \frac{1}{2}$,

$$(3.2) \quad \rho(b; d) = \frac{\sigma_\varepsilon^2}{2\pi} h(0) \zeta(d) \int_{-1}^1 \int_{b-1}^{1+b} |v-w|^{2d-1} K(v) K(w-b) dv dw,$$

(b) if $d = 0$, $\rho(b; d) = \frac{\sigma_\varepsilon^2}{2\pi} h(0) \int_{-1}^1 K(v) K(v-b) dv$.

Proof. The proof is omitted since it follows by standard arguments following Robinson (1997). \square

Proposition 3.2 indicates that the covariance structure is independent of the points at which the regression function is estimated and only depends on the distance among the points where the regression function is estimated.

The next proposition deals with the correlation structure of $\widehat{r}(q)$ as $b(q_1, q_2) \rightarrow 0$ and when $b(q_1, q_2) \rightarrow \infty$ as $n \rightarrow \infty$. In what follows, D will denote a positive finite constant.

Proposition 3.3. *Under C1 – C3, for some $\alpha \in (0, 2]$, as $n \rightarrow \infty$,*

$$(a) \quad \frac{\rho(b(q_1, q_2); d)}{\rho(b(q_1, q_1); d)} - 1 = -D |b(q_1, q_2)|^\alpha + o(|b(q_1, q_2)|^\alpha) \quad \text{as } b(q_1, q_2) \rightarrow 0$$

$$(b) \quad \rho(b(q_1, q_2); d) \log(b(q_1, q_2)) = o(1) \quad \text{as } b(q_1, q_2) \rightarrow \infty.$$

Proposition 3.4. *Assuming C1 – C3, for any finite collection q_j , $j = 1, \dots, p$, such that $\check{n} < q_j < n - \check{n}$ and $|q_{j_1} - q_{j_2}| \geq nz > 0$, as $n \rightarrow \infty$,*

$$\check{n}^{\frac{1}{2}-d} \rho^{-\frac{1}{2}}(0; d) (\widehat{r}(q_j) - r(q_j))_{j=1, \dots, p} \xrightarrow{d} \mathcal{N}(0, \text{diag}(1, \dots, 1)).$$

The results of Propositions 3.2 and 3.4 indicates that $\widehat{r}(q) - r(q)$, where $r(q) = E(y_t | x = x_q)$, has asymptotically stationary increments. This observation together with Propositions 3.3 and 3.1 are key to obtain the asymptotic distribution of

$$\widetilde{T} = \sup_{\check{n} < q < n - \check{n}} |\widehat{r}(q) - r(q)|.$$

Theorem 3.5. *Let $v_n = (-2 \log a)^{1/2}$. Assuming that C1 – C3 holds, then*

$$\Pr \left\{ v_n \left(\check{n}^{\frac{1}{2}-d} \rho^{-\frac{1}{2}}(0; d) \widetilde{T} - \zeta_n \right) \leq x \right\} \xrightarrow[n \uparrow \infty]{} \exp(-2e^{-x}), \text{ for } x > 0,$$

where

(a) If $0 < d < 1/2$, then

$$\zeta_n = v_n + v_n^{-1} \left\{ \left(\frac{1}{2} - \frac{1}{\alpha} \right) \log \log a^{-1} + \log \left((2\pi)^{-\frac{1}{2}} 2^{\frac{2-\alpha}{2\alpha}} E^{\frac{1}{\alpha}} \mathcal{J}_\alpha \right) \right\}$$

for some $0 < E < \infty$, where α is as given in Proposition 3.3,

$$0 < \mathcal{J}_\alpha \equiv \lim_{a \rightarrow 0} \int_0^\infty e^s \Pr \left\{ \sup_{0 \leq t \leq [a]^{-1}} Y(t) > s \right\} ds < \infty$$

and $Y(t)$ is a stationary mean zero Gaussian process with covariance structure

$$\text{Cov}(Y(t_1), Y(t_2)) = |t_1|^\alpha + |t_2|^\alpha - |t_2 - t_1|^\alpha.$$

(b) If $d = 0$, then

$$\zeta_n = v_n + v_n^{-1} \left\{ \log \left(\frac{W_1(K)}{\pi^{1/2}} \right) + \frac{1}{2} \log(-\log a) \right\},$$

where

$$W_1(K) = \frac{K^2(-1) + K^2(1)}{2 \int_{-1}^1 K^2(x) dx}$$

when $W_1(K) > 0$, and otherwise

$$\zeta_n = v_n + v_n^{-1} \log \left[\frac{1}{\pi} \left(\frac{W_2(K)}{2} \right)^{1/2} \right],$$

where $W_2(K) = \left(\int_{-1}^1 K'(x)^2 dx \right) / \left(\int_{-1}^1 K(x)^2 dx \right)$ and $K'(x) = \frac{d}{dx} K(x)$.

Proof. The proof of this theorem follows step by step that of Theorem 4.3 below. So, we have decided to omit its proof. Observe that due to results in Wang et al. (2003), we can regard the distribution of the sequence of random variables $\{u_t\}_{t \in \mathbb{Z}}$ as Gaussian. In particular, we have that

$$\sup_{1 \leq s \leq n} \left| \sum_{t=1}^s K_t u_t - \sum_{t=1}^s K_t \bar{u}_t \right| = o_p \left(n^{d+1/4} \right),$$

where $\bar{u}_t = \sum_{j=0}^{\infty} \vartheta_j \bar{\varepsilon}_{t-j}$ and $\{\bar{\varepsilon}_t\}_{t \in \mathbb{Z}}$ is a zero mean iid sequence of standard normal random variables, and that in what follows we shall abbreviate $K_t(0)$ by K_t . \square

We now comment a bit on the previous theorem. As indicated in the introduction and (2.5), the asymptotic behaviour is based on the uniform behaviour of the nonparametric estimator $\hat{r}(q)$, and so on the construction of nonparametric uniform confidence bands. The latter is an old issue. For the kernel density estimator, say $\hat{p}(x)$, of the probability density function $p(x)$, Woodroffe (1967), under suitable conditions, showed that

$$(3.3) \quad \Pr \left\{ b_n^{1/2} \sup_{x \in \Xi_n} |\hat{p}(x) - p(x)| - c_n < z \right\} \xrightarrow{d} \exp(-2e^{-z}),$$

where $z \in \mathbb{R}^+$, $b_n = (-2\tilde{n} \log a)$ and $\Xi_n = \{0, a \log n, 2a \log n, \dots, 1\}$ is a particular grid (dense) of points in $[0, 1]$. The latter result was extended by Bickel and Rosenblatt (1973) when the supremum is over all $x \in \mathbb{R}$. But more importantly, as we see in Theorem 3.5, the possibility of strong dependence on $\{u_t\}_{t \in \mathbb{Z}}$ will imply that the normalization constants b_n and c_n in (3.3) depend crucially, in a nontrivial fashion, on the degree of dependence of the errors $\{u_t\}_{t \in \mathbb{Z}}$, measured by the so-called long memory parameter in contrast to the case when the errors $\{u_t\}_{t \in \mathbb{Z}}$ are, say, strong mixing. Later in Theorem 3.8 we will see the even more surprising result that the asymptotic distribution of the test depends on the combined strength of dependence of $\{u_t\}_{t \in \mathbb{Z}}$ and $\{x_t\}_{t \in \mathbb{Z}}$ and in several cases of interest also on the rate of convergence to zero of the bandwidth parameter a , when the regressors are stochastic.

In the context of regression models, (3.3) was extended, among others, by Johnston (1982) for iid errors u_t . However, in this paper we have relaxed this condition on u_t and in particular we have dropped any mixing dependence condition on $\{u_t\}_{t \in \mathbb{Z}}$. Recall that when we allow the data not to be weakly dependent, mixing conditions are incompatible with this, see for instance Ibragimov and Rozanov (1978). However, contrary to Csörgő and Mielniczuk (1995), we do not take the

supremum in a grid of points Ξ_n but, more in the line of Bickel and Rosenblatt (1973) on a much finer dense set of the interval $[0, 1]$. More specifically, on the natural set $\{n^{-1}, 2n^{-1}, \dots, 1\}$. Notice that if for instance the uniform kernel were chosen, then we will be in the same situation as in Bickel and Rosenblatt (1973), that is the supremum to compute $\tilde{\mathcal{T}}$, and thus the test \mathcal{T} , will be on the whole interval $[0, 1]$. This is because, in that case, $\hat{r}(x) = \hat{r}(q)$ for all $x \in [0, 1]$ such that $(q-1)/n < x \leq q/n$.

Theorem 3.5 indicates that the limiting distribution of $\tilde{\mathcal{T}}$ corresponds to the type of Gumbel distribution. We now give the main result of this section.

Theorem 3.6. *Under the same conditions of Theorem 3.5,*

$$\Pr \left\{ v_n \left(\tilde{n}^{\frac{1}{2}-d} \rho^{-\frac{1}{2}}(0; d) \mathcal{T} - \zeta_n \right) \leq x \right\} \xrightarrow[n \uparrow \infty]{} \exp(-2e^{-x}), \text{ for } x > 0.$$

Proof. The proof is obvious from Theorem 3.5 since $\hat{\theta} - \theta = O_p(n^{d-1/2})$ by Proposition 3.1 and then by C3. \square

Although Theorem 3.6 gives the asymptotic justification of \mathcal{T} in (2.5), we observe that for the former the normalization ζ_n needed to achieve the asymptotic distribution depends not only on d but on \mathcal{J}_α which may be very difficult to compute except for the special cases $\alpha = 1$ or 2 . In particular, for the latter cases, we know that $\mathcal{J}_2 = v_n + v_n^{-1} \log(\pi^{-1}(E/2)^{1/2})$ and $\mathcal{J}_1 = v_n + v_n^{-1} \log\left\{ (E/\pi)^{1/2} + 2^{-1} \log \log a^{-1} \right\}$, where E is a constant easy to obtain. More specifically, in our context, although d can be estimated, we face one potential difficulty to obtain the uniform bounds of $\hat{r}(q)$. As we observe from (the proof of) Proposition 3.3, α depends on the kernel $K(\cdot)$ as well as on d , so that to obtain \mathcal{J}_α does not seem an easy task.

3.2. STOCHASTIC REGRESSORS.

Before we give our main results for the asymptotic distribution of the test \mathcal{T}_s in (2.12) when x_t is stochastic, we introduce first some notation and definitions. We will abbreviate $K_t(x) - E(K_t(x))$ by $\tilde{K}_t(x)$. Also, let

$$H_j(\xi) = (-1)^j e^{\xi^2/2} \frac{\partial^j}{\partial \xi^j} e^{-\xi^2/2}$$

be the j th *Hermite's* polynomial. Recall that the Hermite polynomials satisfy

$$(3.4) \quad E(H_j(\xi_t) H_k(\xi_s)) = j! (E(\xi_t \xi_s))^j \mathcal{I}(j = k).$$

Finally, because $E\tilde{K}_t^2(x) < D$ by S2, we have the decomposition

$$(3.5) \quad \tilde{K}_t(x) = \sum_{j=1}^{\infty} \frac{b_{jn}(x)}{j!} H_j(\xi_t), \text{ with } b_{jn}(x) = \int_{\mathbb{R}} \tilde{K}_t(x) H_j(\xi_t) \phi(\xi_t) d\xi_t,$$

and where $\phi(\xi_t)$ is the probability density function of the standard normal random variable.

Let us introduce the following condition:

S3: (a) The instantaneous filter $G(\cdot)$ is continuously differentiable and the set $\{\xi \mid G'(\xi) = 0\}$ is finite.

(b) The probability density function of any finite-dimensional joint distribution of $\{x_s\}_{s=1}^S$ is continuously differentiable and uniformly bounded. Also denote by $p(x)$ the probability density function of x_t .

Then, we have the following result.

Proposition 3.7. *Assuming that S1– S3 hold, we have that*

(a) *if $d + d_\xi < 1/2$, then assuming C3, for any finite collection $\{x^q\}_{q=1}^Q \in \mathcal{X}$,*

$$(3.6) \quad \tilde{n}^{1/2} \left\{ \left(\hat{\mu}_{np}(x^q) - \hat{\mu}_p \right) \right\}_{q=1}^Q \xrightarrow{d} \mathcal{N} \left(0; \text{diag} \left(v^1, \dots, v^Q \right) \right),$$

where $v^q = p^{-1}(x^q) E u_t^2 \int_{\mathbb{R}} K^2(v) dv$, whereas

(b) *if $d + d_\xi > 1/2$, then*

$$\hat{\mu}_{np}(x) - \hat{\mu}_p = \frac{x\phi(x)}{p(x)} \left(\frac{1}{n} \sum_{t=1}^n \xi_t u_t - \frac{1}{n} \sum_{t=1}^n \xi_t \frac{1}{n} \sum_{t=1}^n u_t \right) (1 + o_p(1))$$

if $n^{1-2(d+d_\xi)} a^{-1} \rightarrow 0$, and if $n^{2(d+d_\xi)-1} a \rightarrow 0$, then (3.6) holds true.

The results of the previous proposition needs some comments. First, we observe that similarly to the properties of the *LSE*, the asymptotics or behaviour depends on whether the autocorrelation function of $z_t = \xi_t u_t$ is summable or not. The second is that in the situation of nonsummability of the autocorrelation of the process $\{z_t\}_{t \in \mathbb{Z}}$, the behaviour depends on the rate of the bandwidth parameter “ a ” to zero. When $n^{2(d+d_\xi)-1} a \rightarrow \infty$, we have that the asymptotic distribution is unknown.

Before we give the results for the behaviour of the test, let

$$\tilde{\mathcal{T}}_s = \sup_{x \in \mathcal{X}} \left| \left(\frac{\hat{p}(x)}{\hat{\sigma}_u^2 \int_{\mathbb{R}} K^2(v) dv} \right)^{1/2} (\hat{\mu}_p - \hat{\mu}_{np}(x)) \right|,$$

where

$$\hat{p}(x) = \frac{1}{\tilde{n}} \sum_{t=1}^n K_t(x); \quad \hat{\sigma}_u^2 = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{\mu}_p)^2.$$

Remark 3.1. *The motivation to modify \mathcal{T}_s to $\tilde{\mathcal{T}}_s$ is to obtain that asymptotically $\hat{p}^{1/2}(x) / \left(\hat{\sigma}_u^2 \int_{\mathbb{R}} K^2(v) dv \right)^{1/2} (\hat{\mu}_p - \hat{\mu}_{np}(x))$ are identically distributed.*

Then, we have the following result.

Theorem 3.8. *Assuming that S1– S3 hold, we have that*

(a) *if $d + d_\xi < 1/2$, and C3 holds, then,*

$$(3.7) \quad Pr \left\{ v_n \left(\tilde{n}^{1/2} \tilde{\mathcal{T}}_s - \zeta_n \right) \leq x \right\} \xrightarrow{n \uparrow \infty} \exp(-2e^{-x}),$$

for $x > 0$, where $v_n = (-2 \log 2a)^{1/2}$ and

$$\zeta_n = v_n + v_n^{-1} \log \left[\frac{1}{\pi} \left(\frac{W_2(K)}{2} \right)^{1/2} \right],$$

where $W_2(K) = \int_{\mathbb{R}} K'(x)^2 dx / \int_{\mathbb{R}} K(x)^2 dx$, whereas

(b) *if $d + d_\xi \geq 1/2$, then*

$$\mathcal{T}_s = \sup_{x \in \mathcal{X}} \left| \frac{x\phi(x)}{p^{1/2}(x)} \right| \left(\frac{1}{n} \sum_{t=1}^n \xi_t u_t - \frac{1}{n} \sum_{t=1}^n \xi_t \frac{1}{n} \sum_{t=1}^n u_t \right) (1 + o_p(1))$$

if $n^{1-2(d+d_\xi)} a^{-1} \rightarrow 0$, and if $n^{2(d+d_\xi)-1} a \rightarrow 0$, then (3.7) holds true.

Proof. The proof follows by Proposition 3.7 and then proceeding as in the proof of Woodroffe’s (1967) Theorem 5, after noticing that by Proposition 3.7 and standard chaining rule, we have that

$$\sup_{x \in \mathcal{X}} |\hat{p}(x) - p(x)| = o_p(1)$$

and that $\hat{\sigma}_u^2 \rightarrow_P \sigma_u^2$ by ergodicity of $\{u_t\}_{t \in \mathbb{Z}}$, and thus it is omitted. \square

Remark 3.2. Notice that $|x\phi(x)/p^{1/2}(x)|$ is bounded in \mathcal{X} by S3.

Similarly to the conclusions draw from Theorem 3.6, Theorem 3.8 indicates the asymptotic distribution of $\tilde{\mathcal{I}}_s$ may be unknown and it can vary depending on the chosen bandwidth. Under these circumstances, bootstrap algorithms appear to be a sensible procedure to go forward.

