BOOTSTRAP TEST FOR BREAKS OF A REGRESSION MODEL WITH DEPENDENT DATA

J. HIDALGO

Abstract. The paper describes and examines a test for breaks in a nonparametric regression function with dependent errors. We show that after normalization the limit distribution of the test is a Gumbel distribution. However, as (a) the normalization constants are model dependent and difficult to estimate and (b) the rate of convergence of the finite sample distribution of the statistic to the asymptotic one is very slow, see Hall (1979), inferences based on the asymptotic distribution may be difficult to perform or not be very reliable. For those reasons, we describe and examine the bootstrap analogue of the test by mean of bootstrapping the model in the "frequency domain", showing its asymptotic validity under suitable regularity conditions.


1. INTRODUCTION

This paper examines a test for smoothness or lack of breaks in a regression model in the presence of dependent observations. We focus on nonparametric models, both for the regression function and the dependence structure of the data. Our framework and assumptions are, for the most part, similar to those given in Robinson (1997), so that we allow for long-memory, short memory and negative dependence simultaneously. We shall consider the following regression model

\[(1.1) \quad y_t = r \left( \frac{t}{n} \right) + u_t; \quad t = 1, \ldots, n,\]

where the errors \(u_t\) are assumed to follow a covariance stationary linear process. Our primary objective is to develop a test for

\[(1.2) \quad H_0 : r_+ (x) = r_- (x), \quad \forall x \in (0, 1)\]

where \(r_+ (x) = \lim_{z \to x^+} r (z)\) and \(r_- (x) = \lim_{z \to x^-} r (z)\), being the alternative

\[(1.3) \quad H_1 : \exists x \in (0, 1), \quad |r_+ (x) - r_- (x)| > 0,\]

that is the negation of the null. Although it seems possible to extend the results to situations for which we allow under the alternative, that \(r_+ (z)\) and/or \(r_- (z)\) may diverge to infinity as \(z \to x \pm\), we will only consider explicitly the hypothesis testing described in (1.2) – (1.3) for which the jump, if there is any, is of bounded magnitude. On the other hand, our work extends others, see for a latter reference Müller and Stadtmüller (1999) and references therein, to dependent data.

We shall base our test for \(H_0\) on the comparison of (nonparametric) estimates of the regression function, specifically of \(r_+ (x)\) and \(r_- (x)\). In our context, there are several reasons why the implementation of a bootstrap algorithm for the test is of interest. For example, Bühlmann (1998) has indicated and identified some of advantages of the bootstrap when compared to other approaches such as Edgeworth
expansions. On the other hand, because our test converges in distribution to an extreme-value distribution, in particular to the Gumbel distribution, we envisage some other reasons for which the bootstrap approach appears more convenient as we now argue.

Although Theorem 2.4 below gives the (asymptotic) justification of the test, following Hall (1979), the rate of convergence of the finite sample distribution to the asymptotic one is very slow. In particular, the rate of convergence is of logarithmic order. Moreover, the work by Konakov and Piterbarg (1984), see also Hall (1991), suggests that this rate is the best possible one. So, critical values obtained from the asymptotic distribution function can be a poor approximation to those computed from the actual finite sample distribution of the test. To obtain better approximations to the finite sample distribution, there are mainly two solutions or strategies, namely via Edgeworth expansions or via bootstrap techniques. However, with regard to the first possible solution, as Konakov and Piterbarg (1984) showed, the order of approximation between the Edgeworth expansion and the finite sample distributions is only slightly improved. In particular, it is of order $O(n^{-\delta})$ for any $\delta > 0$, so that, the rate is still very slow. In addition, using theory on large deviations, Hall (1990) has given a theoretical justification on the otherwise empirical observation that Edgeworth corrections do not do a good job when they are compared to bootstrap schemes at the tails of the distribution, which is precisely the most interesting region when testing. On the other hand, Hall (1991) has showed that bootstrap schemes do a better job in that the accuracy of the bootstrap distribution to the finite sample distribution of the suprema is $O\left((na)^{-1/2}\log^2 n\right)$, where $a = a(n)$ is a sequence of positive numbers which converge to zero as $n \to \infty$, and which will be identified as the bandwidth parameter in our kernel regression estimates. In addition, as we can see from Theorem 2.4 below, to obtain the asymptotic distribution of the test, some of the normalization constants needed are model dependent and difficult to obtain, so that the implementation of the test can be quite difficult for a given data set, albeit that new (estimated) critical values have to be computed with a new data set. These reasons suggest that the second possible solution, that is bootstrap techniques, may be very useful and helpful method, not only to obtain reliable inferences for realistic sample sizes we encounter in real applications, but to implement the test.

The previous comments motivate the second objective of the paper, that is to implement a valid bootstrap version of the test. In a time series context as ours, bootstrap algorithms are often based on the Moving Block Bootstrap as proposed by Künsch (1988) or the subsampling scheme as in, for example, Politis and Romano (1994). Assuming that the data has an AR($\infty$) representation, another algorithm is the sieve-bootstrap, introduced apparently in Kreiss (1988) and further examined by Bühlmann (1997). The latter bootstrap is based on a series approximation of the AR($\infty$) representation of the process by an AR($p_n$), where $p_n \to \infty$ with the sample size $n$. The latter parameter $p_n$ can be chosen by some criteria function such as the AIC in applications with real data. Because the apparent better finite sample behaviour of the latter bootstrap when compared to the former ones observed in Bühlmann (1997), Bühlmann (1998) employs the sieve-bootstrap to bootstrapping the estimates of a nonparametric regression model similar to ours. However, since we want to relax the dependence structure of the data to allow for, possibly, long-memory, the previous bootstraps may not be valid for the statistics we are interested in. Specifically, when investigating the supremum of a sequence of random variables. Thus, in this paper we propose a modification of the sieve-bootstrap by performing our bootstrap in the frequency domain showing that it is a valid algorithm. In particular, the bootstrap is based on the "discrete" Crâmer
representation of \( u_t \) together with Barlett approximation of the discrete Fourier transform of \( u_t \) in terms of that for the innovations of the process \( u_t \). It is worth noting that our bootstrap method draws some similarities as when estimating the spectral density function either in the frequency domain or by an AR approximation of the correlation structure of \( u_t \) as in Berk (1974).

The paper is organized as follows. In the next section, we describe the test for \( H_0 \) given in (1.2) and establish its asymptotic properties. Section 3 presents a bootstrap algorithm for the test, showing its validity. Section 4 provides the proofs of our results which make use of a series of lemmas in Section 5. Finally, Section 6 concludes and gives some possible extensions.

2. THE TEST FOR SMOOTHNESS AND CONDITIONS

We begin by introducing the following conditions:

\textbf{C1:} \( \{ u_t \} \) is a covariance stationary linear process defined as

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

\textbf{C2:} \( \{ \varepsilon_t \} \) is an ergodic process that satisfies (a) \( E (\varepsilon_t | F_{t-1}) = 0 \),

(b) \( E (\varepsilon_t^2 | F_{t-1}) = E (\varepsilon_t^2) = \sigma_\varepsilon^2 \) a.s., (c) \( E (|\varepsilon_t|^4 | F_{t-1}) = \mu_\varepsilon < \infty \) for \( \ell = 3, 4 \)

and (d)

\[
cum (\varepsilon_{t_1}, \varepsilon_{t_2}, \varepsilon_{t_3}, \varepsilon_{t_4}) = \begin{cases} \kappa, & t_1 = t_2 = t_3 = t_4, \\ 0, & \text{otherwise}, \end{cases}
\]

with \( |\kappa| < \infty \), where \( F_t \) is the \( \sigma \)-algebra of events generated by \( \varepsilon_s, s \leq t \).

\textbf{C3:} The process \( u_t \) has an absolutely continuous spectral distribution function, its spectral density, \( f (\lambda) \), being of the form

\[ f (\lambda) = g (\lambda) h (\lambda), \quad \lambda \in (-\pi, \pi], \]

where (i) \( h \) is an even nonnegative function that is twice continuously differentiable and positive for all \( \lambda \in [0, \pi] \), and we denote \( H = h (0) \); (ii) \( g \) is an integrable function such that

\[
\gamma_j = \int_{-\pi}^{\pi} g (\lambda) \cos (j \lambda) \, d\lambda
\]

\[ \sim \theta (d) j^{2d-1} \] as \( j \to \infty \), \( d \in \left(-\frac{1}{2}, \frac{1}{2}\right) \setminus \{0\}, \)

\[ = 2\pi \Delta_{j\lambda}, \quad d = 0, \]

where \( \Delta_{ab} \) is the Kronecker delta, \( \theta (d) = 2\Gamma (1 - 2d) \cos \left(\frac{\pi}{2} (d - d)\right) \) and also

\[ g (0) = 0, \quad -\frac{1}{2} < d < 0. \]

The case \( d = 0 \) refers to short-memory, whereas the cases \( 0 < d < \frac{1}{2} \) and \( -\frac{1}{2} < d < 0 \) refer to long-memory and antipersistence dependence, respectively.

\textbf{C4:} For all \( x \in [0, 1] \), the regression function \( r (x) \) satisfies

\[
\lim_{y \to x} \left| \frac{r (y) - r (x) - Q (x)}{|x - y|} \right| = o (1)
\]

where \( 0 < \tau \leq 2 \) and \( \sup_{x \in [0, 1]} |Q (x)| \leq D \).

Conditions \( C1 - C3 \) are identical to those in Robinson (1997), except that we required \( h (\lambda) \) to be twice continuously differentiable in \([0, \pi]\). Although the latter is not needed in its present format, when examining the asymptotic distribution of \( T \), given in (2.2) below, this condition eases the derivation and the proof of the
validity of the bootstrap algorithm described in Section 3. Condition C4 is only slightly stronger than functions \( r(x) \) which are Lipschitz continuous of order \( \tau \) if \( 0 < \tau \leq 1 \), or \( r(x) \) is differentiable with derivative satisfying a Lipschitz condition of degree \( \tau - 1 \), if \( 1 < \tau \leq 2 \). For instance, if \( \tau = 2 \), C4 means that \( r(x) \) is twice continuously differentiable.

We now describe the kernel estimators of \( r_+ (x) \) and \( r_- (x) \). To that end, we shall employ the so-called one-sided kernels as proposed by Rice (1984). These kernels were introduced to obtain consistent estimates of curves when they are estimated at the boundary of their compact support. These type of kernels appears necessary by looking at (1.2) and (1.3), since to implement the test we need to estimate \( r_+ (\cdot) \) and \( r_- (\cdot) \), that is the value of \( r(z) \) at \( z+ \) and \( z- \), respectively. To that end, let \( K_+ (x) \) and \( K_- (x) \) be one-sided kernels, that is kernel functions which take values for \( x > 0 \) or \( x < 0 \), respectively. Define \( K_\pm^{(m)}(x) = \partial^m K_\pm(x) / \partial x^m \), \( K_\pm^{(0)}(x) = K_\pm(x), \) \( m = 0, 1, \ldots \). The kernels \( K_\pm(x) \) satisfy the following condition:

\[
\text{C5: } K_+ : [0, 1] \to \mathbb{R} \text{ and } K_- : [-1, 0] \to \mathbb{R}, \text{ where } K_+^{(0)}(x) = K_-^{(0)}(-x), \\
K_+^{(1)}(x) = -K_-^{(1)}(-x), K_+^{(0)}(0) = 0, \int_0^1 K_+^{(1)}(x) \, dx = 1 \text{ and } \int_a^1 x K_+^{(1)}(x) \, dx = 0.
\]