We finish this section discussing the differences we encounter with more “general” models with heteroscedasticity in the case of strong dependence in the error term. So, let us consider regression model

$$y_t = r(x_t; \theta) + \sigma(x_t) u_t,$$

where $\{x_t\}_{t=1}^n$ is as in S1. In this case, our results given in Proposition 3.7 and thus in Theorem 3.8 are a bit different and the conclusions are even more surprising. To that end, we first observe that in general $E(\sigma(x_t) \partial r(x_t; \theta) / \partial \theta) \neq 0$. So, denoting by $\hat{\theta}$ the *nonlinear least squares estimator* of θ , it is easy to show that under suitable regularity conditions, when $d > 0$,

$$\begin{aligned} \hat{\theta} - \theta &= \left(\sum_{t=1}^n \left(\frac{\partial r(x_t; \theta)}{\partial \theta} \frac{\partial r(x_t; \theta)}{\partial \theta'} \right) \right)^{-1} \sum_{t=1}^n \frac{\partial r(x_t; \theta)}{\partial \theta} \sigma(x_t) u_t (1 + o_p(1)) \\ &= \left(E \left(\frac{\partial r(x_t; \theta)}{\partial \theta} \frac{\partial r(x_t; \theta)}{\partial \theta'} \right) \right)^{-1} E \left(\sigma(x_t) \frac{\partial r(x_t; \theta)}{\partial \theta} \right) \frac{1}{n} \sum_{t=1}^n u_t (1 + o_p(1)) \\ &= V \frac{1}{n} \sum_{t=1}^n u_t (1 + o_p(1)). \end{aligned}$$

On the other hand, assuming that the function $r(x; \theta)$ is twice continuous differentiable in x , by standard kernel manipulations and proceeding as with the proof of Proposition 3.7, we have that

$$\begin{aligned} &\hat{\mu}_{np}(x) - r(x; \theta) \\ &= \frac{1}{\tilde{n}} \sum_{t=1}^n \sigma(x_t) K_t(x) u_t \left(\frac{1}{\hat{p}(x)} - \frac{1}{E\hat{p}(x)} \right) + \frac{1}{E\hat{p}(x)} \frac{1}{\tilde{n}} \sum_{t=1}^n \sigma(x_t) K_t(x) u_t + O_p(a^2). \end{aligned}$$

It is evident that the first term on the right has a smaller stochastic order of magnitude than the second one, so that we can conclude that

$$\hat{\mu}_{np}(x) - r(x; \theta) = \frac{1}{p(x)} \frac{1}{\tilde{n}} \sum_{t=1}^n \sigma(x_t) K_t(x) u_t (1 + o_p(1))$$

because $E\hat{p}(x) - p(x) = O(a^2)$. By standard arguments as those given in the proof of Proposition 3.7, because

$$\frac{1}{\tilde{n}} \sum_{t=1}^n \sigma(x_t) K_t(x) u_t = E(\sigma(x_t) K_t(x)) \frac{1}{n} \sum_{t=1}^n u_t (1 + o_p(1))$$

then, using the results for $\hat{\theta} - \theta$,

$$\begin{aligned} \hat{\mu}_{np}(x) - r(x; \hat{\theta}) &= \left\{ \frac{1}{p(x)} E(\sigma(x_t) K_t(x)) - V \frac{\partial r(x; \theta)}{\partial \theta} \right\} \frac{1}{n} \sum_{t=1}^n u_t (1 + o_p(1)) \\ &= \left\{ \sigma(x) - V \frac{\partial r(x; \theta)}{\partial \theta} \right\} \frac{1}{n} \sum_{t=1}^n u_t (1 + o_p(1)) \end{aligned}$$

because $a^{-1} E(\sigma(x_t) K_t(x)) = \sigma(x) p(x) + O(a^2)$. So denoting

$$\check{\mathcal{I}}_s = \sup_{x \in \mathcal{X}} \left| \left(\sigma(x) - V \frac{\partial r(x; \theta)}{\partial \theta} \right)^{-1} n^{1/2-d} \left(r(x; \hat{\theta}) - \hat{\mu}_{np}(x) \right) \right|$$

we have then the following result.

Proposition 3.9. *Assuming that S1– S3 hold, we have that*

$$\check{\mathcal{T}}_s \xrightarrow{d} \mathcal{N} \left(0, \frac{\sigma_\varepsilon^2}{2\pi} h(0) \zeta(d) \int_0^1 \int_0^1 |w-v|^{2d-1} dw dv \right).$$

4. THE BOOTSTRAP APPROACH

Since Efron (1979), bootstrap methods have become an often tool in empirical applications and it has attracted considerable effort to its development, being perhaps two the main motivations. First, bootstrap methods are capable of approximating the finite sample distribution of statistics more effectively than those based on their asymptotic counterparts. The second being that they allow computing valid asymptotic quantiles of the limiting distribution in situations where *a*) the limiting distribution is unknown or *b*) if known, the practitioner is unable to compute its quantiles. The basic idea of the bootstrap is in a broad sense to treat the estimated model as the true probability of the model or population. Depending on the underlying distributional properties of the data, different schemes have been adopted and proposed.

It is clear from the results of Theorems 3.6 and 3.8 that Efron's (1979) original bootstrap will not work. So, we need to employ schemes devise in a time series context. Among others, we could cite the moving block bootstrap (*MBB*) described in Künsch (1989) or the sieve-bootstrap of Bühlmann (1997) following ideas in Kreiss (1988). The aforementioned bootstrap schemes could be labeled as bootstrap procedures in the time domain. However, some alternative bootstraps have been proposed and considered using the frequency domain. Among others, we can cite Dahlhaus and Janas (1996), Theiler et al. (1992) or Prichard and Theiler (1994) or more recently Hidalgo (2003) in the context of regression models. Basically, the procedures are based on resampling the periodogram of the data $\{u_t\}_{t=1}^n$, as in the first two aforementioned papers, or similarly the modulus of its *discrete Fourier transform*, as in the latter three works. More specifically, the bootstraps approximate the modulus $\left| \sum_{j=0}^{\infty} v_j e^{ij\lambda} \right|^{-2} = \left| \sum_{j=0}^{\infty} \vartheta_j e^{ij\lambda} \right|^2$, where $\{\vartheta_j\}_{j=0}^{\infty}$ are the coefficients of the *MA*(∞) representation of $\{u_t\}_{t \in \mathbb{Z}}$,

$$u_t = \sum_{j=0}^{\infty} \vartheta_j \varepsilon_{t-j}; \text{ with } \vartheta_0 = 1.$$

The rationality of these bootstraps is that the normalized periodogram and/or *discrete Fourier transform* evaluated at the Fourier frequencies could be considered as uncorrelated, so that we can employ Efron's original bootstrap. However, the latter bootstrap may not be valid in some circumstances. One example being on bootstrapping the sample mean, so that in view of Theorem 3.8 part (*b*), say, it will not be valid for our problem of bootstrapping the distribution of \mathcal{T}_s in (2.12).

For the justification of the bootstrap in the time domain, one common assumption is that the data satisfy some mixing condition. For instance, when the data exhibit strong dependence, there is not justification for the validity of the sieve-bootstrap. Although results are given in Hall et al. (1998) for the subsampling of Politis and Romano's (1994) algorithm, it requires $v_t = H(u_t)$ with u_t to be Gaussian and $H(u) = u$ or $u^2 - E(u^2)$. However, as we will comment after Theorem 4.3, the subsampling is not a suitable or convenient method, even if it were valid, from an empirical and practical point of view.

Due to this, in this section we describe and examine a bootstrap algorithm in the frequency domain which, unlike the ones previously mentioned, it will approximate the transfer function $\sum_{j=0}^{\infty} \vartheta_j e^{ij\lambda}$ and more importantly it will be valid to bootstrap, say, the sample mean, which in our context plays an important role in all the asymptotics as Theorem 3.6, and also Theorem 3.8, indicates. The bootstrap is similar to that proposed by Hurvich and Zeger (1987), although they did not give its justification. In addition, our conditions are weaker than theirs and it applies to more general models. On the other hand, two differences of our bootstrap procedure with the block bootstrap, say, are that (a) it is not a subset of the original data, and (b) the bootstrap data, say $\{u_t^*\}_{t=1}^n$, is covariance stationary in the sense that $Cov^*(u_t^*, u_s^*)$ is a function of $|t - s|$. Herewith, by $Cov^*(z_1, z_2)$ or, say $E^*(z)$, we mean the covariance or expectation conditional on the data. In addition, the bootstrap is valid for data which do not satisfy the standard mixing conditions, such as strong mixing. That is, we shall allow our data to exhibit strong dependence. The bootstrap described below has some similarities with the sieve-bootstrap as it approximates the transfer function $\sum_{j=0}^{\infty} \vartheta_j e^{ij\lambda}$ instead of its modulus as that in, say, Theiler et al. (1992) or Hidalgo (2003). More specifically, our bootstrap is based on the “discrete” Cramér representation of $\{u_t\}_{t=1}^n$ and the Bartlett’s approximation of the *discrete Fourier transform* of $\{u_t\}_{t=1}^n$ by that of the innovations $\{\varepsilon_t\}_{t=1}^n$. The proposed bootstrap enjoys other properties, like the sieve-bootstrap, such as the ability to cope in a natural form with missing observations or unequal spaced data, which is very difficult to implement using subsampling or the *MBB*. On the other hand, the difference between the sieve-bootstrap and ours parallels to that existing when the practitioner is interested to estimate the spectral density function by either (a) approximating the dependence structure of the data by an *AR*(p_n) model, as in Berk (1974), and $p_n \rightarrow \infty$ with n or (b) the average periodogram, see Brillinger (1981).

Finally it is worth recalling that in our framework, the resampling or bootstrap scheme must satisfy two basic requirements. First, the conditional distribution, given the data, of the bootstrap analogue of say \mathcal{T} , and denoted by \mathcal{T}^* , consistently estimates the distribution of \mathcal{T} under H_0 and local alternatives. That is, $\mathcal{T}^* \rightarrow_{d^*} \mathcal{T}$ in probability, where “ \rightarrow_{d^*} ” in probability means convergence in bootstrap distribution according to Giné and Zinn (1989). And second, as suggested elsewhere, to have good power \mathcal{T}^* must also converge in bootstrap distribution, possibly to a different one, under the alternative hypothesis. However, in our case, as Theorem 4.3 indicates, the bootstrap distribution converges to that of \mathcal{T} under both the null and alternative hypothesis.

4.1. DETERMINISTIC REGRESSORS.

When the regressors x_t are deterministic the bootstrap algorithm for \mathcal{T} in (2.5), is described in the following *7 STEPS*.

STEP 1: Obtain the residuals $\hat{u}_t = y_t - \hat{\theta}' \tilde{x}_t$, $t = 1, \dots, n$.

Our second step describes how to obtain $w_\varepsilon^*(\lambda_j)$, $j = 1, \dots, \tilde{n}$, with $\tilde{n} = [n/2]$.

STEP 2: Let $\{\varepsilon_t^*\}_{t=1}^n$ be a random sample with replacement from the standard normal random variable and obtain its *discrete Fourier transform*

$$\eta_j^* := w_\varepsilon^*(\lambda_j) = \frac{1}{n^{1/2}} \sum_{t=1}^n \varepsilon_t^* e^{-it\lambda_j}, \quad j = 1, \dots, \tilde{n}.$$

Remark 4.1. *The previous step reminds very much the wild bootstrap. We conjecture that using results in Csörgő and Mason (1989) we could use the empirical distribution function of $\{\tilde{u}_t\}_{t=1}^n$, where $\tilde{u}_t = \tilde{\sigma}_u^{-1} \hat{u}_t$, $\tilde{\sigma}_u^2 = n^{-1} \sum_{t=1}^n \hat{u}_t^2$, instead of the normal random variable. However, the latter eases considerably the proof for*

the validity of the bootstrap. Finally, observe that, by periodicity, $\eta_j^* = \bar{\eta}_{-j+n}^*$ for $j = 1, \dots, \tilde{n}$, where \bar{a} denotes the complex conjugate of the number a .

STEP 3: We estimate d, \hat{d} , by say Robinson's (1995b) *GSE*, defined as

$$(4.1) \quad \hat{d} = \arg \min_{d \in [0, \Delta]} \tilde{R}(d),$$

where $0 < \Delta < 1/2$, and

$$\tilde{R}(d) = \log \left(\frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_{\widehat{u}\widehat{u}}(\lambda_j) \right) - 2d \sum_{j=1}^m \log \lambda_j$$

for integer $m \in [1, \tilde{n})$ and where $I_{\widehat{u}\widehat{u}}(\lambda) = |w_{\widehat{u}}(\lambda)|^2 / (2\pi)$ is the periodogram of $\{\widehat{u}_t\}_{t=1}^n$, with $m^{-1} + mn^{-1} \rightarrow 0$ as $n \rightarrow \infty$.

Define the estimator of $\sigma_\varepsilon^2 h(\lambda) / 2\pi$ by

$$(4.2) \quad \widehat{h}(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^m \left| 1 - e^{-i(\lambda+\lambda_j)} \right|^{2\widehat{d}} I_{\widehat{u}\widehat{u}}(\lambda + \lambda_j).$$

Remark 4.2. We can regard $\widehat{h}(\lambda)$ as an estimator of $2\pi g(\lambda) |1 - e^{-i\lambda}|^{2d} h(\lambda)$. When the coefficients $\bar{\vartheta}_k$ in C1 are given by $\bar{\vartheta}_k = \Gamma(k+d) / (\Gamma(d)\Gamma(k+1))$, that is $g(\lambda) = |1 - e^{-i\lambda}|^{-2d}$, then $\widehat{h}(\lambda)$ becomes an estimator of $2\pi h(\lambda) = |B(\lambda)|^2$. However, because $2\pi g(\lambda) |1 - e^{-i\lambda}|^{2d} h(\lambda) =: \left| \sum_{j=0}^{\infty} \nu_j e^{ij\lambda} \right|^2$ is a symmetric twice continuously differentiable function in $\lambda \in [0, \pi]$, we have that $\{\nu_j\}_{j \geq 0}$ satisfies the same conditions as $\{b_j\}_{j \geq 0}$ in C1. Hence in what follows, and without loss of generality, to simplify the notation we shall not make any distinction between $\{b_j\}_{j \geq 0}$ and $\{\nu_j\}_{j \geq 0}$.

Define the truncated version of $|1 - e^{i\lambda}|^{-2d}$,

$$\tilde{g}(\lambda; d) = \sum_{\ell=1-n}^{n-1} \gamma(\ell; d) e^{i\ell\lambda},$$

where

$$\gamma(\ell; d) = \frac{(-1)^\ell \Gamma(1-2d)}{\Gamma(\ell-d+1)\Gamma(1-\ell-d)}.$$

STEP 4: Consider $M = [n/4m]$ and compute

$$u_t^* = \frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} \tilde{g}^{1/2}(\lambda_j; \widehat{d}) \widehat{B}(\lambda_j) \eta_j^*, \quad t = 1, \dots, n,$$

where

$$(4.3) \quad \widehat{B}(\lambda_j) = 1 + \widehat{b}_1 e^{-i\lambda_j} + \dots + \widehat{b}_M e^{-iM\lambda_j}$$

$$\widehat{b}_\ell = n^{-1} \sum_{j=-\tilde{n}+1}^{\tilde{n}-1} \widehat{A}(\lambda_j) e^{i\ell\lambda_j}, \quad \ell = 1, \dots, M$$

with

$$\widehat{A}(\lambda) = \exp \left\{ \sum_{r=1}^M \widehat{c}_r e^{-ir\lambda} \right\}$$

and

$$\widehat{c}_r = \tilde{n}^{-1} \sum_{\ell=1}^{\tilde{n}} \log \left(\widehat{h}(\lambda_\ell) \right) \cos(r\lambda_\ell); \quad r = 0, \dots, M.$$

Remark 4.3. *It is worth mentioning that the way to obtain the bootstrap observations u_t^* in STEP 4, together with the definition of $\hat{h}(\lambda)$ in (4.2), has some similarities with the autoregressive-aided bootstrap in Kreiss and Paparoditis (2003).*

$\hat{B}(\lambda)$ in (4.3) is an estimator of $B(\lambda)$ in (2.7), whereas $\hat{A}(\lambda)$ is an estimator of $A(\lambda) = \exp\{\sum_{r=1}^{\infty} c_r e^{-ir\lambda}\}$, with $c_r = \pi^{-1} \int_0^\pi \log h(\lambda) \cos(r\lambda) d\lambda$, which comes from the so-called canonical spectral decomposition of $h(\lambda)$, see for instance Brillinger (1981, p. 78 – 79).

STEP 5: Compute

$$y_t^* = \hat{\theta}' \tilde{x}_t + \left(2\pi e^{\hat{c}_0}\right)^{1/2} u_t^*; \quad t = 1, \dots, n.$$

Note that by the well-known Kolmogorov formula, $2\pi e^{\hat{c}_0}$ is an estimator of the variance of ε_t , σ_ε^2 , that is the one-step mean square prediction error.

STEP 6: Compute $\hat{r}^*(t)$, $t = \tilde{n} + 1, \dots, n - \tilde{n} - 1$, as in (2.4) but with y_t replaced by y_t^* with the same bandwidth parameter “ a ”. Also denote the bootstrap least squares estimator of $\hat{\theta}$ by $\hat{\theta}^*$.

Remark 4.4. *We should emphasize that we use the same bandwidth parameter “ a ” to estimate nonparametrically the regression model with the observed and bootstrap data. This is very important and crucial to validate the bootstrap algorithm to make inferences. See the comments we make after Theorem 4.3 for a more theoretical, and also practical, reason.*

Our final step is:

STEP 7: Compute the bootstrap version of \mathcal{T} as

$$(4.4) \quad \mathcal{T}^* = \sup_{\tilde{n} < t < n - \tilde{n}} \left| \hat{\theta}^{*'} \tilde{x}_t - \hat{r}^*(t) \right|.$$

Remark 4.5. *It is worth mentioning that the bootstrap approach will be also valid if instead of computing $\{\hat{u}_t\}_{t=1}^n$ as in STEP 1, we would have obtained them using the alternative hypothesis. That is, if we would have computed the residuals as*

$$(4.5) \quad \hat{u}_t = y_t - \hat{r}(t), \quad t = 1, \dots, n.$$

The latter may induce a better power for the test. However, the motivation to keep STEP 1 is merely to simplify, and make then clearer, the already lengthy and technical proofs of our results. The reason is twofold. The first one is due to the fact that when we obtain the residuals \hat{u}_t as in (4.5), the rate of convergence to zero of $\hat{u}_t - u_t$ can be very slow as was seen by Robinson (1997). Because of that, it implies that extra technical arguments will be needed to validate the bootstrap \mathcal{T}^ in, say, (4.4). The second motivation is more of a theoretical reason and/or its practical implication. Indeed, if (4.5) were to be used to obtain the residuals \hat{u}_t , we would have to modify STEP 1 as*

STEP 1*: Obtain the centred residuals $\hat{u}_t = \tilde{u}_t - n^{-1} \sum_{t=1}^n \tilde{u}_t$, $t = 1, \dots, n$, where $\tilde{u}_t = y_t - \hat{r}(t)$ with

$$(4.6) \quad \hat{r}(t) = \begin{cases} \hat{r}_+(t), & t = 1, \dots, \tilde{n} \\ \frac{1}{2}(\hat{r}_+(t) + \hat{r}_-(t)), & \tilde{n} < t < n - \tilde{n} \\ \hat{r}_-(t), & t = n - \tilde{n}, \dots, n, \end{cases}$$

and $\hat{r}_+(t)$ and $\hat{r}_-(t)$ are given by

$$\hat{r}_\pm(t) = \frac{1}{\tilde{n}} \sum_{s=1}^n y_s K_\pm\left(\frac{t-s}{\tilde{n}}\right)$$

and $K_+(x)$ and $K_-(x)$ be one-sided kernels integrating 1, that is kernel functions taking values for $x > 0$ and $x < 0$, respectively.