Kernels \( K_+ \), and therefore \( K_- \), satisfying C5 can be obtained from any function \( \ell(x) \) with domain in \([0, 1]\), as \( K_+ (x) = \ell(x) (c_1 + c_2 x) \) where the constants \( c_1 \) and \( c_2 \) are the solutions to \( \int_0^1 \ell(x) (c_1 + c_2 x) \, dx = 1 \) and \( \int_1^a \ell(x) (c_1 + c_2 x) \, dx = 0 \). As an example, let \( \ell(x) = 1 \), then \( K_+ (x) = x (18 - 24x) \). It is worth mentioning that as in Robinson (1997), C5 can be modified to allow for Kernels with infinite support, e.g. with domain in \( \mathbb{R}_\pm \), with the extra conditions that \( K_+ (x) = O \left( (1 + x^2)^{-1} \right) \) and \( K_-^{(1)}(x) = O \left( (1 + x^{1+\rho})^{-1} \right) \) for some \( \rho > 0 \). However, the proofs are already quite lengthy and technical and Kernels with compact support ease some of their arguments. Finally, it is worth noting that other types of kernels seem possible, such as Fan’s (1993) local kernel polynomial or Gasser-Müller’s (1984) kernel.

Our kernel regression estimators of \( r_+ (x) \) and \( r_- (x) \) at a point \( x_q = q/n \), \( q = [na] + 1, \ldots, n - [na] - 1 \), are given by

\[
\tilde{r}_+(q) := \tilde{r}_+(x_q) = \frac{1}{n a} \sum_{t=q}^{n} y_t K_+ \left( \frac{t-q}{na} \right), \quad \tilde{r}_-(q) := \tilde{r}_-(x_q) = \frac{1}{n a} \sum_{t=1}^{q} y_t K_- \left( \frac{t-q}{na} \right), \tag{2.1}
\]

respectively, where \( a = a(n) \) is a bandwidth parameter which satisfies:

\[
\text{C6: } \text{As } n \to \infty, \ (i) \ (na)^{-1} + n^{-2d} a^{2-2d} \to 0 \text{ and } (ii) \ n^{\frac{1}{2} -d} a^{\frac{1}{2} + d + \tau} < D.
\]

The second term on the left in C6(i) is vacuous if we assume that \( d \geq 0 \). Part (ii) needs more explanation. Part (ii) of C6 differs from the analogue assumed by Robinson (1997). We, contrary to the latter work, do not need to assume that \( n^{\frac{1}{2} -d} a^{\frac{1}{2} + d + \tau} \to 0 \) as \( n \to \infty \). This allows us to be able to choose the optimal bandwidth parameter \( a \), in the sense of that value \( a \) which minimizes the MSE of the nonparametric regression estimate. More precisely, suppose that \( d = 0 \) and that \( \tau = 2 \) in C4. Then, it is known that the optimal choice of \( a \) satisfies \( a = D n^{-2/5} \) for some finite positive constant \( D \). Note that this choice \( a \) would be that obtained if some cross-validation criterion were employed for the choice of the bandwidth parameter. Also, observe that for the level of smoothness of \( r(x) \), that is \( \tau \) in C4, the bandwidth parameter \( a \) converges to zero slower as the value of \( d \) increases.

Because under \( H_0, \ r_+(x) = -r_-(x) \) for all \( x \in [0, 1] \), we should expect that \( \tilde{r}_+(q) \approx \tilde{r}_-(q) \), for all \( q = \tilde{n} + 1, \ldots, n - \tilde{n} - 1 \), where henceforth we abbreviate \([na]\) by \( \tilde{n} \). This suggests that the test for jumps or breaks can be based on the difference between the kernel estimates \( \tilde{r}_+(q) \) and \( \tilde{r}_-(q) \) given in (2.1). That is, it indicates
that the test can be based on

\[ T_r = \sup_{\hat{\theta} < r < \hat{\theta} - \frac{1}{\hat{n}}} |\hat{\theta}_+ (q) - \hat{\theta}_- (q)|. \]

Before we examine the properties of \( T_r \) given in (2.2), we will first examine the covariance of \( \hat{\theta}_+ (q) - \hat{\theta}_- (q) \) at two different points \( x_{q_1} \leq x_{q_2} \). Define \( b(q_1, q_2) = (q_2 - q_1) / \hat{n} \). Also, herewith, \( D \) denotes a positive finite constant.