STEPS 2* to 4*: As those in *STEPS 2 to 4*.

STEP 5*: Compute $\widehat{r}(t)$ as in (4.6) but with a bandwidth parameter e such that $e \rightarrow 0$ and $a = o(e)$, and

$$y_t^* = \widehat{r}(t) + \left(2\pi e^{\widehat{c}_0}\right)^{1/2} u_t^*; \quad t = 1, \dots, n.$$

We now comment on *STEP 5**. The motivation to use a pilot bandwidth parameter e such that $a = o(e)$ is to provide a correct adjustment of the bias of the kernel nonparametric estimator of $r(\cdot)$. The reason is because by standard arguments $\varphi(q) = E(\widehat{r}(q) - r(q)) = o(\check{n}^{d-1/2})$, whereas the bootstrap bias constructed from the resampled data is $\varphi^*(q) = E^*(\widehat{r}^*(q) - \widehat{r}(q))$. Now a new condition regarding the rate of convergence to zero of the extra bandwidth parameter e is required. In particular, we would need that:

As $n \rightarrow \infty$, (i) $(ne)^{-1} \rightarrow 0$, (ii) $n^{\frac{1}{2}-d} e^{\frac{5}{2}-d} \leq D < \infty$.

STEP 6*: Compute $\widehat{r}^*(q)$, $q = \check{n} + 1, \dots, n - \check{n} - 1$, as in (2.4) but with y_t replaced by y_t^* and the same bandwidth parameter a employed in *STEP 1** to compute $\widehat{r}(t)$ in (4.6).

And finally,

STEP 7*: The same as *STEP 7*.

So, apart from the technical problems that *STEP 1** will induce, there exists the potential problem of the necessity to choose another bandwidth parameter, being the latter not a very desirable requirement. Nevertheless, as the Monte Carlo experiment in the next section indicates, the power properties of the bootstrap algorithm described in *STEPS 1 to 7* appears to be very good. So, no major power loss is expected with the simpler proposed bootstrap.

The next proposition examines the behaviour of \widehat{d} given in (4.1).

Proposition 4.1. Under *C1*, $\widehat{d} - d = O_p\left(m^{-1/2} + (m/n)^2 + m^{2d}/n\right)$.

Proof. The proof is omitted as it follows similarly to that of Robinson's (1997) Theorem 3, after noting that in our situation we do not have his terms $I_{\xi\xi}$ and $I_{\chi\chi}$ and taking $a^{-1}k(v) = 1$ for all v there. \square

We shall slightly modify Condition *C1* to:

C1': *C1* holds with $\bar{\vartheta}_k = \Gamma(k+d) / (\Gamma(d)\Gamma(k+1))$ there.

The modification of *C1* to *C1'* implies that $g(\lambda) = |1 - e^{-i\lambda}|^{-2d}$ in (2.8). Also

C4: $m^4/n^3 + n^2/m^3 \rightarrow 0$ as $n \rightarrow \infty$.

Proposition 4.2. Assuming *C1'*, *C2* – *C4*, we have that for $0 \leq q \leq n - \check{n}$,

$$\frac{1}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} |E^*(u_t^* u_{t+v}^*) - \delta_{|v|}| = o_p(1).$$

Theorem 4.3. Assuming *C1'*, *C2* – *C4*, under the maintain hypothesis,

$$\Pr \left\{ v_n \left(\check{n}^{\frac{1}{2}-\widehat{d}} \rho^{-\frac{1}{2}} \left(\theta; \widehat{d} \right) \mathcal{T}^* - \zeta_n \right) \leq x \mid \mathcal{U}_n \right\} \xrightarrow{P} \exp(-2e^{-x}), \quad \text{for } x > 0,$$

where $\mathcal{U}_n = \{u_t\}_{t=1}^n$ and v_n and ζ_n were defined in Theorem 3.5.

We now comment on Theorem 4.3. We see that to obtain a critical value, say $x(\beta)$, for which $\exp(-2e^{-x(\beta)}) = 1 - \beta$, is the same as to find the value, say z_n^* , such that

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{T}^* \leq z_n^* \mid \mathcal{U}_n \} = 1 - \beta.$$

Now, this value z_n^* depends on ζ_n and v_n and thus on the choice of the bandwidth parameter a . In fact, because we cannot compute the normalization constant ζ_n ,

in practice we would compute z_n^* . The latter implies that for the bootstrap to be valid, we need z_n^* to satisfy that $|z_n^*/z_n - 1| \xrightarrow{p} 0$, where z_n is such that

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{T} \leq z_n \} = 1 - \beta,$$

e.g. the corresponding value with the observed data. But, for $|z_n^*/z_n - 1| \xrightarrow{p} 0$ to hold true, it is obvious that we need the normalization constants ζ_n and v_n to be the same for both \mathcal{T} and \mathcal{T}^* , or at least their ratios converging to one. This is obviously possible *if only if* the bandwidth parameters a are the same when estimating the regression function with both the original $\{y_t\}_{t=1}^n$ and bootstrap $\{y_t^*\}_{t=1}^n$ data. The latter indicates that if we employ the subsampling algorithm, then the same bandwidth parameter is to be used when estimating nonparametrically the regression function with the full sample size and the subsample. This does not appear to be either advisable or sensible, especially in small samples, since a can be relatively too large when we use the subsample. In particular, in our situation we can end up with the situation where $a^{-1} \simeq c$, implying that the nonparametric estimator of $r(t)$ can be very flat or even constant. Finally, the scheme will involve another “parameter” to be used, namely the size of the subsample which is not clear how to be chosen.

4.2. STOCHASTIC REGRESSORS.

The bootstrap for stochastic regressors is exactly the same as with the deterministic ones. The only differences are in *STEPS 1* and *5* to *7*. Namely,

STEP 1: Obtain the residuals $\hat{u}_t = y_t - \hat{\mu}_p$, $t = 1, \dots, n$.

STEPS 2-4: As in Section 4.1.

STEP 5: Compute

$$y_t^* = \hat{\mu}_p + \left(2\pi e^{\hat{c}_0}\right)^{1/2} u_t^*; \quad t = 1, \dots, n.$$

STEP 6: Compute $\hat{\mu}_{np}^*(x)$, $x \in \mathcal{X}$ as in (2.11) but with y_t replaced by y_t^* with the same bandwidth parameter “ a ”. Also denote the bootstrap estimator of $\hat{\mu}_p$ by $\hat{\mu}_p^*$. That is, $\hat{\mu}_p^* = n^{-1} \sum_{t=1}^n y_t^*$.

Remark 4.6. *It should be emphasized once again that the bandwidth parameter “ a ” employed to estimate the sample mean with the observed and bootstrap data has to be the same. This is very important to validate the used of the bootstrap to make inferences. See the comments we made after Proposition 3.7 and the results given in Theorem 3.8.*

Our final step is:

STEP 7: Compute the bootstrap version of \mathcal{T} as

$$\mathcal{T}_s^* = \sup_{x \in \mathcal{X}} \left| \hat{\mu}_p^* - \hat{\mu}_{np}^*(x) \right|,$$

and also let

$$\tilde{\mathcal{T}}_s^* = \sup_{x \in \mathcal{X}} \left| \left(\frac{\hat{p}(x)}{\hat{\sigma}_{u^*}^2 \int_{\mathbb{R}} K^2(v) dv} \right)^{1/2} (\hat{\mu}_p^* - \hat{\mu}_{np}^*(x)) \right|,$$

where

$$\hat{\sigma}_{u^*}^2 = \frac{1}{n} \sum_{t=1}^n (y_t^* - \hat{\mu}_p^*)^2.$$

Proposition 4.4. *Assuming that S1–S3 hold, we have that*

(a) *if $d + d_\xi < 1/2$, then assuming C3, for any finite collection $\{x^q\}_{q=1}^Q \in \mathcal{X}$,*

$$(4.7) \quad \tilde{n}^{1/2} \left\{ (\hat{\mu}_{np}^*(x^q) - \hat{\mu}_p^*) \right\}_{q=1}^Q \xrightarrow{d^*} \mathcal{N} \left(0; \text{diag} (v^1, \dots, v^Q) \right),$$

where $v^q = p^{-1}(x^q) Eu_t^2 \int_{\mathbb{R}} K^2(v) dv$, whereas

(b) if $d + d_\xi > 1/2$, then

$$\widehat{\mu}_{np}^*(x) - \widehat{\mu}_p^* = \frac{x\phi(x)}{p(x)} \left(\frac{1}{n} \sum_{t=1}^n \xi_t u_t^* - \frac{1}{n} \sum_{t=1}^n \xi_t \frac{1}{n} \sum_{t=1}^n u_t^* \right) (1 + o_p^*(1))$$

if $n^{1-2(d+d_\xi)} a^{-1} \rightarrow 0$, and if $n^{2(d+d_\xi)-1} a \rightarrow 0$, then (4.7) holds true.

Proof. The proof follows proceeding step by step as that of Proposition 3.7 and the arguments given for the proof of Proposition 4.2, and thus it is omitted. \square

Theorem 4.5. *Assuming that S1 – S3 hold, we have that*

(a) if $d + d_\xi < 1/2$ then under C3, we have that

$$(4.8) \quad Pr \left\{ v_n \left(\tilde{n}^{\frac{1}{2}} \widetilde{\mathcal{T}}_s^* - \zeta_n \right) \leq x | \mathcal{U}_n \right\} \xrightarrow{P} \exp(-2e^{-x}),$$

for $x > 0$ and ζ_n is as defined in Theorem 3.8,

whereas (b) if $d + d_\xi > 1/2$, then

$$\mathcal{T}_s^* = \sup_{x \in \mathcal{X}} \left| \frac{x\phi(x)}{p^{1/2}(x)} \right| \left(\frac{1}{n} \sum_{t=1}^n \xi_t u_t^* - \frac{1}{n} \sum_{t=1}^n \xi_t \frac{1}{n} \sum_{t=1}^n u_t^* \right) (1 + o_p^*(1))$$

if $n^{1-2(d+d_\xi)} a^{-1} \rightarrow 0$, and if $n^{2(d+d_\xi)-1} a \rightarrow 0$, then (4.8) holds true.

Proof. The proof follows by Proposition 4.4 and Woodroffe's (1967) Theorem 5, and thus it is omitted. \square

4.3. MODIFICATION AND MOTIVATION OF THE BOOTSTRAP.

We finish Section 4 giving now the intuition/motivation and also a modification of the bootstrap scheme described in Section 4.1. We begin with the intuition. Suppose that in C1, $\bar{\vartheta}_k = \mathcal{I}(k=0)$, that is $u_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}$. Then, using the identity

$$(4.9) \quad u_t = \frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} w_u(\lambda_j)$$

and Bartlett's approximation of $w_u(\lambda_j)$, that is $w_u(\lambda_j) \approx B(-\lambda_j) w_\varepsilon(\lambda_j)$, see Brockwell and Davis's (1991) Theorem 10.3.2, where “ \approx ” should be read as “approximately”, we obtain that u_t can be approximated by

$$\frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} B(-\lambda_j) w_\varepsilon(\lambda_j).$$

However, in this paper we allow for the possibility of strong dependence. For instance, that u_t follows the fractional difference model

$$u_t = (1 - L)^{-d} \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}; \quad b_0 = 1.$$

So, the previous arguments would suggest to employ the Bartlett's approximation

$$(4.10) \quad w_u(\lambda_j) \approx (1 - e^{-i\lambda_j})^{-d} B(-\lambda_j) w_\varepsilon(\lambda_j)$$

and thence the approximation

$$(4.11) \quad u_t \approx \frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} (1 - e^{-i\lambda_j})^{-d} B(-\lambda_j) w_\varepsilon(\lambda_j).$$

However, the lack of smoothness of $(1 - e^{-i\lambda_j})^{-d}$ around $\lambda_j = 0$ and results given in Robinson's (1995a) Theorem 1 at frequencies λ_j for fixed j indicate that for those frequencies the approximation in (4.10) seems to be invalid. Observe

that these frequencies are precisely the more relevant ones when examining the asymptotic behaviour of the least squares estimator $\hat{\theta}$ given in (2.3) or the sample mean, see Yajima (1991) and also the behaviour of $\hat{r}(q/n)$ in (2.4). For that reason, we have replaced $(1 - e^{-i\lambda_j})^{-d}$ by $\tilde{g}^{1/2}(-\lambda_j; d)$ in *STEP 4*. That is, we modified (4.11) by

$$(4.12) \quad u_t \approx \tilde{u}_t =: \frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} \tilde{g}^{1/2}(-\lambda_j; d) B(-\lambda_j) w_\varepsilon(\lambda_j).$$

It is very easy to show that the right side of (4.12) preserves (asymptotically) the covariance structure of $\{u_t\}_{t \in \mathbb{Z}}$. Indeed, after noting that $C1$ implies that $E(w_\varepsilon(\lambda_j) w_\varepsilon(-\lambda_k)) = \sigma_\varepsilon^2 \mathcal{I}(j = k)$ and recalling (2.8), we have that

$$E(\tilde{u}_t \tilde{u}_s) = \frac{\sigma_\varepsilon^2}{n} \sum_{j=1}^n e^{i(t-s)\lambda_j} \tilde{g}(\lambda_j; d) |B(\lambda_j)|^2 \approx \int_{-\pi}^{\pi} f(\lambda) e^{i(t-s)\lambda} d\lambda := E(u_t u_s),$$

because $\tilde{g}(\lambda_j; d) |B(\lambda_j)|^2 \approx |1 - e^{-i\lambda_j}|^{-2d} |B(\lambda_j)|^2 = 2\pi f(\lambda_j) / \sigma_\varepsilon^2$ and $\frac{2\pi}{n} \sum_{j=1}^n f(\lambda_j) e^{i\ell\lambda_j} \rightarrow \int_0^{2\pi} f(\lambda) e^{i\ell\lambda} d\lambda = E(u_0 u)$. Moreover the last displayed equality indicates that the covariance structure of u_t^* only depends on $|t - s|$.

Thus, what we did in *STEP 4* was to replace d and $B(-\lambda_j)$ respectively in the right side of (4.12) by consistent estimators, whereas *STEP 2* computes the bootstrap for the *discrete Fourier transform* $w_\varepsilon(\lambda_j)$, $j = 1, \dots, \tilde{n}$, which it could have been computed as in Hidalgo (2003) if instead of looking at bootstrapping of \mathcal{T} in (2.5) we would have been interested in bootstrapping of $\hat{\theta}$ or even \mathcal{T} but at a finite number of points.

Next, we shall describe a modification of the bootstrap given in *STEPS 1* to *7*. The alternative bootstrap is as follows.

STEPS 1' and 2': They are the same as *STEPS 1* and *3*, respectively.

STEP 3': Compute

$$\tilde{u}_t = \frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} \tilde{g}^{1/2}(-\lambda_j; -\hat{d}) w_{\hat{u}}(\lambda_j), \quad t = 1, \dots, n,$$

and fit an *AR*(p_n) model to $\{\tilde{u}_t\}_{t=1}^n$, where $p_n = o(n)$. Denote the residuals by

$$\hat{\varepsilon}_t = \tilde{u}_t - \sum_{q=1}^{p_n} \hat{\phi}_q \tilde{u}_{t-q} \quad t = p_n + 1, \dots, n,$$

being $\hat{\phi}_q$, $q = 1, \dots, p_n$, the *LSE* in $\tilde{u}_t = \sum_{q=1}^{p_n} \phi_q \tilde{u}_{t-q} + v_t$.

Let $\{\tilde{\varepsilon}_t^*\}_{t=1-n}^n$ be a random sample with replacement from the empirical distribution function of the standardized residuals

$$\tilde{\varepsilon}_t = \tilde{\sigma}_\varepsilon^{-1} \left(\hat{\varepsilon}_t - \frac{1}{n - p_n} \sum_{t=1}^{n-p_n} \hat{\varepsilon}_t \right), \quad \tilde{\sigma}_\varepsilon^2 = \frac{1}{n - p_n} \sum_{t=1}^{n-p_n} \left(\hat{\varepsilon}_t - \frac{1}{n - p_n} \sum_{t=1}^{n-p_n} \hat{\varepsilon}_t \right)^2.$$

STEP 4': Compute

$$\tilde{u}_t^* = \sum_{q=1}^{p_n} \hat{\phi}_q \tilde{u}_{t-q}^* + \tilde{\sigma}_\varepsilon \tilde{\varepsilon}_t^*; \quad t = 1 - n, \dots, n$$

with $\tilde{u}_t^* = 0$ for $t \leq -n$.

STEP 5': Compute the bootstrap observations $\{u_t^*\}_{t=1}^n$ as

$$u_t^* = \frac{1}{n^{1/2}} \sum_{j=1}^n e^{it\lambda_j} \tilde{g}^{1/2}(-\lambda_j; \hat{d}) w_{\tilde{u}^*}(\lambda_j), \quad t = 1, \dots, n.$$

STEPS 6' to 8': They are the same as *STEPS 5* to *7* respectively.

5. MONTE-CARLO EXPERIMENT

In this section we present a small Monte-Carlo experiment of the test given in (2.12) to shed some light on its finite sample performance. Throughout the experiment we have considered three sample sizes, $n = 64, 128$ and 256 . For each sample size, we have performed 1000 replications with $B = 499$ as the number of bootstrap samples. We have considered the regression model

$$y_t = E[y_t | x_t] + u_t, \quad t = 1, \dots, n,$$

where $\{u_t\}_{t=1}^n$ is generated as a *FARIMA* $(0, d, 0)$ Gaussian process with unit variance, with $d = 0.1, 0.2, 0.3$ and 0.4 , using an algorithm due to Davies and Harte (1987). For our regressor, we compute $\{x_t\}_{t=1}^n$ as a *FARIMA* $(0, d_\xi, 0)$ Gaussian process with unit variance, with $d_\xi = 0.0, 0.2$ and 0.3 , using the same algorithm due to Davies and Harte. In all of our experiments we have chosen the bandwidth parameters m and a deterministically. More specifically, we chose $m = n/4, n/8$ and $n/16$ and $a = 0.5n^{-1/4}$. The motivation for the choice of the bandwidth parameter a comes from the Monte Carlo experiment given in Hidalgo (1997).

To analyze the size accuracy of the test we have considered the above displayed regression model with $E[y_t | x_t] = 1$. The results are given in *TABLES 5.1 to 5.3*.

TABLES 5.1 to 5.3 ABOUT HERE

The first conclusion that we can draw from *TABLES 5.1 to 5.3* is that the choice of the bandwidth parameter m appears not to be very relevant or sensitive. Also, the empirical size of the test is very similar and very satisfactory even for sample sizes as small as $n = 64$.

Next, we examine the power performance of the test. To this end, we considered the following two linear regression models

$$\begin{aligned} y_t &= 1 + \beta x_t + u_t, \quad t = 1, \dots, n, \\ y_t &= 1 + \sin(x_t) + u_t, \quad t = 1, \dots, n, \end{aligned}$$

with $\beta = 0.5$ and $\beta = 1.0$. The reason to take two different values of β is to calibrate the power function as the alternative divert from the null. The results are given in *TABLES 5.4* and *5.5*.

TABLES 5.4 and 5.6 ABOUT HERE

The main conclusion is that we observe that as the “distance” between the null and alternative hypothesis increases, the power increases as *TABLES 5.4* and *5.5* illustrate. Also, this increase in power is uniform across the sample sizes. On the other hand, the above three tables also indicate that the power increases with the sample size, which is a desirable property too, as it will be an indication of the consistency of the test.

To conclude this section, we shall point out that although we have chosen the bandwidths deterministically and the results appear not to be very sensitive to that choice, however, in empirical examples it would be convenient to employ some automatic criterion to choose the bandwidth m . One approach that can be adopted is similar to that in Robinson (1991) and employed in Lobato and Robinson (1997). More specifically, we can choose m as

$$m(0) = \frac{1}{2} \left(\frac{3n}{4\pi} \right)^{3/4} \left| \frac{\xi''(0)}{2\xi(0)} \right|^{-3/8}$$

and $\xi(\lambda)$ is the spectral density function of an *AR*(1) model with parameter ρ , that is $\xi(\lambda) = (2\pi)^{-1} (1 + \rho^2 - 2\rho \cos(\lambda))^{-1}$, although more general $\xi(\lambda)$ functions can

be adopted, see Lobato and Robinson (1997) for a discussion. Alternatively, we could have chosen

$$m = \frac{1}{2} \left(\frac{3n}{4\pi} \right)^{3/4} \frac{1}{\tilde{n}} \sum_{j=0}^{\tilde{n}} \left| \frac{\xi''(\lambda_j)}{2\xi(\lambda_j)} \right|^{-3/8}$$

which is in a sense the average pointwise bandwidths

$$m(\lambda_j) = 2^{-1} (3n/4\pi)^{3/4} \left| \xi''(\lambda_j) / (2\xi(\lambda_j)) \right|^{3/8}.$$

However for simplicity and because the asymptotic distribution of $\hat{\theta}$ in (2.3) depends only on the behaviour of the spectral density function at zero frequency, we could opt for $m(0)$. In practice as ρ is not known, $m(0)$ would be replaced by

$$\begin{aligned} m^{**}(0) &= \frac{1}{2} \left(\frac{3n}{4\pi} \right)^{3/4} \left| \frac{\widehat{\xi}''(0)}{2\widehat{\xi}(0)} \right|^{-3/8} \\ &= \frac{1}{2} \left(\frac{3n}{4\pi} \right)^{3/4} \left| \frac{-\widehat{\rho}}{(1-\widehat{\rho})^2} \right|^{-3/8}, \end{aligned}$$

where $\widehat{\rho}$ is the least squares estimator of ρ in $u_t = \rho u_{t-1} + \text{error}$. Because in practice $m^{**}(0)$ might be smaller than 1 or greater than \tilde{n} , we truncate $m^{**}(0)$ as

$$m^* = \begin{cases} \underline{m} & \text{if } m^{**}(0) < \underline{m} \\ \left(\frac{3n}{4\pi} \right)^{3/4} \left| \frac{-\widehat{\rho}}{(1-\widehat{\rho})^2} \right|^{-3/8} & \underline{m} < m^{**}(0) < \overline{m} \\ \overline{m} & \text{if } \overline{m} < m^{**}(0), \end{cases}$$

where, say, $\underline{m} = 0.06n^{3/4}$ and $\overline{m} = 1.2n^{3/4}$. However, other procedures are possible as those described in Henry and Robinson (1996) or Hurvich (2001).

6. PROOFS OF THE MAIN RESULTS

6.1. Proof of Proposition 3.3.

We shall begin with the case $\mathbf{d} > \mathbf{0}$. Abbreviating $b(q_1, q_2)$ by b , Proposition 3.2 implies that

$$(6.1) \quad \rho(b; d) = h(0) \theta(d) \int_{-1}^1 \int_{b-1}^{1+b} |v-w|^{2d-1} K(v) K(w-b) dv dw + o(1).$$

After noting that for $\ell = |v-w| > 0$ and $d > 0$, $\ell^{2d-1} = \frac{2}{\pi} \Gamma(2d) \cos(d\pi) \int_0^\infty \lambda^{-2d} \cos(\ell\lambda) d\lambda$, we have that the first term on the right of (6.1) is

$$\begin{aligned} & \int_0^1 \int_{b-1}^{1+b} \left(\int_0^\infty \lambda^{-2d} \left(e^{i|v-w|\lambda} + e^{-i|v-w|\lambda} \right) d\lambda \right) K(v) K(w-b) dv dw \\ &= \frac{1}{2} \int_{-\infty}^\infty |\lambda|^{-2d} \int_{-1}^1 \int_{-1}^1 \left(\left(e^{i(v-w)\lambda} e^{-ib\lambda} + e^{-i(v-w)\lambda} e^{ib\lambda} \right) K(v) K(w) dv dw \right) d\lambda \\ &= \frac{1}{2} \int_{-\infty}^\infty |\lambda|^{-2d} \left| \int_{-1}^1 K(v) e^{iv\lambda} dv \right|^2 (e^{-ib\lambda} + e^{ib\lambda}) d\lambda, \end{aligned}$$

where in the first equality we have used the reflection formula of the gamma function together with the change of variables $w-b \rightarrow w$. Hence, using that $2 \cos x = e^{ix} + e^{-ix}$, we conclude that

$$\rho(b; d) - \rho(0; d) = \int_{-\infty}^\infty |\lambda|^{-2d} \{ \cos(|\lambda|b) - 1 \} \left| \int_{-1}^1 K(v) e^{i\lambda v} dv \right|^2 d\lambda.$$

Next, because $\left| \int_{-1}^1 K(v) e^{i\lambda v} dv \right|^2 > 0$ and different than a constant and $|\cos x| \leq 1$, we obtain that $\rho(b; d) - \rho(0; d) < 0$ for $|b| > 0$. Moreover, as $b \rightarrow 0$, $|\rho(b; d) - \rho(0; d)| = O(|b|^\alpha) + o(|b|^\alpha)$, for some $0 < \alpha \leq 2$, so that we can conclude that

$$(6.2) \quad \text{Corr}(b) = \frac{\rho(b; d)}{\rho(0; d)} = 1 - D|b|^\alpha + o(|b|^\alpha); \quad b \rightarrow 0$$

On the other hand, as $b \rightarrow \infty$, it is easily shown that $\text{Corr}(b) \log b = o(1)$ from the definition of $\text{Cov}(b)$ given in (6.1) because $d \in (0, 1/2)$. Note that the latter agrees with Robinson's (1997) Proposition 2, since for $|q_2 - q_1| > zn$, $z > 0$, we have that $b \geq za^{-1}$ and $a \rightarrow 0$ by C3. This concludes the proof for $d > 0$.

Next, when $\mathbf{d} = \mathbf{0}$, the proof is omitted as it follows by standard arguments using Bickel and Rosenblatt's (1973) Theorem B.1. \square

6.2. Proof of Proposition 3.4.

Proceeding similarly as with Robinson's (1997) Theorem 1 and Lemma 1, it suffices to show that

$$(6.3) \quad E(\hat{r}(q) - r(q)) = O(a^2 + n^{-1}).$$

Observe that by uniform integrability of u_t^2 and that Propositions 3.2 and 3.3 imply that $\text{Cov}(\hat{r}(q_1), \hat{r}(q_2)) \rightarrow 0$ when $|q_1 - q_2| \geq nz > 0$, we conclude that the covariance of the asymptotic distribution of the estimators is zero by Theorem A of Serfling (1980, p.14). On the other hand, under H_0 and by standard kernel manipulations, we obtain that (6.3) holds true. \square

6.3. Proof of Proposition 3.7.

By standard algebra, we have that $\hat{\mu}_{np}(x) - \mu$ is

$$\frac{1}{\check{n}} \sum_{t=1}^n u_t K_t(x) \left(\frac{1}{\hat{p}(x)} - \frac{1}{E\hat{p}(x)} \right) + \frac{1}{E\hat{p}(x)} \frac{1}{\check{n}} \sum_{t=1}^n u_t \tilde{K}_t(x) + \frac{1}{n} \sum_{t=1}^n u_t,$$

where $\tilde{K}_t(x) = K_t(x) - E(K_t(x))$. So, $\hat{\mu}_{np}(x) - \hat{\mu}_p$ is

$$\begin{aligned} & \frac{1}{E\hat{p}(x)} \left\{ \frac{1}{\check{n}} \sum_{t=1}^n u_t \tilde{K}_t(x) - \frac{1}{\check{n}} \sum_{t=1}^n u_t K_t(x) \frac{\sum_{t=1}^n \tilde{K}_t(x)}{\check{n}\hat{p}(x)} \right\} \\ &= -\frac{1}{n} \sum_{t=1}^n u_t \frac{\sum_{t=1}^n \tilde{K}_t(x)}{\check{n}\hat{p}(x)} + \frac{1}{E\hat{p}(x)} \frac{1}{\check{n}} \sum_{t=1}^n u_t \tilde{K}_t(x) \left(1 - \frac{1}{\hat{p}(x)\check{n}} \sum_{t=1}^n \tilde{K}_t(x) \right) \\ (6.4) \quad &= \frac{1}{p(x)} \left\{ \frac{1}{\check{n}} \sum_{t=1}^n u_t \tilde{K}_t(x) - \frac{1}{n} \sum_{t=1}^n u_t \frac{1}{\check{n}} \sum_{t=1}^n \tilde{K}_t(x) \right\} (1 + o_p(1)) \end{aligned}$$

because following Ho (1996), see also Hidalgo (1997), we have that

$$\frac{1}{\check{n}} \sum_{t=1}^n \tilde{K}_t(x) = O_p \left(\frac{1}{\check{n}^{1/2}} + \frac{1}{n^{1/2-d_\varepsilon}} \right) \quad \text{and} \quad E\hat{p}(x) = p(x) + O(a^2).$$

Next, using (3.5), we also have that

$$(6.5) \quad \frac{1}{\check{n}} \sum_{t=1}^n u_t \tilde{K}_t(x) = \sum_{j=1}^{\infty} \frac{b_{jn}(x)}{j!a} \frac{1}{n} \sum_{t=1}^n u_t H_j(\xi_t).$$

We consider part (a) first. Because by (3.4), we have that $\sum_{t=1}^n u_t H_j(\xi_t) = O_p(n^{1/2})$ and $\sum_{t=1}^n u_t = O_p(n^{d+1/2})$, proceeding as in Ho (1996), we have that

the right side of (6.4), so it is $\tilde{n}^{1/2} (\widehat{\mu}_{np}(x) - \widehat{\mu}_p)$, is governed by the behaviour of

$$\frac{1}{p(x)} \frac{1}{\tilde{n}^{1/2}} \sum_{t=1}^n u_t \tilde{K}_t(x).$$

So, we have that (3.4) implies that at two points, say x^1 and x^2 ,

$$\begin{aligned} & \tilde{n}E \left\{ (\widehat{\mu}_{np}(x^1) - \widehat{\mu}_p) (\widehat{\mu}_{np}(x^2) - \widehat{\mu}_p) \right\} \\ &= \frac{1}{p(x^1)p(x^2)} \sum_{j=1}^{\infty} \frac{b_{jn}(x^1)b_{jn}(x^2)}{j!a} \left\{ Eu_t^2 EH_j^2(\xi_t) \right. \\ & \quad \left. + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) E(u_t u_{t+k}) E^j(\xi_t \xi_{t+k}) \right\} (1 + o(1)) \\ &= \frac{Eu_t^2}{p(x^1)p(x^2)} \sum_{j=1}^{\infty} \frac{b_{jn}(x^1)b_{jn}(x^2)}{j!a} EH_j^2(\xi_t) + o(a), \end{aligned}$$

after using (3.12) and (3.13) in Ho (1996). But, (3.5) indicates that

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{b_{jn}(x^1)b_{jn}(x^2)}{j!a} EH_j^2(\xi_t) &= a^{-1} E(\tilde{K}_t(x^1) \tilde{K}_t(x^2)) \\ &= p(x^1) \int_{\mathbb{R}} K^2(x) dx \mathcal{I}(x^1 = x^2) + o(a) \end{aligned}$$

by standard kernel manipulations. So the left side of (3.6) holds if we show that

$$(6.6) \quad \frac{1}{\tilde{n}^{1/2}} \sum_{t=1}^n u_t \tilde{K}_t(x) \xrightarrow{d} \mathcal{N}\left(0, p(x) Eu_t^2 \int_{\mathbb{R}} K^2(v) dv\right),$$

say. The proof of (6.6) proceeds similarly to that of Robinson and Hidalgo's (1997) Theorem 1 and so it is only sketched. Denoting $K_s^* = \sum_{t=1}^n \tilde{K}_t(x) \vartheta_{t-s}$, with $\vartheta_j = 0$ if $j < 0$, we have that the left side of (6.6) is

$$\frac{1}{\tilde{n}^{1/2}} \sum_{s=-N}^n K_s^* \varepsilon_s + \frac{1}{\tilde{n}^{1/2}} \sum_{s=-\infty}^{-N-1} K_s^* \varepsilon_s$$

with $N = N_n$ yet to be chosen. Now, after standard algebra, the second moment of the second term of the last displayed expression is bounded by

$$CE\varepsilon_s^2 \frac{1}{\tilde{n}} \left(\sum_{t=1}^n (E\tilde{K}_t^2(x))^{1/2} \right)^2 \sum_{s=N}^{\infty} \vartheta_s^2 \leq Cn \sum_{s=N}^{\infty} \vartheta_s^2$$

which tends to 0 as $n \rightarrow \infty$ for suitable sequence N_n by Condition S1. So, to conclude the proof of part (a), it suffices to how that

$$\sum_{s=-N}^n K_s^{*2} E\varepsilon_s^2 \xrightarrow{P} p(x) Eu_t^2 \int_{\mathbb{R}} K^2(v) dv; \quad E \sum_{s=-N}^n K_s^{*2} E(\varepsilon_s^2 \mathcal{I}(|K_s^* \varepsilon_s| > v | \{x_t\}_{t=1}^n)) \rightarrow 0$$

for all $v > 0$. Recall that $Eu_t^2 = E\varepsilon_t^2 \sum_{j=0}^{\infty} \vartheta_j^2$. But, the proof of the last two displayed expressions follow the same steps as those of Robinson and Hidalgo's (1997) Proposition 2 and so it is omitted. This concludes the proof of part (a).

Next, we examine part (b). To that end, for simplicity we assume that the Hermite rank is 1. That is $b_{1n}(x) \neq 0$. So, using (6.5) and Proposition 1 of Ho

(1996), we have that

$$(6.7) \quad \frac{1}{\tilde{n}} \sum_{t=1}^n \tilde{K}_t(x) = \sum_{j=2}^{\infty} \frac{b_{jn}(x)}{j!a} \frac{1}{n} \sum_{t=1}^n H_j(\xi_t) + \phi(x) x \frac{1}{n} \sum_{t=1}^n \xi_t (1 + o_p(1))$$

$$\frac{1}{\tilde{n}} \sum_{t=1}^n u_t \tilde{K}_t(x) = \sum_{j=2}^{\infty} \frac{b_{jn}(x)}{j!a} \frac{1}{n} \sum_{t=1}^n u_t H_j(\xi_t) + \phi(x) x \frac{1}{n} \sum_{t=1}^n u_t \xi_t (1 + o_p(1)).$$

Now, by standard arguments, $\sum_{t=1}^n \xi_t = O_p(n^{d_\xi+1/2})$, and $\sum_{t=1}^n u_t \xi_t = O_p(n^{d_\xi+d})$, whereas as proceeding as in part (a), the first terms on the right of (6.7) are $O_p(\tilde{n}^{-1/2})$.