**Proposition 2.1.** Assuming \( C1 - C6 \), under \( H_0 \), for any \( \hat{n} < q_1 \leq q_2 < n - \hat{n} \), as \( n \to \infty \),

\[ \hat{n}^{1/2d} \text{Cov} (\hat{\theta}_+ (q_1) - \hat{\theta}_- (q_1), \hat{\theta}_+ (q_2) - \hat{\theta}_- (q_2)) = \rho_+ (b; d) + \rho_- (b; d) - \rho_\pm (b; d) - \rho_\mp (b; d) \]

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

(a) if \( 0 < d < \frac{1}{4} \),

\[ \rho_+ (b; d) = H \Theta (d) \int_0^1 \int_{b - 1}^{b + b} |v - w|^{2d - 1} K_+ (v) K_+ (w - b) \, dv \, dw, \]

(b) if \( d = 0 \), \( \rho_+ (b; d) = \rho_- (b; d) = 4^{-1} H \int_0^1 K_+^2 (v) \, dv \) and \( \rho_\pm (b; d) = \rho_\mp (b; d) = 0 \),

(c) if \( -\frac{1}{4} < d < 0 \),

\[ \rho_+ (b; d) = H \Theta (d) \int_0^1 \int_{b - 1}^{b + b} |v - w|^{2d - 1} K_+ (v) \{K_+ (w - b) - K_+ (v - b)\} \, dv \, dw, \]

\[ \rho_- (b; d) = H \Theta (d) \int_0^1 \int_{b - 1}^{b + b} |v - w|^{2d - 1} K_- (v) \{K_- (w - b) - K_- (v - b)\} \, dv \, dw, \]

\[ \rho_\pm (b; d) = H \Theta (d) \int_0^1 \int_{b - 1}^{b + b} |v - w|^{2d - 1} K_+ (v) \{K_- (v) - K_- (v - b)\} \, dv \, dw, \]

\[ \rho_\mp (b; d) = H \Theta (d) \int_0^1 \int_{b - 1}^{b + b} |v - w|^{2d - 1} K_- (v) \{K_+ (v) - K_+ (v - b)\} \, dv \, dw. \]

**Proof.** The proof of this result or any other is confined to Section 4. \( \square \)

The next proposition deals with the correlation structure of \( \hat{\theta}_+ (q) \) and \( \hat{\theta}_- (q) \) as \( b(q_1, q_2) \to 0 \) and when \( b(q_1, q_2) \to \infty \). This will play an important role when studying the limit distribution of our test \( T_r \) given in (2.2).

**Proposition 2.2.** Under \( C1 - C6 \), as \( n \to \infty \),

\[ \text{(a) } \frac{\text{Cov} (b(q_1, q_2))}{\text{Cov} (b(q_1, q_1))} = 1 - Db^a (q_1, q_2) + o (b^a (q_1, q_2)) \quad \text{as } b(q_1, q_2) \to 0 \]

\[ \text{(b) } \text{Cov} (b(q_1, q_2)) \log b(q_1, q_2) = o (1) \quad \text{as } b(q_1, q_2) \to \infty. \]

The next proposition will examine the finite dimensional distributions of \( \hat{\theta}_+ (q) - \hat{\theta}_- (q) \) when the latter is considered as a process indexed by \( q \).
Proposition 2.3. Assuming $C1 - C6$, under $H_0$, for any finite collection $q_j$, $j = 1, ..., p$, such that $\bar{n} < q_j < n - \bar{n}$ and $|q_j - q_{j+1}| \geq nz > 0$, as $n \to \infty$,
\[
\hat{\theta}_n^{-d} \hat{\theta}_n^{-d} (0; d) (\hat{T}_n (q_j) - \hat{T}_n (q_{j+1}))_{j=1, ..., p} \to N (0, \text{diag} (1, ..., 1)),
\]
where $\rho (0; d) = \rho_+ (0; d) + \rho_- (0; d) - \rho_{\pm} (0; d) - \rho_{\pm} (0; d)$.

Some comments on the results of these propositions are worth mentioning. The first issue that we observe is that even when the bandwidth parameter $a$ is chosen optimally, there is not asymptotic bias. This is in clear contrast to standard kernel regression estimation results, for which a bias term appears in the asymptotic distribution, when $a$ is chosen to minimize the $MSE$, e.g. when $a$ is chosen as in $C6$. These results will have some rather important consequences when implementing the bootstrap analogue of our test $T_r$ given in (2.2). In particular, the most immediate consequence is that the same bandwidth parameter $a$ can be employed all throughout to obtain the bootstrap analogue of (1.1). On the other hand, Proposition 1 indicates that the covariance structure is independent of the points at which the regression function is estimated and only depends on how “far” are the points at which $r (x)$ is estimated.

We now give the main result of this section.

Theorem 2.4. Under $H_0$, and assuming that $C1 - C6$ holds, as $n \to \infty$,
\[
\text{Prob} \left\{ v_n \left( \hat{\theta}_n^{-d} \hat{\theta}_n^{-d} (0; d) T_r - \xi_n \right) \leq x \right\} \to \exp (-2e^{-x}) , \text{ for } x > 0 ,
\]
where $v_n = (-2 \log a)^{1/2}$ and
(a) If $0 < d < 1/2$,
\[
\xi_n = v_n + v_n^{-1} \left\{ \left( \frac{1}{2} - \frac{1}{d} \right) \log \log a^{-t} + \log \left( \frac{2}{d} \right) - d J_n \right\}
\]
where
\[
0 < \alpha = \lim_{a \to 0} \int_0^\infty e^s \text{Pr} \left\{ \sup_{0 \leq t \leq |a|^{-1}} \frac{\alpha}{t} Y (t) > s \right\} ds < \infty,
\]
and $Y (t)$ is a stationary mean zero gaussian process with covariance structure
\[
\text{Cov} \left( Y (t_1), Y (t_2) \right) = |t_1|^\alpha + |t_2|^\alpha - |t_2 - t_1|^\alpha ,
\]
for some $0 < D < \infty$, where $0 < \alpha \leq 2$ is as given in Proposition 2.2.
(b) If $d = 0$, then
\[
\xi_n = v_n + v_n^{-1} \left\{ \log \left( \frac{W_1 (K)}{\pi^{1/2}} \right) + \frac{1}{2} \log \log a^{-t} \right\},
\]
where
\[
W_1 (K) = \frac{1}{2} (K_+^2 (0) + K_+^2 (1))
\]
if $W_1 (K) > 0$, and otherwise
\[
\xi_n = v_n + v_n^{-1} \log \left( \frac{1}{\pi} \left( \frac{W_2 (K)}{2} \right)^{1/2} \right),
\]
where $W_2 (K) = 2^{-t} \int_0^1 K_+^{(1)} (x)^2 dx$.