So, because $n^{-1} \sum_{t=1}^n u_t = O_p(n^{d-1/2})$, when $n^{1-2(d_\xi+d)} a^{-1} \rightarrow 0$, we have that

$$\hat{\mu}_{np}(x) - \hat{\mu}_p = x\phi(x) \left(\frac{1}{n} \sum_{t=1}^n u_t \xi_t - \frac{1}{n} \sum_{t=1}^n \xi_t \frac{1}{n} \sum_{t=1}^n u_t \right) (1 + o_p(1)),$$

whereas if $n^{2(d_\xi+d)-1} a \rightarrow 0$, then

$$\hat{\mu}_{np}(x) - \hat{\mu}_p = \frac{1}{p(x)} \sum_{j=2}^{\infty} \frac{b_{jn}(x)}{j!a} \frac{1}{n} \sum_{t=1}^n u_t H_j(\xi_t) (1 + o_p(1))$$

and thus we obtain (3.6), proceeding as with part (a). \square

In what follows, we abbreviate $\gamma(t; d)$ and $\hat{\gamma}(t; \hat{d})$ by $\gamma(t)$ and $\hat{\gamma}(t)$ respectively.

6.4. Proof of Proposition 4.2.

Because $\{\varepsilon_t^*\}_{t=1}^n$ is a zero mean *iid* sequence of $\mathcal{N}(0, 1)$, Brockwell and Davis's (1991) Proposition 10.3.3 implies that $E^*(\eta_j^* \eta_{-l}^*) = \mathcal{I}(j = l)$, and hence we obtain that $E^*(u_t^* u_{t+v}^*) - \delta_v = n^{-1} 2\pi e^{\hat{c}_0} \sum_{j=1}^n e^{iv\lambda_j} \left(\sum_{|\ell| < n} \hat{\gamma}(\ell) e^{-i\ell\lambda_j} \right) \left| \hat{B}(\lambda_j) \right|^2 - \delta_v$, and where $\delta_t = E(u_0 u_t)$. Now, because $2\pi e^{\hat{c}_0} - \sigma_\varepsilon^2 = O_p(n^{-1/2})$ by Hidalgo and Yajima (2002), in what follows without loss of generality we shall assume that $\sigma_\varepsilon^2 = 1$ and take $2\pi e^{\hat{c}_0} = 1$. So, it suffices to show that

$$(6.8) \quad \frac{1}{\tilde{n}^{2d}} \sum_{v=q+1}^{\tilde{n}+q} \left| \frac{1}{n} \sum_{j=1}^n e^{iv\lambda_j} \left(\sum_{|\ell| < n} \hat{\gamma}(\ell) e^{-i\ell\lambda_j} \right) \left| \hat{B}(\lambda_j) \right|^2 - \delta_v \right| = o_p(1).$$

Now using that $\sum_{j=1}^n e^{i\ell\lambda_j} = n\mathcal{I}(0, n, \dots)$, by definition of $\hat{B}(\lambda_j)$ in (4.3), $\left| \hat{B}(\lambda_j) \right|^2 = \sum_{k=-M}^M \hat{\beta}_k e^{-ik\lambda_j}$, where $\hat{\beta}_k = \sum_{\ell=0}^{M-|k|} \hat{b}_\ell \hat{b}_{\ell+|k|}$; $|k| \leq M$, we obtain that the expression inside the modulus in (6.8) is

$$(6.9) \quad \sum_{|\ell| < n; |v+\ell| \leq M} \hat{\gamma}(\ell) \hat{\beta}_{v+\ell} - \delta_v.$$

Similarly using that $|B(\lambda_j)|^2 = \sum_{p=-\infty}^{\infty} \beta_p e^{-ip\lambda_j}$, where $\beta_p = (2\pi)^{-1} \int_{-\pi}^{\pi} |B(\lambda)|^2 e^{ip\lambda} d\lambda$, when $\left| \hat{B}(\lambda_j) \right|^2$ replaced by $|B(\lambda_j)|^2$, it implies that (6.8) becomes

$$(6.10) \quad \frac{1}{\tilde{n}^{2d}} \left| \sum_{v=q+1}^{\tilde{n}+q} \sum_{|\ell| < n} \hat{\gamma}(\ell) \sum_{q=0}^{\infty} \beta_{v+\ell+qn} - \delta_v \right| = o_p(1).$$

We prove (6.8) in two steps. In step (a), we shall show that (6.10) holds, and in step (b), that $\tilde{n}^{-2d} \left| \sum_{v=q+1}^{\tilde{n}+q} \sum_{|\ell| < n} \hat{\gamma}(\ell) \sum_{q=0}^{\infty} \beta_{v+\ell+qn} - \sum_{|\ell| < n; |v+\ell| \leq M} \hat{\gamma}(\ell) \hat{\beta}_{v+\ell} \right|$ is also $o_p(1)$. We begin the proof with step (a). Because $C1'$ implies that $\beta_p =$

$o\left(|p|_+^{-2}\right)$ and by definition of $\delta_v = \sum_{\ell=-\infty}^{\infty} \gamma(\ell) \beta_{v+\ell}$, see Fuller's (1996) Corollary 3.5.1., the expression inside the absolute value of (6.10) is, as we now show,

$$(6.11) \quad \sum_{|\ell|<n} \widehat{\gamma}(\ell) \beta_{v+\ell} - \delta_v + \sum_{|\ell|<n} \widehat{\gamma}(\ell) \beta_{v+\ell+n} + o(n^{-1}) \sum_{|\ell|<n} |\widehat{\gamma}(\ell)| \\ = \sum_{|\ell|<n} (\widehat{\gamma}(\ell) - \gamma(\ell)) \beta_{v+\ell} - \sum_{|\ell|\geq n} \gamma(\ell) \beta_{v+\ell} + o_p\left(n^{2d-1} + \frac{\mathcal{I}(d>0)}{|v|_+^{2d-2}}\right)$$

uniformly in $|v| \leq n - \check{n}$. First, by Lemma 7.7 part (a), $\sup_{\ell \geq 1} \gamma^{-1}(\ell) |\widehat{\gamma}(\ell) - \gamma(\ell)| = O_p\left(|\widehat{d} - d|\right)$ for $d > 0$ and $\sup_{\ell \geq 1} \ell |\widehat{\gamma}(\ell)| = O_p\left(|\widehat{d} - d|\right)$ for $d = 0$ and that by (2.9), $|\gamma(\ell)| = O(\ell^{2d-1})$. Hence, the third term on the left of (6.11) is $o_p(n^{2d-1})$, whereas proceeding similarly the second term on the left of (6.11) is $o\left(|v|_+^{2d-2} \mathcal{I}(d > 0) + n^{-1}\right)$.

Next, the second term on the right of (6.11) is $o(n^{2d-2})$ uniformly in $|v| \leq n$ by $C1'$, whereas by Lemma 7.7 with the convention that $\sum_c^e = 0$ if $e < c$, the first term on the right of (6.11) is bounded in absolute value by

$$D \left| \widehat{d} - d \right| \left\{ \sum_{|\ell|<2^{-1}v} + \sum_{2^{-1}v<|\ell|<n} \right\} |\ell|^{2d-1} |\beta_{v+\ell}| = o_p\left(|v|_+^{2d-1} \log^{-1} n\right)$$

because $|\beta_j| = o(j^{-2})$ and Proposition 4.1. So (6.11) is $o_p\left(|v|_+^{2d-1} \log^{-1} n\right)$ and hence (6.10) holds true because $d < 1/2$. This concludes the proof of step **(a)**.

To conclude the proof, we need to show step **(b)**. To that end, and in view of the proof of (6.10), it suffices to show that

$$(6.12) \quad \frac{1}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \left\{ \sum_{\substack{|\ell|<n \\ |\ell+v|\leq M}} \widehat{\gamma}(\ell) (\widehat{\beta}_{\ell+v} - \beta_{\ell+v}) - \left\{ \sum_{\substack{0<\ell<n \\ M<|\ell+v|}} + \sum_{\substack{-n<\ell=0 \\ M<|\ell+v|}} \right\} \right\} \gamma(\ell) \beta_{\ell+v} = o_p(1).$$

If $\mathbf{d} = \mathbf{0}$, the second and third terms on the left of (6.12) are bounded in absolute value by $D \sum_{v=M}^n |\beta_v| = o(1)$ because $\beta_v = o(v^{-2})$, whereas for $\mathbf{d} > \mathbf{0}$, they are

$$\frac{D}{\check{n}^{2d}} \sum_{v=q+1}^{(\check{n}+q) \wedge \check{n}} \sum_{M-v<\ell}^n |\gamma(\ell) \beta_{\ell+v}| + \frac{1}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \sum_{\ell=1; M<|v-\ell|}^n |\gamma(\ell) \beta_{v-\ell}| \\ = o(1) + \frac{D}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \left\{ \sum_{\ell=1; M<|v-\ell|}^M + \sum_{\ell=M+1; M<|v-\ell|}^{|v|/2} + \sum_{\ell=1+\frac{1}{2}|v|; M<|v-\ell|}^{v+M} + \sum_{\ell=v+M}^n \right\} |\gamma(\ell) \beta_{\ell-v}| \\ = \frac{D}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} (M^{2d} + v^{2d}) |\beta_v| + \frac{D}{M \check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \left\{ v^{2d-1} + (v+M)^{2d-1} \right\} + o(1),$$

which is $o(1)$ because $|\gamma(k)| = O\left(|k|_+^{2d-1}\right)$ and $|\beta_k| = o\left(|k|^{-2}\right)$ and $M\check{n}^{-1} = o(1)$ by $C3$. So, to complete the proof it remains to show that the first term on the left of (6.12) is also $o_p(1)$. But by Lemma 7.7 part (a), that term is

$$\frac{1}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \sum_{|\ell|<n; |\ell+v|\leq M} \gamma(\ell) (\widehat{\beta}_{\ell+v} - \beta_{\ell+v}) (1 + o_p(1)).$$

Now by definition of $\widehat{\beta}_{\ell+v}$ and $\beta_{\ell+v}$, the last displayed expression is governed by (6.13)

$$\frac{1}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \sum_{|\ell|<n; |\ell+v|\leq M} \gamma(\ell) \left\{ \sum_{p=0}^{M-|\ell+v|} \left(\widehat{b}_p \widehat{b}_{p+|\ell+v|} - b_p b_{p+|\ell+v|} \right) - \sum_{p=M-|\ell+v|+1}^{\infty} b_p b_{p+|\ell+v|} \right\}.$$

Because $\sum_{p=M-|\ell+v|+1}^{\infty} |b_p b_{p+|\ell+v|}| \leq DM^{-1} (M - |\ell + v|)_+^{-2}$, it implies that the second term of (6.13) is bounded in absolute value by

$$\frac{D}{M\check{n}^{2d}} \sum_{v=1}^M \sum_{\ell=0}^{M-v} \frac{|\gamma(\ell)|}{(M-v-\ell)^2} + \frac{D}{M\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \sum_{\ell=\max(1, v-M)}^{M+v} \frac{|\gamma(\ell)|}{(\ell-v+M)^2} = o(1).$$

Finally, the first term of (6.13). Now because by $C1'$ and (2.9),

$$\sum_{v=q+1}^{\check{n}+q} \sum_{|\ell|<n; |\ell+v|\leq M} |\gamma(\ell)| = \sum_{v=q+1}^M \sum_{\ell=-M-v}^{M-v} |\gamma(\ell)| + \sum_{v=(M+1)\vee q}^{\check{n}+q} \sum_{\ell=v-M}^{v+M} |\gamma(\ell)| = O(M\check{n}^{2d}),$$

then that, uniformly in p , by Lemma 7.6 parts (a) and (b), we have that $\widehat{b}_p - b_p = \varkappa_p + o_p(M^{-1})$, we conclude that the first term of (6.13) is dominated by (6.14)

$$\frac{D}{\check{n}^{2d}} \sum_{v=q+1}^{\check{n}+q} \sum_{|\ell|<n; |\ell+v|\leq M} \sum_{p=0}^{M-|\ell+v|} \{ \varkappa_p \varkappa_{p+|\ell+v|} + b_p \varkappa_{p+|\ell+v|} + b_{p+|\ell+v|} \varkappa_p \} |\gamma(\ell)| + o_p(1).$$

Next, by Hidalgo and Yajima's (2002) Theorem 1,

$$\frac{1}{\check{n}^{1/2}} \sum_{\ell=1}^{\check{n}} \frac{\check{h}(\lambda_\ell; d) - \widetilde{h}(\lambda_\ell)}{h(\lambda_\ell)} \cos(r\lambda_\ell) \xrightarrow{d} \mathcal{N}(0, \cdot).$$

So, because u_t^2 is uniformly integrable, Serfling's (1980, p.14) Theorem A, together with Lemma 7.3 part (a), implies that $E|\widehat{c}_r - \widetilde{c}_{r,n}| = O(n^{-1/2})$. From here and noting the definition of \varkappa_p and $\sum_p (|b_p| + |b_p^*|) < \infty$, it is easy to conclude that (6.14) is $o_p(1)$ by Markov's inequality. This completes the proof of step (b) and the proposition. \square

6.5. Proof of Theorem 4.3.

Because the asymptotic independence of the distributions of \max_q and \min_q , and the asymptotic distributions of $\sup_i X_i$ and $\inf_i -X_i$ are the same, it suffices to show that, for $x > 0$,

$$(6.15) \quad \Pr \left\{ v_n \left(\sup_{\check{n} \leq q \leq n - \check{n}} \check{n}^{\frac{1}{2}-\widehat{d}} \rho^{-\frac{1}{2}} \left(0; \widehat{d} \right) (\widehat{r}^*(q) - \widehat{r}(q)) - \zeta_n \right) \leq x | \mathcal{U}_n \right\} \xrightarrow{P} \exp(e^{-x}).$$

To that end, we will show that $\check{n}^{\frac{1}{2}-\widehat{d}} \rho^{-\frac{1}{2}} \left(0; \widehat{d} \right) (\widehat{r}^*(q) - \widehat{r}(q))$ converges, in bootstrap sense, to the Gaussian process $\mathcal{G}(q)$ in $\mathbb{D}[0, \infty)$, whose correlation structure is that given in (6.2) which satisfies conditions (v) and (vi) of Bickel and Rosenblatt's (1973) Theorem A1, for some $\alpha > 0$. See also Pickands's (1969) equations (1.2) and (2.1). From here the limiting distribution in (6.15) holds by Bickel and Rosenblatt's (1973) Theorem 1, for some $\alpha > 0$.

Because $E^* u_t^* = 0$, $\check{n}^{\frac{1}{2}-\widehat{d}} \rho^{-\frac{1}{2}} \left(0; \widehat{d} \right) E^* (\widehat{r}^*(q) - \widehat{r}(q)) \rightarrow_P 0$ by Proposition 3.2 and that by Proposition 4.1 and C3, $n^{\widehat{d}-d} - 1 = o_p(1)$ so that $\rho \left(0; \widehat{d} \right) - \rho(0; d) = o_p(1)$. Next, Lemmas 7.8 and 7.9 imply that the finite dimensional distributions converges in probability to those of $\mathcal{G}(q)$. Notice that Lemma 7.8 is the Lindeberg's condition. Next, by Proposition 3.3, the correlation structure of $\mathcal{G}(u)$ satisfies (6.2)

when $b \rightarrow 0$, whereas $\text{Corr}(b) = o(\log^{-1} b)$ as $b \rightarrow \infty$. So, the function $\text{Corr}(b)$ in (6.2) satisfies the conditions in Bickel and Rosenblatt (1973) or Pickands (1969).

To complete the proof of the theorem, it remains to verify the tightness condition. To that end, denote

$$(6.16) \quad X_n^*(\tilde{q}) = \frac{1}{\tilde{n}^{\frac{1}{2}+d}} \sum_{t=1}^n u_t K\left(\frac{t}{\tilde{n}} - \tilde{q}\right), \quad \tilde{q} = \frac{1}{\tilde{n}}, \frac{2}{\tilde{n}}, \dots, [a]^{-1}.$$

So, $X_n^*(\tilde{q})$, say, is a process in $\mathbb{D}\left[0, [a]^{-1}\right]$ equipped with Skorohod's metric, where we extend $\mathbb{D}\left[0, [a]^{-1}\right]$ to $\mathbb{D}[0, \infty)$ by writing $X_n^*(\infty) = X_n^*\left([a]^{-1}\right)$. By Pollard (1984, *Ch.V*), we need to show tightness in $\mathbb{D}[0, D]$ for any finite $D > 0$. On the other hand, we observe that Lemma 7.8 implies that the process $X_n(\tilde{q})$ has independent and stationary increments, that is for $\tilde{q} \in [c_1, d_1]$ and $\tilde{q} \in [c_2, d_2]$ and $[c_1, d_1] \cap [c_2, d_2] = \emptyset$, $X_n(\tilde{q})$ is (asymptotically) independent with the same finite dimensional distributions.

Because $\mathcal{G}(\bullet)$ has continuous paths, by Billingsley's (1968) Theorem 16.6, it suffices to show the Kolmogorov's moment condition

$$E\left(|X_n^*(\tilde{q}_2) - X_n^*(\tilde{q})|^\beta |X_n^*(\tilde{q}) - X_n^*(\tilde{q}_1)|^\beta\right) \leq H_n(\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}|^{\frac{1+\beta}{2}} |\tilde{q} - \tilde{q}_1|^{\frac{1+\beta}{2}}$$

for some $\delta > 0$, $\beta > 0$ and where $0 \leq \tilde{q}_1 < \tilde{q} < \tilde{q}_2 \leq D$ and $H_n(\tilde{q}_2, \tilde{q}_1) = O_p(1)$. As usual, observe that we can consider only the situation for which $\tilde{n}^{-1} < \tilde{q}_2 - \tilde{q}_1$, since otherwise the left side is trivially zero. Because for any $0 \leq a < b < c \leq D$, $|c - b||b - a| \leq |c - a|^2$ by Cauchy-Schwarz inequality, the last displayed inequality holds true if

$$(6.17) \quad E |X_n^*(\tilde{q}_2) - X_n^*(\tilde{q}_1)|^2 \leq H_n(\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^2.$$

It suffices to consider $|\tilde{q}_2 - \tilde{q}_1| < 1$, the case $|\tilde{q}_2 - \tilde{q}_1| \geq 1$ is trivial since the left side of (6.17) is bounded.

By definition, $X_n^*(\tilde{q}_2) - X_n^*(\tilde{q}_1)$ is

$$(6.18) \quad \frac{1}{\tilde{n}^{\frac{1}{2}+d}} \left\{ \sum_{t=\tilde{n}-(q_2-q_1)+1}^{\tilde{n}} u_{t+q_2}^* K_t + \sum_{t=1}^{\tilde{n}-(q_2-q_1)} u_{t+q_2}^* (K_t - K_{t+q_1-q_2}) - \sum_{t=1}^{q_2-q_1} u_{t+q_1}^* K_t \right\}.$$

Because $\tilde{n}\tilde{q} = q$, the contribution into the left of (6.17) due to the third term of (6.18) is bounded by $H_n(\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{3+2d}$ because by C2, $|K_t| \leq Dt/\tilde{n}$ and Proposition 4.2 implies that $\sum_{t,s=1}^q |E^*(u_{t+q_1}^* u_{s+q_1}^*)| = D \sum_{t,s=1}^q |E(u_t u_s)| (1 + o_p(1)) = q^{1+2d} (1 + o_p(1))$.

Observe that $H_n(\tilde{q}_2, \tilde{q}_1) = D\tilde{n}^{2(d-\hat{d})} = O_p(1)$ because by Proposition 4.1 and C3, $|\hat{d} - d| = o_p(\log^{-1} n)$. So, the last term of (6.18) satisfies the inequality (6.17). Similarly, we obtain that the contribution due to the first term of (6.18) is $D\tilde{n}^{2(d-\hat{d})} |\tilde{q}_2 - \tilde{q}_1|^2 (1 - (1 - (\tilde{q}_2 - \tilde{q}_1)))^{2d} \leq H_n(\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{2+2d}$ because $0 < \tilde{q}_2 - \tilde{q}_1 < 1$. Finally, C2 part (b) implies that the contribution due to the second term of (6.18) into the left side of (6.17) is because $d < 1/2$ bounded by $D(\tilde{q}_2 - \tilde{q}_1)^2 \tilde{n}^{-1-2\hat{d}} \sum_{t,s=1}^{\tilde{n}-(q_2-q_1)} |t-s|^{2d-1} \leq H_n(\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^2$. So, (6.17) holds true and hence it concludes the proof of the theorem because by Lemma 7.8, the correlation structure converges in probability to that of Proposition 3.3. \square

7. AUXILIARY LEMMAS

In what follows $\varphi(\lambda_j)$ will be abbreviated as φ_j for a generic $\varphi(\lambda)$ function. Denote $\check{h}_\ell(d) = (2m+1)^{-1} \sum_{j=-m}^m \left| 2 \sin\left(\frac{\lambda_{\ell+j}}{2}\right) \right|^{2d} I_{uu, \ell+j}$, where $\phi_j = \log |2 \sin(\lambda_j/2)|$. We have that Taylor's expansion up to the $2nd$ term implies that

$$(7.1) \quad \check{h}_\ell(\widehat{d}) - \check{h}_\ell(d) = \frac{2(\widehat{d}-d)}{2m+1} \sum_{j=-m}^m \phi_{\ell+j}^{2d} \log \phi_{\ell+j} I_{uu, \ell+j} + D \left| d - \widehat{d} \right|^2 \frac{\log^2 n}{m} \sum_{j=-m}^m \phi_{\ell+j}^{2\widehat{d}} I_{uu, \ell+j},$$

where, herewith, \widehat{d} denotes an intermediate point between d and \widehat{d} . Also denote $q_\ell(p) = (2m+1)^{-1} \sum_{j=-m}^m \frac{h_{\ell+j}}{h_\ell} \left(I_{\varepsilon\varepsilon, \ell+j} - \frac{\sigma_\varepsilon^2}{2\pi} \right) \log^p \phi_{\ell+j}$.

Lemma 7.1. *Assuming C1', C2 and C3, uniformly in $r \leq M$,*

$$(7.2) \quad \frac{r}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \frac{\check{h}_\ell(\widehat{d}) - \check{h}_\ell(d)}{h_\ell} \cos(r\lambda_\ell) = \frac{2r(\widehat{d}-d)}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} q_\ell(1) \cos(r\lambda_\ell) + o_p\left(\left|\widehat{d}-d\right| \log r\right).$$

Proof. First, proceeding as with the proof of Robinson's (1995b) expression (4.8), we have that

$$(7.3) \quad E \left| \frac{1}{2m+1} \sum_{j=-m}^m \frac{\log \phi_{\ell+j}}{h_\ell} \left(\phi_{\ell+j}^{2d} I_{uu, \ell+j} - \frac{\sigma_\varepsilon^2}{2\pi} h_{\ell+j} \right) - q_\ell(1) \right| = O\left(\frac{\log n}{m^{2/3}}\right).$$

On the other hand, by standard arguments, $(2m+1)^{-1} \sum_{j=-m}^m \log \phi_{\ell+j} h_{\ell+j} = h_\ell \log \phi_\ell + O(\ell^{-2} \log \ell)$. So by Markov's inequality, we conclude that the contribution of the first term on the right of (7.1) into the left of (7.2) is

$$2(d-\widehat{d}) \left\{ \frac{r}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} q_\ell(1) \cos(r\lambda_\ell) + \frac{\sigma_\varepsilon^2 r}{2\pi \tilde{n}} \sum_{\ell=1}^{\tilde{n}} \cos(r\lambda_\ell) \log \phi_\ell \right\} + O_p\left(\frac{r \left| d - \widehat{d} \right| \log n}{m^{2/3}}\right).$$

However because by Yong's (1974) Theorem III-23,

$$\frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \cos(r\lambda_\ell) \log^p \phi_\ell = O\left(\frac{\log r}{r}\right),$$

we conclude that the contribution of the first term on the right of (7.1) into the left of (7.2) is its right side. Next, proceeding as above, by Proposition 3.2 and C1, the contribution of the second term on the right of (7.1) into the left of (7.2) is $o_p\left(\left|\widehat{d}-d\right|^2\right)$. \square

Lemma 7.2. *Assuming C1' and C2, uniformly in $\ell \leq \tilde{n}$,*

$$(7.4) \quad \widehat{h}_\ell - \check{h}_\ell(\widehat{d}) = O_p(1) v_n,$$

where v_n satisfies that $E|v_n|^q = O\left(\left(n^{2d} \log^2 n / (m(\ell+m))\right)^q\right)$, for $q = 1, 2$.

Proof. Writing $\varphi_j(p) = \phi_j^{2d} \log^p \phi_j$, by Taylor's expansion, the left side of (7.4) is

$$(7.5) \quad \left\{ \sum_{p=0}^{\beta} \frac{(\widehat{d}-d)^p}{2m+1} \sum_{j=-m}^m \varphi_{\ell+j}(p) + K \left| \widehat{d}-d \right|^{\beta+1} \frac{\log^\beta n}{m} \sum_{j=-m}^m \phi_{\ell+j}^{2\widehat{d}} \right\} (I_{\widehat{u}\widehat{u}, \ell+j} - I_{uu, \ell+j}).$$

Now, by definition of \widehat{u}_t , we have that a typical element of $I_{\widehat{u}\widehat{u},j} - I_{uu,j}$ is

$$\frac{(\widehat{\theta}_k - \theta_k)}{2\pi n} \sum_{t,s=1}^n \left\{ u_t \left(\frac{s}{n}\right)^k + u_s \left(\frac{t}{n}\right)^k \right\} e^{i(t-s)\lambda_j} + \left| \frac{(\widehat{\theta}_k - \theta_k)}{(2\pi n)^{1/2}} \sum_{t=1}^n \left(\frac{t}{n}\right)^k e^{it\lambda_j} \right|^2.$$

So, we have that a typical term of the first term of (7.5) is $(\widehat{d} - d)^P / (2\pi)$ times

$$(7.6) \quad \frac{(\widehat{\theta}_k - \theta_k)^2}{mn} \left(\sum_{j=-m}^m \varphi_{\ell+j}(p) \sum_{t,s=1}^n \left(\frac{t}{n}\right)^k \left(\frac{s}{n}\right)^k e^{i(t-s)\lambda_{\ell+j}} \right)$$

$$(7.7) \quad + \frac{(\widehat{\theta}_k - \theta_k)}{mn} \left(\sum_{j=-m}^m \varphi_{\ell+j}(p) \sum_{t,s=1}^n \left(\frac{t}{n}\right)^k u_s e^{i(t-s)\lambda_{\ell+j}} \right).$$

Next, because by standard algebra, $\sum_{t=1}^n (t/n)^k e^{it\lambda_j} = 1 + (e^{i\lambda_j} - 1)^{-1}$, and then Taylor's expansion, $|e^{i\lambda_j} - 1|^{-1} \leq Kn/j$, the second moment of (7.6) is $O\left((n^{2d} \log^2 n / (m(\ell+m)))^2\right)$ by Proposition 3.1. Proceeding similarly and noting that $E \left| n^{-1/2} \sum_{s=1}^n u_s e^{is\lambda_{\ell+j}} \right|^2 = O\left(\lambda_{\ell+j}^{-2d}\right)$ by Giraitis, Hidalgo and Robinson's (2002) Lemma 4.4, the second moment of (7.7) is also $O\left(n^{2d} \log^2 n / (m(\ell+m))\right)$. Next proceeding as above, the second term of (7.5) is also $O_p(1) v_n$ by Proposition 3.3 and choosing β large enough. \square

Let $\widetilde{h}_\ell = (2m+1)^{-1} \sum_{j=-m}^m h_{\ell+j}$ and define

$$c_{r,n} = \frac{1}{\widetilde{n}} \sum_{\ell=1}^{\widetilde{n}} \log(h_\ell) \cos(r\lambda_\ell); \quad \widetilde{c}_{r,n} = \frac{1}{\widetilde{n}} \sum_{\ell=1}^{\widetilde{n}} \log(\widetilde{h}_\ell) \cos(r\lambda_\ell).$$

Lemma 7.3. *Let v_n be such that $E|v_n| = O(1)$. Assuming $C1'$, $C2$ and $C4$, uniformly in $r \leq M$*

$$(7.8) \quad (a) \quad \widehat{c}_r - \widetilde{c}_{r,n} = \frac{1}{\widetilde{n}} \sum_{\ell=1}^{\widetilde{n}} \frac{\check{h}_\ell(d) - \widetilde{h}_\ell}{h_\ell} \cos(r\lambda_\ell) + O_p\left(n^{-1/2}\right) v_n$$

$$(b) \quad \widetilde{c}_{r,n} - c_{r,n} = O(M^{-2}); \quad (c) \quad c_{r,n} - c_r = O(n^{-1}).$$

Proof. Part (b) follows by Taylor's expansion and because standard arguments imply that $\widetilde{h}_\ell - h_\ell = O(M^{-2})$, whereas part (c) follows immediately by Brillinger (1981, p.15). Next, part (a). Using the definitions of \widehat{c}_r and $\widetilde{c}_{r,n}$, we have that

$$(7.9) \quad \widehat{c}_r - \widetilde{c}_{r,n} = \frac{1}{\widetilde{n}} \sum_{\ell=1}^{\widetilde{n}} \left(\log \widehat{h}_\ell - \log \widetilde{h}_\ell \right) \cos(r\lambda_\ell).$$

Proceeding as in the proof of Theorem 1 of Hidalgo and Yajima (2002), the right side of (7.9) is

$$\begin{aligned}
& \frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \frac{\check{h}_\ell(d) - \tilde{h}_\ell}{\tilde{h}_\ell} \cos(r\lambda_\ell) + \frac{D}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \left(\frac{\check{h}_\ell(d) - \tilde{h}_\ell}{\tilde{h}_\ell} \right)^2 (1 + o_p(1)) \\
(7.10) \quad & - \frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \frac{\check{h}_\ell(\hat{d}) - \check{h}_\ell(d)}{\tilde{h}_\ell} \cos(r\lambda_\ell) + \frac{D}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \left| \frac{\check{h}_\ell(\hat{d}) - \check{h}_\ell(d)}{\tilde{h}_\ell} \right|^2 (1 + o_p(1)) \\
& - \frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \frac{\hat{h}_\ell - \check{h}_\ell(\hat{d})}{\tilde{h}_\ell} \cos(r\lambda_\ell) + \frac{D}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \left| \frac{\hat{h}_\ell - \check{h}_\ell(\hat{d})}{\tilde{h}_\ell} \right|^2 (1 + o_p(1)).
\end{aligned}$$

The first moment of the second term of (7.10) is $O_p(m^{-1}) = o(n^{-1/2})$ by an obvious extension of Hidalgo and Robinson's (2002) Proposition 2.1, whereas the last two terms of (7.10) are, by Lemma 7.2, $O(n^{-1/2})$.

Finally, to conclude the proof it suffices to show that the third and fourth terms of (7.10) are $O_p(n^{-1/2})v_n$. From the proof of Lemma 7.1 and that standard arguments imply that $\tilde{h}_\ell = h_\ell + O(M^{-2})$, it suffices to examine the behaviour of

$$(7.11) \quad \frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} q_\ell(1) \cos(r\lambda_\ell)$$

by C5. Recall (7.3). Proceeding as with the proof of Hidalgo and Yajima's (2002) Theorem 1, we have that $E \left| \sum_{\ell=1}^{\tilde{n}} q_\ell(1) \right|^2 = O(n)$. So, we conclude by standard arguments that (7.11) is $O_p(n^{-1/2})v_n$. From here, it is obvious to observe that the fourth term of (7.10) is also $O_p(n^{-1/2})v_n$, because by standard results, see Brillinger (1981), $E |q_\ell(1)|^2 = O(m^{-1})$.

Thus, we have shown that the right of (7.9) is, uniformly in $r \leq M$,

$$\frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \frac{\check{h}_\ell(d) - \tilde{h}_\ell}{\tilde{h}_\ell} \cos(r\lambda_\ell) + O_p(n^{-1/2})v_n,$$

where $E|v_n| = O(1)$. But, the first term of the last displayed expression is

$$\frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \frac{\check{h}_\ell(d) - \tilde{h}_\ell}{h_\ell} \cos(r\lambda_\ell) + \frac{1}{\tilde{n}} \sum_{\ell=1}^{\tilde{n}} \left(\frac{\check{h}_\ell(d) - \tilde{h}_\ell}{\tilde{h}_\ell} \right) \left(1 - \frac{\tilde{h}_\ell}{h_\ell} \right) \cos(r\lambda_\ell)$$

whose second term has a first absolute moment bounded by $\tilde{n}^{-1} \sum_{\ell=1}^{\tilde{n}} \left| \frac{h_\ell - \tilde{h}_\ell}{h_\ell} \right| E \left| \frac{\check{h}_\ell(d) - \tilde{h}_\ell}{\tilde{h}_\ell} \right| = o(n^{-1/2})$ by Hidalgo and Yajima's (2002) Lemma 2 and the proof of their Lemma 1.

Now use that $M^{-2} |\hat{d} - d| = O_p(n^{-1/2})$ by C4 and Proposition 4.1 to conclude. \square

Lemma 7.4. *Assuming C1', C2 and C4, $E \left| \sum_{\ell=1}^{\tilde{n}} \frac{\check{h}_\ell(d) - \tilde{h}_\ell}{h_\ell} \cos(r\lambda_\ell) \right|^2 = O(n)$.*

Proof. It is a standard extension of Hidalgo and Yajima's (2002) Theorem 1. \square

Let us define

$$\tilde{A}_{\ell,n} = \exp \left\{ \sum_{r=1}^{M-1} \tilde{c}_{r,n} e^{-ir\lambda_\ell} \right\}; \quad A_{\ell,n} = \exp \left\{ \sum_{r=1}^{M-1} c_{r,n} e^{-ir\lambda_\ell} \right\}; \quad A_\ell^* = \exp \left\{ \sum_{r=1}^{M-1} c_r e^{-ir\lambda_\ell} \right\}.$$

Lemma 7.5. *Let v_n be such that $E|v_n| = O(n^{1/2}/m)$. Assuming C1', C2 and C3, uniformly in ℓ ,*

$$(a) \quad \hat{A}_\ell - \tilde{A}_{\ell,n} = O_p(1)v_n, \quad (b) \quad \tilde{A}_{\ell,n} - A_{\ell,n} = O(M^{-2}), \quad A_{\ell,n} - A_\ell^* = O(m^{-1}).$$

Proof. We begin by examining (b). Let $\tilde{a}_{\ell,n} = \log(\tilde{A}_{\ell,n})$ and $a_{\ell,n} = \log(A_{\ell,n})$. First, Taylor expansion of $\log(\tilde{h}_\ell/h_\ell)$ implies that $\tilde{a}_{\ell,n} - a_{\ell,n}$ is

$$\begin{aligned} & \sum_{r=1}^{M-1} \frac{1}{\tilde{n}} \sum_{p=1}^{\tilde{n}} \left\{ \left(\frac{\tilde{h}_p - h_p}{h_p} \right) + \left(\frac{\tilde{h}_p - h_p}{h_p} \right)^2 (1 + o(1)) \right\} \cos(r\lambda_p) e^{-ir\lambda_\ell} \\ &= \sum_{r=1}^{M-1} \frac{1}{\tilde{n}} \sum_{p=1}^{\tilde{n}} \left(\frac{\tilde{h}_p - h_p}{h_p} \right) \cos(r\lambda_p) e^{-ir\lambda_\ell} + O\left(\frac{1}{M^3}\right) \\ &= \frac{1}{2\tilde{n}} \sum_{p=1}^{\tilde{n}} \left(\frac{\tilde{h}_p - h_p}{h_p} \right) \sum_{r=1}^{M-1} (e^{ir\lambda_{p-\ell}} + e^{ir\lambda_{p+\ell}}) + O\left(\frac{1}{M^3}\right) = O\left(\frac{1}{M^2}\right), \end{aligned}$$

where in the first equality we have used that $h_p^{-1}\tilde{h}_p - 1 = O(M^{-2})$ by standard arguments and in the last equality, the proofs of Lemmas 1 and 4 of Hidalgo and Yajima (2002). To conclude the proof of part (b), use that $|A_{\ell,n}| \leq D$ and Taylor expansion of the exponential function. Proceeding similarly and that $c_{r,n} - c_r = O(n^{-1})$ by Lemma 7.3 part (c), the second part of (b) follows easily.