Remark 2.1. The uniform kernel $K_+ (x) = 1, 0 < x < 1$ falls under the first case, while the triangular or Hanning-Tukey kernels fall under the second case.
Remark 2.2. The techniques of proof of the result may readily be adapted to prove limit theorems as that of Woodroofe (1967) or Van-Ness and Woodroofe (1967) for the maximum deviation at the points $x_{j\cdot n\log n}$, at $j = 0, 1, \ldots, [a]^{-1}/\log n$, respectively, where $v_n$ and $\xi_n$ are as defined in Theorem 2.4.

Theorem 2.4 indicates that the limiting distribution of $T_r$ corresponds to the type of Gumbel distribution which appears as one of the three possible limit distributions in extreme-value statistics theory, see Theorem 8.13.1 of Bingham et al. (1987).

A desirable and important characteristic of any test is its consistency, that is, the test is consistent. This is con

Theorem 2.6.

Corollary 2.5. Assuming C1 – C6, under $H_0$ in (2.6), as $n \to \infty$,

$$\text{Prob} \left\{ v_n \left( \frac{n}{\bar{n}} - d \right) \leq x \right\} \to \exp \left( -2e^{-\left( x - |\bar{n}\rho| K \right)} \right)$$

for $x > 0$, where $v_n$ and $\xi_n$ are as defined in Theorem 2.4 and $\rho(K) = \max_{\ell=1, \ldots, \bar{n}} \int_0^{\ell/n} K_+ (v) \, dv$.

From Corollary 2.5, one would expect that for fixed alternatives

$H_1 : \exists \delta \in [0, 1]$ such that $r_+ (x^0) = r_- (x^0) + r, \quad |r| > 0$,

we should have

$$\lim_{n \to \infty} \text{Prob} \left\{ v_n \left( \frac{n}{\bar{n}} - d \right) \leq x \right\} = \lim_{n \to \infty} \exp \left( -e^{-\left( x - |\bar{n}\log a|^{1/2} \rho(K) \right)} \right) = 0,$$

that is, the test is consistent. This is confirmed in the next theorem.

Theorem 2.6. Assuming C1 – C6, the test $T_r$ is consistent.

Although Theorem 2.4 gives asymptotic justification to perform the test for $H_0$, we observe that one of the normalization constants needed to achieve the asymptotic distribution depends on $d$. Also, and perhaps more importantly, $\xi_n$ depends, albeit on $d$, on $J_n$, which except in the case for which $\alpha = 1$ or 2, it may be very difficult to compute. In particular, for the former cases, we have that $J_2 = v_n + v_n^{-1} \log \left( \pi^{-1} (D/2)^{1/2} \right)$ and $J_4 = v_n + v_n^{-1} \log \left( (D/\pi)^{1/2} + 2 - 1 \log \log a^{-1} \right)$.

More specifically, in our context, although $d$ can be estimated, we face two potential difficulties when implementing the test. First to obtain $\alpha$, which depends on $K_+$ as well as on $d$, as we can observe from Proposition 2.2. And second, even if the latter were known, $J_n$ has to be obtained by Monte-Carlo simulations, albeit that this implies that it is data dependent. Under these circumstances, bootstrap methods appear to be a sensible way to proceed. Moreover, as was mentioned in the introduction, among the various motivations for the use of bootstrap methods is that they approximate the tails of the finite sample distributions better or more effectively than rival methods such as those based on Edgeworth expansions. In our context, this appears to be the case, apart from the fact that they may have a better order of improvement, see Hall (1991). Because of all these reasons, in the next section we describe a bootstrap scheme for the test $T_r$ given in (2.2).
3. THE BOOTSTRAP APPROACH

Since Efron (1978), bootstrap methods have become an often tool in empirical applications. The basic idea of the bootstrap is, given a stretch of data $Z_T = \{z_t, t = 1, ..., T\}$ say, to treat the data as if it were the true population, and to carry out Monte-Carlo experiments in which pseudo-data is drawn from $Z_T$. Based on the underlying distributional properties of $\{z_t, t = 1, ..., T\}$, different schemes have been adopted and proposed.

In this section, we propose a bootstrap scheme in the frequency domain which will be valid simultaneously for short and long memory dependent data, and thus for the bootstrap of the test given in (2.2). The bootstrap scheme described in STEPS 1 to 7 below is similar to the sieve-bootstrap but instead of bootstrapping in the time domain, we employ frequency domain techniques. The difference between the latter and ours is parallel to that existing when we are interested to estimate the spectral density function by (a) approximating the dependence structure of the data by an $AR(p_n)$ model as in Berk (1974) or by (b) the average periodogram, see Brillinger (1981).

The resampling or bootstrap scheme must satisfy two basic requirements. First, given the sample $Y = \{y_1, ..., y_n\}$, the conditional distribution of the bootstrap statistic, say, $T^*_r$ consistently estimates the distribution of $T_r$ under $H_0$ and local/contiguous alternatives $H_a$. That is, under $H_0 \cup H_a$, $T^*_r \rightarrow^d T_r$ in probability, where $\rightarrow^d$ in probability means convergence in bootstrap distribution according to Giné and Zinn (1989). A second requirement is that under $H_1$, the bootstrap distribution of the test must also converge in bootstrap distribution, although possibly to a different one than under $H_0$. To achieve the first requirement, one key condition is that the resampling algorithm should preserve the correlation structure of the original data $u_t$, whereas the second requirement would be achieved if we were capable to bootstrap from the null hypothesis.