Next, we examine part (a). Defining $\hat{a}_\ell = \log(\hat{A}_\ell)$ and that $4Mm = n$, we have

$$(7.12) \quad \begin{aligned} \hat{a}_\ell - \tilde{a}_{\ell,n} &= \sum_{r=1}^{M-1} (\hat{c}_r - \tilde{c}_{r,n}) e^{-ir\lambda_\ell} \\ &= \sum_{r=1}^{M-1} \frac{1}{\tilde{n}} \sum_{p=1}^{\tilde{n}} \frac{\tilde{h}_p(d) - \tilde{h}_p}{h_p} \cos(r\lambda_p) e^{-ir\lambda_\ell} + O_p\left(\frac{n^{1/2}}{m}\right), \end{aligned}$$

by Lemma 7.3, and where proceeding as in the proof of Theorem 2 of Hidalgo and Yajima (2002), and in particular the proof of the first term on the right of their equation (46), the first absolute moment of the first term is $O(n^{1/2}/m)$.

But $\hat{A}_\ell - A_{\ell,n} = (\exp(\hat{a}_\ell - a_{\ell,n}) - 1) A_{\ell,n}$ and $|A_{\ell,n} - A_\ell| = o(1)$. Then, a simple application of delta methods implies that $\hat{A}_\ell - A_{\ell,n} = O_p(1) v_n$ by C5 and where $E|v_n| = O(n^{1/2}/m)$. This completes the proof of the lemma. \square

Denote by b_p^* the p th Fourier coefficient of $A(\lambda)$, and define $\tilde{b}_{v,n} = n^{-1} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} \tilde{A}_{\ell,n} e^{iv\lambda_\ell}$ and $b_{v,n} = \frac{1}{2M} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} A_{\ell,n} e^{iv\lambda_\ell}$.

Lemma 7.6. *Assuming C1', C2 and C4, as $n \rightarrow \infty$,*

$$(7.13) \quad (a) \tilde{b}_v - \tilde{b}_{v,n} = \varkappa_v + o_p(M^{-1}), \quad (b) \tilde{b}_{v,n} - b_v = O(M^{-2}).$$

uniformly in v , where $\varkappa_v = \sum_{r=1}^{M-1} (\hat{c}_r - \tilde{c}_{r,n}) b_{v-r}^$.*

Proof. We begin with (a). From the definition of $\hat{a}_\ell - \tilde{a}_{\ell,n}$ and Taylor expansion of $\hat{A}_\ell - \tilde{A}_{\ell,n}$, the left of (7.13) is

$$(7.14) \quad \left(\frac{1}{\tilde{n}} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} (\hat{a}_\ell - \tilde{a}_{\ell,n}) \tilde{A}_{\ell,n} e^{iv\lambda_\ell} \right) + \frac{1}{2\tilde{n}} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} |\hat{a}_\ell - \tilde{a}_{\ell,n}|^2 |\tilde{A}_{\ell,n}| (1 + o_p(1))$$

By the proof of Lemma 7.5 and standard inequalities, we have that $|\hat{a}_\ell - \tilde{a}_{\ell,n}|^2 = O_p(n/m^2)$, so that the second term of (7.14) is $o_p(M^{-1})$ by C4 because $|\tilde{A}_{\ell,n}| \leq D$.

Next, we examine the first term of (7.14). From the definition of $\hat{a}_\ell - \tilde{a}_{\ell,n}$ in (7.12), the first term of (7.14), after adding and subtracting $A_{\ell,n}$, is

$$\begin{aligned} & \sum_{r=1}^{M-1} (\hat{c}_r - \tilde{c}_{r,n}) \left\{ \frac{1}{n} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} A_{\ell,n} e^{i(v-r)\lambda_\ell} + \frac{1}{n} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} (\tilde{A}_{\ell,n} - A_{\ell,n}) e^{i(v-r)\lambda_\ell} \right\} \\ &= \sum_{r=1}^{M-1} (\hat{c}_r - \tilde{c}_{r,n}) b_{|v-r|}^* + O\left(\frac{\log M}{M^2}\right) \sum_{r=1}^{M-1} |\hat{c}_r - \tilde{c}_{r,n}|, \end{aligned}$$

where in the last equality we have used Lemma 6.5 part (b), that the ℓ th Fourier coefficient of $A(\lambda)$ is b_ℓ^* , and Brillinger (1981, p.15) and then C2 imply that

$$(7.15) \quad \frac{1}{n} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} A_\ell e^{i(v-r)\lambda_\ell} - \int_{-\pi}^{\pi} A(\lambda) e^{i(v-r)\lambda} d\lambda = O(n^{-1}),$$

that $|A_\ell^* - A_\ell| = O(M^{-2})$ because $c_r = O(r^{-3})$ and Lemma 6.5 part (c) imply that $\left| n^{-1} \sum_{\ell=1-\tilde{n}}^{\tilde{n}} A_{\ell,n} e^{i(v-r)\lambda_\ell} \right| = b_{|v-r|}^* + O(M^{-2})$.

Now, observing that the first absolute moment of the first term on the right of (7.8) is $O(n^{-1/2})$ by Lemma 6.4, we conclude that the first term of (7.14) is

$$\sum_{r=1}^{M-1} (\hat{c}_r - \tilde{c}_{r,n}) b_{|v-r|}^* + O_p(M^{-2}),$$

because Lemmas 6.3 and 6.4 imply that $\sum_{r=1}^{M-1} |\hat{c}_r - \tilde{c}_{r,n}| = o_p(1)$. Note that the orders of magnitude in Lemma 6.3 are uniformly in r .

Next, we examine part (b). We first note that $\tilde{a}_{v,n} - a_v = (\tilde{a}_{v,n} - a_{v,n}) + (a_{v,n} - a_v)$ whose second term on the right is $O(n^{-1})$ by Brillinger (1981, p.15). Now, proceeding as with part (a) and Taylor's expansion of $\tilde{A}_{\ell,n}$ and $A_{\ell,n}$, it is easily obtained that

$$\tilde{b}_{v,n} - b_v = \sum_{r=1}^{M-1} (\tilde{c}_{r,n} - c_{r,n}) b_{|v-r|}^* + O\left(n^{-1} + M^{-2} \sum_{r=1}^{M-1} |\tilde{c}_{r,n} - c_{r,n}|\right)$$

which is $O(M^{-2})$ by summability of b_p^* and Lemma 7.3 part (b). \square

Lemma 7.7. *Assuming C1', C2 and C4, we have that*

$$\begin{cases} \sup_{t \geq 1} t^{1-2d} |\hat{\gamma}(t) - \gamma(t)| = O_p\left(\left|\hat{d} - d\right|\right) & \text{if } d > 0 \\ \sup_{t \geq 1} t |\hat{\gamma}(t)| = O_p\left(\left|\hat{d} - d\right|\right) & \text{if } d = 0. \end{cases}$$

Proof. By definition of $\hat{\gamma}(t)$ and $\gamma(t)$, we have that, uniformly in t ,
(7.16)

$$\hat{\gamma}(t) - \gamma(t) = (\hat{\gamma}(t-1) - \gamma(t-1)) \frac{t-1+d}{t-d} + \frac{\gamma(t-1)}{t-d} (\hat{d} - d) (1 + o_p(1)).$$

Noting that for $d=0$, $\gamma(t) = 0$ for $t \geq 1$ and $\hat{\gamma}(1) = O_p(\hat{d} - d)$, we obtain by induction that $\sup_{t \geq 2} t \hat{\gamma}(t) = O_p(\left|\hat{d} - d\right|)$. When $d > 0$, solving recursively (7.16), we obtain that $\hat{\gamma}(t) - \gamma(t) = D \left(1 + \frac{1}{t-d}\right) \gamma(t-1) (\hat{d} - d) (1 + o_p(1))$, which concludes the lemma. \square

Lemma 7.8. *For any $\check{n} < q_1 \leq q_2 < n - \check{n}$, as $n \rightarrow \infty$,*

$$(7.17) \quad \check{n}^{1-2d} \text{Cov}^*(\hat{r}^*(q_1), \hat{r}^*(q_2)) \xrightarrow{P} \rho(b; d).$$

Proof. The left side of (7.17), recalling that $\delta_t = E(u_0 u_t)$, is

$$\frac{1}{\tilde{n}^{1+2\hat{d}}} \sum_{t=1}^{\tilde{n}} \sum_{s=q_2-q_1+1}^{\tilde{n}+q_2-q_1} \left\{ (E^*(u_{t+q_1}^* u_{s+q_1}^*) - \delta_{t-s}) + \delta_{t-s} \right\} K_{+,t} K_{+,s+q_1-q_2}.$$

However, it suffices to show that the contribution due to $E^*(u_{t+q_1}^* u_{s+q_1}^*) - \delta_{t-s}$ in the last displayed expression converges in probability to zero because by Proposition 3.2, the term due to δ_{t-s} converges to $\rho(b; d)$. The first term is $o_p(1)$ by standard arguments and because $|K_t| \leq D$ by C2, and then by Proposition 4.2. \square

Lemma 7.9. *Assuming C1', C2 and C4,*

$$(7.18) \quad \sum_{j=1}^n E^* \left(n^{-1} \tilde{n}^{-1-2\hat{d}} |\varrho_q(\lambda_j) \eta_j^*|^2 \mathcal{I} \left(n^{-1} \tilde{n}^{-1-2\hat{d}} |\varrho_q(\lambda_j) \eta_j^*|^2 > \phi \right) \right) \xrightarrow{P} 0$$

for all $\phi > 0$, where $\varrho_q(\lambda_j) = \widehat{k}_q(\lambda_j) \widetilde{g}^{1/2}(-\lambda_j; \widehat{d}) \widehat{B}(\lambda_j)$ with $\widehat{k}_q(\lambda_j) = \sum_{t=q+1}^{\tilde{n}+q} K_{t-q} e^{it\lambda_j}$.

Proof. First, we show that

$$(7.19) \quad n^{-1} \tilde{n}^{-1-2\hat{d}} \max_{1 \leq j \leq n} |\varrho_q(\lambda_j)|^2 = O_p(a^{1-2d}).$$

Now, by definition of $|\widehat{B}(\lambda_j)|^2$ and $|B(\lambda_j)|^2$ and the triangle inequality,

$$\sup_{1 \leq j \leq n} \left| |\widehat{B}(\lambda_j)|^2 - |B(\lambda_j)|^2 \right| \leq \sup_{1 \leq j \leq n} \left| \sum_{|k| \leq M} (\widehat{\beta}_k - \beta_k) e^{-ik\lambda_j} \right| + \sup_{1 \leq j \leq n} \left| \sum_{|k| > M} \beta_k e^{-ik\lambda_j} \right|.$$

The second term on the right is $O(M^{-1})$ because C1 implies that $|\beta_k| = o(k^{-2})$, whereas the first term is, by definition of $\widehat{\beta}_k$ and β_k , bounded by

$$\sum_{|k| \leq M} |\widehat{b}_k - b_k|^2 + \sum_{|k| \leq M} |b_k| |\widehat{b}_k - b_k| = O_p(M^{-1})$$

by Lemmas 7.6 and 7.3, see also our arguments of the contribution of \varkappa_p into (6.14). So, because C1 implies that $\sup_{1 \leq j \leq n} |B(\lambda_j)|^2 = \sup_{1 \leq j \leq n} |h(\lambda_j)| < D$, (7.19) holds

because by integrability of $K(\cdot)$, $\left| \tilde{n}^{-1} \sum_{t=1+q}^{q+\tilde{n}} K_{t-q} \right|^2 \leq D$ and then by Lemma 7.7 we have that

$$\max_{1 \leq j \leq n} \left(\frac{1}{n \tilde{n}^{1+2\hat{d}}} \sum_{|\ell| < n} \widehat{\gamma}(\ell) e^{-i\ell\lambda_j} \left| \sum_{t=1+q}^{q+\tilde{n}} K_{t-q} e^{it\lambda_j} \right|^2 \right) = D \frac{\tilde{n}^{1-2\hat{d}}}{n} \sum_{|\ell| < n} |\gamma(\ell)| (1 + o_p(1))$$

which is $O_p(a^{1-2d})$ because $n^{2(\hat{d}-d)} - 1 = o_p(1)$ by Proposition 4.1. So, we conclude that (7.19) holds true, which implies that with probability approaching one the left side of (7.18) is bounded by

$$(7.20) \quad \sum_{j=1}^n E^* \left(n^{-1} \tilde{n}^{-1-2\hat{d}} |\varrho_q(\lambda_j) \eta_j^*|^2 \mathcal{I} \left(|\eta_j^*|^2 > D\phi a^{2d-1} \right) \right) \\ \leq D \sup_{1 \leq j \leq n} E^* \left(|\eta_j^*|^2 \mathcal{I} \left(|\eta_j^*|^2 > D\phi a^{2d-1} \right) \right) \frac{1}{n \tilde{n}^{1+2\hat{d}}} \sum_{j=1}^{n-1} |\varrho_q(\lambda_j)|^2.$$

However, by Proposition 4.1 and that $n^{2(\hat{d}-d)} - 1 = o_p(1)$, the second factor on the right of (7.20) is bounded in probability. On the other hand, because $E^* |\eta_j^*|^4 =$

$n^{-1} \sum_{t=1}^n \widehat{u}_t^4 + (n^{-1} \sum_{t=1}^n \widehat{u}_t^2)^2$, and say $n^{-1} \sum_{t=1}^n \widehat{u}_t^4 = n^{-1} \sum_{t=1}^n u_t^4 (1 + o_p(1))$, then by Proposition 3.4

$$E^* \left(|\eta_j^*|^2 \mathcal{I} \left(|\eta_j^*|^2 > D\phi a^{2d-1} \right) \right) \leq \frac{Da^{2-4d}}{\phi^2} \left\{ \frac{1}{n} \sum_{t=1}^n u_t^4 + \left(\frac{1}{n} \sum_{t=1}^n u_t^2 \right)^2 \right\} (1 + o_p(1)).$$

But, by a well-know argument, see Stout's (1974) Theorem 3.7.8, u_t is also ergodic by $C1$, which implies that the right side of the last displayed inequality converges to zero because $\phi > 0$, $d < 1/2$ and $a \rightarrow 0$ by $C3$. So, we conclude that (7.20) $\rightarrow_p 0$ for all $\phi > 0$ and the proof of the lemma. \square

REFERENCES

- [1] BERK, K. N. (1974): "Consistent autoregressive spectral estimates," *The Annals of Statistics*, **2**, 489-502.
- [2] BICKEL, P.J. AND ROSENBLATT, M. (1973): "On some global measures of the deviations of density function estimates," *Annals of Statistics*, **1**, 1071-1095.
- [3] BIERENS, H. AND PLOBERGER, W. (1997): "Asymptotic theory of integrated conditional moment test," *Econometrica*, **65**, 1153-1174.
- [4] BILLINGSLEY, P. (1968): *Convergence of probability measures*. Wiley and Sons. New York.
- [5] BRILLINGER, D.R. (1981): *Time series, data analysis and theory*. Holden-Day. San Francisco.
- [6] BROCKWELL, P.J. AND DAVIS, R.A. (1991): *Time series: theory and methods*. Springer. New York.
- [7] BÜHLMANN, P. (1997): "Sieve-bootstrap for time series," *Bernoulli*, **3**, 123-148.
- [8] CSÖRGÖ, S. AND MASON, D.M. (1989): "Bootstrapping empirical functions," *Annals of Statistics*, **17**, 1447-1471.
- [9] CSÖRGÖ, S. AND MIELNICZUK, J. (1995): "Nonparametric regression under long-range dependent normal errors," *Annals of Statistics*, **23**, 1000-1014.
- [10] DAHLHAUS, R., AND D. JANAS (1996): "A frequency domain bootstrap for ratio statistics in time series analysis," *Annals of Statistics*, **24**, 1934-1963.
- [11] DAVIES, R.B. AND HARTE, D.S. (1987): "Tests for Hurst Effect," *Biometrika*, **74**, 95-101.
- [12] DELGADO, M.A. AND GONZALEZ-MANTEIGA, W. (2001): "Significance testing in nonparametric regression based on the bootstrap," *Annals of Statistics*, **29**, 1469-1507.
- [13] DELGADO, M.A., HIDALGO, J. AND VELASCO, C. (2005): "Distribution free goodness-of-fit tests for linear processes," *Annals of Statistics*, **33**, 2568-2609.
- [14] EFRON, B. (1979): "Bootstrap methods: Another look at the jackknife," *Annals of Statistics*, **7**, 1-26.
- [15] EUBANK, R. AND SPIEGELMAN, S. (1990): "Testing the goodness of fit of a linear model via nonparametric functional forms," *Journal of the American Statistical Association*, **85**, 387-392.
- [16] FAN, Y. AND LI, Q. (1996): "Consistent model specification test: omitted variables and semiparametric functional forms," *Econometrica*, **64**, 865-890.
- [17] FULLER, W.A. (1996): *Introduction to statistical time series*. Wiley and Sons. New York.
- [18] GINÉ, E. AND ZINN, J. (1989): "Necessary conditions for the bootstrap of the mean," *Annals of Statistics*, **17**, 684-691.
- [19] HALL, P., JING, B.-Y. AND LAHIRI, S.N. (1998): "On the sampling window method for long-range dependent data," *Statistica Sinica*, **8**, 1189-1204.
- [20] HÄRDLE, W. AND MAMMEN, E. (1993): "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, **21**, 1926-1947.
- [21] HENRY, M. AND ROBINSON, P.M. (1996): "Bandwidth choice in Gaussian semiparametric estimation of long-range dependence," in *Athens Conference on Applied Probability and Time Series Analysis, Vol. II, 220-232: Time Series Analysis. In memory of E.J. Hannan*, Ed: P.M. Robinson and M. Rosenblatt. Springer Verlag, New York.
- [22] HIDALGO, J. (1997): "Nonparametric estimation with strongly dependent multivariate time series," *Journal of Time Series Analysis*, **18**, 95-122.
- [23] HIDALGO, J. (1999): "Nonparametric tests for model selection with time series data," *Test*, **8**, 365-398.
- [24] HIDALGO, J. (2000): "Nonparametric test for causality with long-range dependence," *Econometrica*, **68**, 1465-1490.
- [25] HIDALGO, J. (2003): "An alternative bootstrap to moving blocks for time series regression models," *Journal of Econometrics*, **117**, 369-399.