We now describe the bootstrap algorithm. To that end, and for illustration purposes, suppose that $u_t$ were observable and we were interested to bootstrap the distribution of the sample mean

$$\bar{u} = \frac{1}{n} \sum_{t=1}^{n} u_t.$$ 

As was mentioned above, one of the key requirements for the validity of the resampling scheme is that the bootstrap observations, say $u^*_1, ..., u^*_n$, should have (bootstrap) mean zero and preserve the covariance structure of the original data $u_1, ..., u_n$. Depending on the hypothetical structure of the serial dependence of $u_t$, different schemes have been proposed. In a parametric framework, say that $u_t$ follows an $ARMA(p, q)$ model, Kreiss and Franke (1992) proposed a residual-based bootstrap approach. However, because to give a correct specification of the model is not easy, some nonparametric schemes have been proposed. Two popular ones are the Moving Block Bootstrap as proposed by Künsch (1986) or the subsampling scheme as proposed by Politis and Romano (1994). More recently, in a setup similar to ours, Bühlmann (1998) proposed to employ the sieve-bootstrap introduce in Kreiss (1988) and developed in Bühlmann (1997). The bulk of the theoretical work, and in particular of the aforementioned papers, has been done when $d = 0$, that is when the data is short-memory dependent or that $u_t$ satisfies some type of mixing condition. One exception is Lahiri (1992) who employed Moving Block Bootstrap schemes. However, as Bühlmann (1998) indicated, the Moving Block or subsampling methods tend to work less effectively than the latter bootstrap. Similar outcome occurs when bootstrapping the F-test in a (linear) regression model with dependent errors, in the sense that the outcome of the test depends very
much on the length of the block employed when implementing the Moving Block or subsampling bootstraps.

The purpose of this section is to use a different approach, valid simultaneously for both $d = 0$ and $d \neq 0$. The scheme, contrary to the previous aforementioned work, is done in the frequency domain, and in particular we use the "discrete" Crâmer representation of $u_t$. To that end, for a generic sequence $a_t$, denote the "Discrete Fourier Transform" 

$$w_u(\lambda_j) = \frac{1}{n^{1/2}} \sum_{t=1}^{n} a_t e^{-it\lambda_j}. $$

Suppose, for the time being, that $u_t$ is a covariance stationary linear process given by 

$$u_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}; \hspace{1em} b_0 = 1,$$

where $B(z) = \sum_{k=0}^{\infty} b_k z^k$ and $\sum_{k=0}^{\infty} k^2 |b_k| < \infty$. Observe that the latter condition on $b_0$ will imply that $B(z)$ is twice continuously differentiable in $[0, \pi]$. Now, using the identity 

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

where $\lambda_j = (2\pi j)/n$ for integer $j$ and Barlett’s approximation of $w_u(\lambda_j)$, that is $w_u(\lambda_j) \approx B(e^{i\lambda_j}) w_\varepsilon(\lambda_j)$, we obtain that (3.1), e.g. $u_t$, can be approximated by 

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

where "\approx" should be read as "approximately". Observe that we can regard (3.1) as a discrete approximation of Crâmer representation of a stationary process.

Because in this paper we allow for the possibility of long-memory dependence, for instance that $u_t$ follows the fractional difference model 

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

the previous arguments may induce to employ the Barlett’s approximation 

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

and thence the approximation 

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

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 $\lambda_j$ for fixed $j$, indicate that the Barlett’s approximation in (3.2) appears not to be appropriate. Observe that the frequencies $\lambda_j$ for which $j$ is fixed are precisely those which are more relevant in our context. For that reason in the paper we shall replace $(1 - e^{i\lambda_j})^{-d}$ by 

$$\tilde{g}^{1/2}(\lambda_j; d) = \left| \sum_{t=-n+1}^{n-1} \gamma(t; d) e^{-it\lambda_j} \right|^{1/2},$$
where

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

with $\Gamma(\cdot)$ being the gamma function. So, $\tilde{g}(\lambda; d)$ can be regarded as a truncated version of $|1 - e^{i\lambda}|^{-2d}$. That is, with this replacement the approximation given in (3.3), to be used to obtain the bootstrap sample $u^*_1, \ldots, u^*_n$, becomes

$$(3.4) \quad u_t \approx \frac{1}{n^{1/2}} \sum_{j=1}^{n} e^{i(t-s)\lambda} \tilde{g}(\lambda; d) B(e^{i\lambda_t}) w_x(\lambda_j).$$

The right side of (3.4) preserves (asymptotically) the covariance structure of $u_t$. This can be seen after noting that $E(w_x(\lambda_j) w_x(-\lambda_k)) = \sigma^2 T(j = k)$, which implies that

$$E(u_t u_s) \approx \sigma^2 \frac{1}{n} \sum_{j=1}^{n} e^{i(t-s)\lambda} \tilde{g}(\lambda; d) B(e^{i\lambda_t})^2$$

as $\tilde{g}(\lambda; d) B(e^{i\lambda_j})^2 \approx |1 - e^{i\lambda}|^{-2d} |B(e^{i\lambda_j})|^2 = f(\lambda_j)$. Thus, if in the right side of (3.4) $d$ and $B(e^{i\lambda_j})$ were replaced by consistent estimators, say $\hat{d}$ and $\hat{B}(e^{i\lambda_j})$, the problem of how to obtain a bootstrap sample $u^*_t$, $t = 1, \ldots, n$, becomes that of how to perform a valid bootstrap algorithm for the DFT’s $w_x(\lambda_j)$, $j = 1, \ldots, [n/2]$. Note that because $w_x(\lambda_j) = w_{\xi}(\lambda_{j-1})$, where $\xi$ denotes the conjugate of $a$, the resampling of $w_x(\lambda_j)$ has to be done from the first $[n/2]$ frequencies $\lambda_j$.