- [26] HIDALGO, J. AND ROBINSON, P.M. (2002): "Adapting to unknown disturbance autocorrelation in regression with long memory," *Econometrica*, **70**, 1545-1581.
- [27] HIDALGO, J. AND YAJIMA, Y. (2002): "Prediction in the frequency domain under long-range processes with application to the signal extraction problem," *Econometric Theory*, **18**, 584-624.
- [28] HOLLY, A. (1982): "A remark on Hausman's specification test," *Econometrica*, **50**, 749-760.
- [29] HONG, Y. AND WHITE, H. (1995): "Consistent specification testing via nonparametric series regressions," *Econometrica*, **63**, 1133-1159.
- [30] HURVICH, C.M. (2001): "Model selection for broadband semiparametric estimation of long-memory in time series," *Journal of Time Series Analysis*, **22**, 679-709.
- [31] HURVICH, C.M. AND ZEGER, S.L. (1987): "Frequency domain bootstrap methods for time series." *Technical report 87-115*, Graduate School of Business Administration, New York University.
- [32] IBRAGIMOV, I.A. AND ROZANOV, Y.A. (1978): Gaussian random processes. Springer-Verlag, Berlin.
- [33] JOHNSTON, G.J. (1982): "Probabilities of maximal deviation for nonparametric regression function estimates," *Journal of Multivariate Analysis*, **12**, 402-414.
- [34] KOUL, H.L. AND STUTE, W. (1999): "Nonparametric model checks for time series," *Annals of Statistics*, **27**, 204-236.
- [35] KREISS, J.-P. (1998): "Asymptotic statistical inference for a class of stochastic processes," *Habilitationsschrift*, Fachbereich Mathematik, Univ. Hamburg.
- [36] KREISS, J.-P. AND PAPARODITIS, E. (2003): "Autoregressive-aided periodogram bootstrap for time series," *Annals of Statistics*, **31**, 1923-1955.
- [37] KÜNSCH, H.A. (1989): "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, **17**, 1217-1241.
- [38] LOBATO, I. AND ROBINSON P.M. (1997): "A nonparametric test for $I(0)$," *Review of Economic Studies*, **65**, 475-495.
- [39] MARINUCCI, D. AND ROBINSON, P.M. (2003): "Weak convergence of multivariate fractional processes," *Stochastic Processes and its Applications*, **80**, 103-120.
- [40] PICKANDS, J. (1969): "Asymptotic properties of the maximum in a stationary gaussian process," *Transactions of the American Mathematical Society*, **145**, 75-86.
- [41] POLITIS, D.N. AND ROMANO, J.P. (1994): "Large sample confidence regions based on subsamples under minimal assumptions," *Annals of Statistics*, **22**, 2031-2050.
- [42] POLLARD, D. (1984): Convergence of stochastic processes. Springer, New York.
- [43] PRICHARD, D. AND THEILER, J. (1994): "Generating surrogate data for time series with several simultaneously measured variables," Working paper #94-04-023. Santa Fe Institute.
- [44] ROBINSON, P.M. (1991): "Automatic frequency domain inference on semiparametric and nonparametric models," *Econometrica*, **59**, 1329-1363.
- [45] ROBINSON, P.M. (1995a): "Log-Periodogram regression for time series with long range dependence," *Annals of Statistics*, **23**, 1048-1072.
- [46] ROBINSON, P.M. (1995b): "Gaussian semiparametric estimation of long-range dependence," *Annals of Statistics*, **23**, 1630-1661.
- [47] ROBINSON, P.M. (1997): "Large-sample inference for nonparametric regression with dependent errors," *Annals of Statistics*, **25**, 2054-2083.
- [48] ROBINSON, P.M. AND HIDALGO, F.J. (1997): "Time series regression with long-range dependence," *Annals of Statistics*, **25**, 77-104.
- [49] SERFLING, R.J. (1980): Approximation theorems of mathematical statistics. *Wiley and Sons*, New York.
- [50] STUTE, W. (1997): "Nonparametric model checks for regression," *Annals of Statistics*, **25**, 613-641.
- [51] STOUT, W.F. (1974): Almost sure convergence. *Academic Press*. New York.
- [52] THEILER, J., GALDRIKIAN, B., LONGTIN, A., EUBANK, S. AND FARMER, J.D. (1992): "Using surrogate data to detect nonlinearity in time series," A proceedings volume in the Santa Fe Institute Studies in the Sciences of Complexity. Nonlinear Modeling and Forecasting. Ed: Casdagli, M. and Eubank, S., 163-188.
- [53] WANG, Q., LIN, Y.-X. AND GULATI, C.M. (2003): "Strong approximation for long memory processes with applications," *Journal of Theoretical Probability*, **16**, 377-389.
- [54] WOODROOFE, M. (1967): "On the maximum deviation of the sample density," *Annals of Mathematical Statistics*, **38**, 475-481.
- [55] YAJIMA, Y. (1991): "Asymptotic properties of the LSE in a regression model with long-memory stationary errors," *Annals of Statistics*, **19**, 158-177.
- [56] YONG, C.H. (1974): Asymptotic behaviour of trigonometric series. Hong Kong: Chinese University of Hong Kong.

- [57] ZHENG, J.X. (1996): "A consistent test of functional form via nonparametric estimation techniques," *Journal of Econometrics*, **75**, 263-289.

TABLE 5.1

Proportion of rejections in 1000 Monte Carlo experiments under $H_0 : E[y_t | x_t] = \beta_0$. The bandwidth parameters employed are $m = n/16$ and $a = 0.5n^{-1/4}$.

	Size (%)		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$
	$n = 64$					
$d_x = 0.0$	1		1.20	1.20	1.10	1.20
	5		5.10	4.70	4.80	4.90
	10		9.40	8.90	9.00	9.40
$d_x = 0.2$	1		0.80	1.40	1.40	1.40
	5		4.30	4.60	4.40	4.20
	10		8.60	8.60	8.60	8.30
$d_x = 0.3$	1		1.30	1.00	1.30	1.40
	5		4.00	3.60	4.00	4.30
	10		8.00	7.30	7.80	8.20
$n = 128$						
$d_x = 0.0$	1		1.20	1.00	1.20	1.30
	5		4.80	4.80	4.70	4.70
	10		9.40	10.40	9.50	9.30
$d_x = 0.2$	1		1.70	1.50	1.60	1.60
	5		5.50	5.20	5.30	4.80
	10		9.30	9.10	9.20	9.40
$d_x = 0.3$	1		1.40	1.20	1.40	1.40
	5		4.10	4.00	4.60	4.90
	10		8.10	8.60	8.60	8.20
$n = 256$						
$d_x = 0.0$	1		1.30	2.30	4.00	8.00
	5		5.10	6.10	7.30	11.30
	10		9.70	11.20	11.80	14.70
$d_x = 0.2$	1		1.10	1.60	3.70	7.00
	5		5.20	4.30	6.20	9.30
	10		10.00	10.10	11.30	12.80
$d_x = 0.3$	1		0.70	0.90	2.90	6.30
	5		4.70	5.00	6.00	9.00
	10		9.50	9.40	10.30	12.60

The bootstrap critical values are computed using 499 bootstrap samples

TABLE 5.2

Proportion of rejections in 1000 Monte Carlo experiments under $H_0 : E[y_t | x_t] = \beta_0$. The bandwidth parameters employed are $m = n/8$ and $a = 0.5n^{-1/4}$.

	Size (%)		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$
	$n = 64$					
$d_x = 0.0$	1	1.40	1.80	1.20	1.80	
	5	5.30	5.30	4.70	5.40	
	10	10.30	9.90	9.50	10.80	
$d_x = 0.2$	1	0.90	1.00	1.40	1.60	
	5	4.70	4.60	4.30	4.80	
	10	8.20	8.50	8.10	8.90	
$d_x = 0.3$	1	1.00	1.20	1.60	1.80	
	5	3.60	3.60	4.20	4.00	
	10	8.00	7.70	7.50	8.90	
$n = 128$						
$d_x = 0.0$	1	0.90	0.90	1.20	1.40	
	5	4.20	4.10	4.40	5.10	
	10	9.00	10.10	9.60	10.00	
$d_x = 0.2$	1	1.10	1.20	1.10	2.10	
	5	5.30	5.10	5.30	5.60	
	10	9.10	9.10	9.40	10.40	
$d_x = 0.3$	1	1.10	1.00	1.50	1.30	
	5	4.10	4.10	4.80	4.90	
	10	8.70	8.90	8.90	8.60	
$n = 256$						
$d_x = 0.0$	1	1.10	1.20	2.20	5.60	
	5	4.80	5.20	5.30	9.20	
	10	9.90	10.70	10.30	13.00	
$d_x = 0.2$	1	1.10	1.10	1.30	4.40	
	5	5.00	4.30	4.50	7.20	
	10	9.90	9.90	9.80	10.90	
$d_x = 0.3$	1	0.60	0.70	1.20	3.80	
	5	4.50	4.80	5.50	7.30	
	10	9.30	10.00	9.90	11.40	

The bootstrap critical values are computed using 499 bootstrap samples

TABLE 5.3

Proportion of rejections in 1000 Monte Carlo experiments under $H_0 : E[y_t | x_t] = \beta_0$. The bandwidth parameters employed are $m = n/4$ and $a = 0.5n^{-1/4}$.

	Size (%)		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$
	$n = 64$					
$d_x = 0.0$	1	0.90	0.90	1.40	1.80	
	5	4.80	4.70	4.70	5.80	
	10	9.60	9.60	9.70	11.00	
$d_x = 0.2$	1	0.90	1.30	1.30	1.40	
	5	4.50	4.90	4.30	4.20	
	10	8.10	8.00	8.50	8.70	
$d_x = 0.3$	1	1.40	1.30	1.30	1.50	
	5	3.70	3.70	3.80	3.90	
	10	8.30	7.60	7.80	8.40	
$n = 128$						
$d_x = 0.0$	1	0.80	0.70	1.00	1.80	
	5	3.90	4.30	4.70	5.20	
	10	9.20	10.10	9.30	10.20	
$d_x = 0.2$	1	0.90	1.10	1.00	1.70	
	5	5.20	5.00	5.50	5.50	
	10	9.20	9.20	9.20	10.10	
$d_x = 0.3$	1	1.20	1.00	1.10	1.10	
	5	3.90	4.10	5.10	4.70	
	10	8.80	9.10	8.90	9.00	
$n = 256$						
$d_x = 0.0$	1	1.00	1.00	1.20	3.10	
	5	5.00	4.70	4.90	6.30	
	10	9.70	10.30	9.30	10.90	
$d_x = 0.2$	1	1.10	1.30	1.20	2.10	
	5	5.30	4.20	4.00	5.00	
	10	10.00	10.00	9.10	8.70	
$d_x = 0.3$	1	0.80	0.70	0.80	2.10	
	5	4.60	5.20	5.40	5.80	
	10	9.40	9.90	9.90	10.10	

The bootstrap critical values are computed using 499 bootstrap samples

TABLE 5.4

Proportion of rejections in 1000 Monte Carlo experiments under the alternative $E[y_t|x_t] = \beta_0 + 0.5x_t$. The bandwidth parameters employed are $m = n/8$ and $a = 0.5n^{-1/4}$.

$n = 64$		Size (%)	$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	
	$d_x = 0.0$	1	9.20	10.00	12.10	17.70	
		5	27.50	28.40	30.90	39.70	
		10	41.20	41.40	43.70	53.50	
	$d_x = 0.2$	1	8.60	9.70	10.00	16.10	
		5	25.60	26.70	29.40	37.80	
		10	40.10	39.80	42.40	49.70	
	$d_x = 0.3$	1	9.20	9.60	11.60	16.30	
		5	28.60	28.90	29.70	39.30	
		10	40.40	40.70	43.90	53.60	
	$n = 128$		Size (%)	$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$
		$d_x = 0.0$	1	19.70	20.80	23.30	31.90
5			42.00	43.40	46.40	57.80	
10			55.60	56.40	61.50	73.50	
$d_x = 0.2$		1	19.80	20.10	21.70	32.50	
		5	45.00	46.20	48.20	60.10	
		10	59.80	60.70	63.40	74.00	
$d_x = 0.3$		1	21.50	20.90	22.30	32.30	
		5	45.50	46.70	48.60	59.20	
		10	61.80	61.10	62.30	72.00	
$n = 256$			Size (%)	$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$
		$d_x = 0.0$	1	43.10	45.10	49.70	66.90
	5		73.60	73.70	77.00	86.70	
	10		84.40	84.80	85.90	92.90	
	$d_x = 0.2$	1	40.70	40.60	45.50	60.80	
		5	71.00	71.30	73.40	82.30	
		10	82.30	82.10	84.20	89.80	
	$d_x = 0.3$	1	44.90	44.50	45.70	59.10	
		5	72.10	70.90	72.00	78.80	
		10	82.90	83.20	82.90	86.90	

The bootstrap critical values are computed using 499 bootstrap samples

TABLE 5.5

Proportion of rejections in 1000 Monte Carlo experiments under the alternative $E[y_t|x_t] = \beta_0 + x_t$. The bandwidth parameters employed are $m = n/8$ and $a = 0.5n^{-1/4}$.

	Size (%)		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$	
	$n = 64$	$d_x = 0.0$	1	31.10	32.40	35.90	43.90
5			63.50	64.40	67.80	74.20	
10			76.80	77.90	79.80	84.50	
$d_x = 0.2$		1	35.00	36.20	39.10	48.70	
		5	66.30	66.60	68.60	77.70	
		10	79.20	80.70	82.40	88.70	
$d_x = 0.3$		1	41.10	42.90	44.40	54.10	
		5	70.20	70.40	72.80	81.10	
		10	82.10	82.60	83.80	89.60	
$n = 128$		$d_x = 0.0$	1	63.80	66.80	69.70	80.90
			5	90.20	90.60	92.30	95.40
			10	96.00	95.90	97.20	98.30
	$d_x = 0.2$	1	70.50	71.20	74.60	83.40	
		5	92.40	92.30	94.20	96.20	
		10	97.00	96.90	97.70	99.10	
	$d_x = 0.3$	1	73.80	73.30	76.40	84.10	
		5	94.20	93.60	94.30	96.60	
		10	97.60	97.50	97.50	99.20	
	$n = 256$	$d_x = 0.0$	1	96.30	96.20	97.00	98.60
			5	99.50	99.80	99.90	100.0
			10	100.0	100.0	100.0	100.0
$d_x = 0.2$		1	96.80	97.20	97.60	98.70	
		5	100.0	99.90	100.0	100.0	
		10	100.0	100.0	100.0	100.0	
$d_x = 0.3$		1	96.10	96.60	96.30	98.00	
		5	99.80	99.70	99.50	99.60	
		10	100.0	99.90	99.80	99.80	

The bootstrap critical values are computed using 499 bootstrap samples

TABLE 5.6

Proportion of rejections in 1000 Monte Carlo experiments under the alternative $E[y_t|x_t] = \beta_0 + \sin(x_t)$. The bandwidth parameters employed are $m = n/8$ and $a = 0.5n^{-1/4}$.

	Size (%)		$d = 0.1$	$d = 0.2$	$d = 0.3$	$d = 0.4$
	$n = 64$					
$d_x = 0.0$	1		41.80	44.50	47.70	63.00
	5		69.50	71.40	75.20	85.50
	10		81.00	82.90	85.30	92.50
$d_x = 0.2$	1		37.60	39.80	45.30	58.00
	5		67.30	69.50	74.20	84.40
	10		81.50	82.70	86.10	92.70
$d_x = 0.3$	1		37.60	39.10	43.60	58.80
	5		67.90	68.60	71.70	83.00
	10		79.30	80.80	83.60	90.20
$n = 128$						
$d_x = 0.0$	1		71.80	74.30	78.10	90.10
	5		91.20	92.60	94.90	98.30
	10		96.70	96.30	97.90	99.50
$d_x = 0.2$	1		73.80	75.10	78.70	88.60
	5		93.20	93.60	94.90	98.30
	10		97.00	97.10	97.60	99.20
$d_x = 0.3$	1		71.90	72.30	76.60	88.00
	5		92.80	92.80	93.10	97.20
	10		97.10	97.40	97.80	98.40
$n = 256$						
$d_x = 0.0$	1		96.80	97.70	98.40	99.90
	5		100.0	100.0	100.0	100.0
	10		100.0	100.0	100.0	100.0
$d_x = 0.2$	1		97.60	97.40	98.20	99.30
	5		99.80	99.80	99.80	100.0
	10		100.0	99.90	99.80	100.0
$d_x = 0.3$	1		97.00	96.90	96.20	98.10
	5		99.60	99.50	99.60	99.50
	10		99.90	99.80	99.80	99.80

The bootstrap critical values are computed using 499 bootstrap samples