The previous arguments suggest the bootstrap algorithm to bootstrapping $T_r$ in (2.2), and which is described in the following 7 STEPS.

**STEP 1:** Obtain the centred residuals $\tilde{u}_t = \tilde{u}_t - \bar{u}_t - n^{-1} \sum_{t=1}^{n} \tilde{u}_t$, $t = 1, \ldots, n$, where

$$\tilde{u}_t = \begin{cases} y_t - \hat{\gamma}_+(t), & t = 1, \ldots, \tilde{n} \\ y_t - \frac{1}{2}(\hat{\gamma}_+(t) + \hat{\gamma}_-(t)), & t = \tilde{n} + 1, \ldots, n - \tilde{n} - 1 \\ y_t - \hat{\gamma}_-(t), & t = n - \tilde{n}, \ldots, n, \end{cases}$$

and $\hat{\gamma}_+(t)$ and $\hat{\gamma}_-(t)$ as in (2.1).

Our second step is similar to that of Hidalgo (2002), describing how to obtain $w^*_x(\lambda_j)$, $j = 1, \ldots, [n/2]$.

**STEP 2:** (a) Compute the Discrete Fourier Transform of the residuals $\tilde{u}_t$, denoted $w_{\tilde{u}}(\lambda_j)$, and let $v_{\tilde{u}}(\lambda_j) = w_{\tilde{u}}(\lambda_j)/|w_{\tilde{u}}(\lambda_j)|$, $j = 1, \ldots, [n/2]$. Draw independent bootstrap observations $\eta^*_{j,1}$, $j = 1, \ldots, [n/2]$, from the empirical distribution function of

$$\tilde{v}_{\tilde{u}}(\lambda_j) = \sigma^{-1}_{\tilde{u}} (v_{\tilde{u}}(\lambda_j) - \mathbf{\overline{v}}_{\tilde{u}}), \quad j = 1, \ldots, [n/2]$$

where $\mathbf{\overline{v}}_{\tilde{u}} = [n/2]^{-1} \sum_{j=1}^{[n/2]} v_{\tilde{u}}(\lambda_j)$ and $\sigma^2_{\tilde{u}} = [n/2]^{-1} \sum_{j=1}^{[n/2]} |v_{\tilde{u}}(\lambda_j) - \mathbf{\overline{v}}_{\tilde{u}}|^2$. That is, for all $j = 1, \ldots, [n/2]$, 

$$\text{Pr} \{\eta^*_{j,1} = \tilde{v}_{\tilde{u}}(\lambda_{j-1})\} = [n/2]^{-1}, \quad k = 1, \ldots, [n/2],$$

and write $\eta^*_{n-j,1} = \mathbf{\overline{v}}_{\tilde{u}}$.

(b) Let $\tilde{u}^* = (\tilde{u}^*_1, \tilde{u}^*_2, \ldots, \tilde{u}^*_n)'$ be a random sample with replacement from the empirical distribution of the standardized residuals

$$\bar{u}_t = \tilde{u}_t - \bar{u}_t, \quad \bar{u}^2_t = \frac{1}{n} \sum_{t=1}^{n} \tilde{u}^2_t.$$
and obtain the "Discrete Fourier Transform" of $\tilde{u}$ as 

$$\overline{\eta}_{n-j,2} = \eta_{j,2} = \frac{1}{n^{1/2}} \sum_{t=1}^{n} \tilde{u}_t e^{-it\lambda_j}, \quad j = 1, \ldots, [n/2].$$

Denote by $h_1(\lambda)$ the "square-root" of $h(\lambda)$. That is, the function such that $h_1(\lambda) h_1(\lambda) = h(\lambda)$.

**STEP 3:** Let $\tilde{d}$ be an estimator of $d$, say Robinson's (1995b) GSE, defined as

$$\tilde{d} = \arg \min_{\psi \in (\Delta_1, \Delta_2)} \tilde{R}(\psi),$$

where $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$ and

$$\tilde{R}(\psi) = \log \tilde{H}(\psi) - 2\psi \sum_{j=1}^{m} \log \lambda_j$$

with

$$\tilde{H}(\psi) = \frac{1}{n} \sum_{j=1}^{m} \lambda_j^{2\psi} I_{\tilde{u}_n}(\lambda_j)$$

for integer $m \in [1, n/2)$. Let

$$\hat{h}(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^{m} \left| 1 - e^{-i(\lambda+\lambda_j)} \right|^{2\tilde{d}} I_{\tilde{u}_n} (\lambda + \lambda_j)$$

where $I_{\tilde{u}_n}(\lambda) = (2\pi)^{-1} |w_\lambda(\lambda)|$ is the periodogram of $\tilde{u}_t$, with $m^{-1} + mn^{-1} \rightarrow 0$ and $M = [n/4m]$. Then, our estimator of $h_1(\lambda)$ is given by

$$\hat{h}_1(\lambda) = \exp \left\{ \sum_{r=1}^{M} \tilde{c}_r e^{-ir\lambda} \right\},$$

where

$$\tilde{c}_r = \frac{1}{\pi} \int_{0}^{\pi} \log \left( \tilde{h}(\lambda) \right) \cos (r\lambda) d\lambda; \quad r = 0, \ldots, M.$$

The estimator of $h_1(\lambda)$, $\hat{h}_1(\lambda)$, comes from the so-called canonical spectral decomposition of the spectral density function $h(\lambda)$, see Brillinger (1981, pp. 78 – 79).

**Remark 3.1.** As in Robinson (1997), if $d > 0$, the estimator of $d$ given in STEP 3 could have been computed using the observations $y_t$ instead of the residuals $\tilde{u}_t$ given in STEP 1.

**Remark 3.2.** Although in general $g(\lambda) \neq |1 - e^{-i\lambda}|^{-2d}$, C3 implies that this does not cause any problem. The reason being because C3 will imply that the function $h^+(\lambda) = g(\lambda) \left| 1 - e^{-i\lambda} \right|^{2d} h(\lambda)$ would satisfy the same conditions as $h(\lambda)$ in C3.

**STEP 4:** For $c = 1, 2$, compute

$$u_{t,c}^* = \left( \frac{2\pi}{n} \right)^{1/2} \sum_{j=1}^{n} e^{it\lambda_j} \sum_{l=-n+1}^{n-1} \tilde{\gamma}(\ell; \tilde{d}) e^{-it\lambda_j} \left| \hat{h}_1(e^{-i\lambda_j}) \eta_{j,c}^* \right|^{1/2}, \quad t = 1, \ldots, n.$$

**STEP 5:** Compute

$$\tilde{\gamma}(t) = \begin{cases} \tilde{\gamma}_+(t), & t = 1, \ldots, \tilde{n} \\ \frac{1}{2} (\tilde{\gamma}_+(t) + \tilde{\gamma}_-(t)), & t = \tilde{n} + 1, \ldots, n - \tilde{n} - 1 \\ \tilde{\gamma}_-(t), & t = n - \tilde{n}, \ldots, n. \end{cases}$$
and, for $c = 1, 2$,

$$y^*_{t,c} = \tilde{r}_t(t) + 2\pi e^{\tilde{\xi}_0} u^*_{t,c}, \quad t = 1, ..., n,$$

where $\tilde{r}_+(t)$ and $\tilde{r}_-(t)$ are defined as in STEP 1.

Note that by the well-known Kolmogorov formula, $2\pi e^{\tilde{\xi}_0}$ is an estimate of the variance of $\varepsilon_t$, that is the one step prediction error.

**STEP 6**: For $c = 1, 2$, compute $\tilde{r}^*_{t,c}(q)$ and $\tilde{r}^*_{t,-c}(q)$, $q = \hat{n} + 1, ..., n - \hat{n} - 1$, as in (2.1) but with $y_t$ replaced by $y^*_{t,c}$ and the same bandwidth parameter $a$ employed in STEPS 1 and 5 to compute $\tilde{r}_+(t)$ and $\tilde{r}_-(t)$ or $\tilde{r}(t)$.

Our final step is:

**STEP 7**: Compute the bootstrap test as

$$T^*_{r,c} = \sup_{\hat{n} < q < n - \hat{n}} |\tilde{r}^*_{r,c}(q) - \tilde{r}^*_{r,-c}(q)|, \quad c = 1, 2.$$

Some comments on STEPS 5 and 6 are needed. First, we observe that the same bandwidth parameter $a$ can be employed all throughout, that is to estimate the regression model using the observed data as well as to obtain the bootstrap observations $y^*_{t,c}$ and the bootstrap estimate of the model. This is in clear contrast to usual results when bootstrapping kernel regression models where the bandwidth choice to estimate $r_+(q)$ and $r_-(q)$ was the optimal bandwidth $a$ in terms of that which minimizes the MSE, since in this case the estimates will have a non negligible bias. If the latter were the case, the bootstrap approach should be able to estimate in addition the asymptotic bias as well. This would be achieved if in STEP 5 an oversmoothed estimate of $r_+(q)$ and $r_-(q)$ were employed. See for instance Härdle and Marron (1991) for an explanation. That is, STEP 5 would become

**STEP 5’**: Compute

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

and, for $c = 1, 2$,

$$y^*_{t,c} = \tilde{r}_t(t) + 2\pi e^{\tilde{\xi}_0} u^*_{t,c}, \quad t = 1, ..., n,$$

where $\tilde{r}_+(t)$ and $\tilde{r}_-(t)$ are defined as in (2.1) but with a bandwidth parameter $\hat{a}$ which converges to zero slower than $a$, the bandwidth employed to obtain $\tilde{r}_+(t)$ and $\tilde{r}_-(t)$ in STEP 1.

Observe that this could be the case if the bandwidth parameter in STEP 1 were chosen by a cross-validation criterion. On the other hand, if in STEP 1 the bandwidth $a$ had been chosen suboptimally, that is in such a way for which the bias would converge faster to zero than the standard deviation of the estimate, as was the case in Bickel and Rosenblatt (1973), then the same bandwidth could have been used in STEP 1 and STEPS 5 and 6. However, in our context this is not needed as Proposition 2.3 shows, since even with optimal choice of $a$, the asymptotic distribution of $r_+(q)$ and $r_-(q)$ will not generate an asymptotic bias term.

Let us introduce the following condition on the smoothing parameter $m$ in STEP 3.

**C7**: As $n \to \infty$,

$$\log^4 n \left\{ \left( \frac{m}{a} \right)^{1.5} + \frac{\log^2 m}{m a} + \frac{1}{m a} + a + a^4 n^{-1.5 / 2} m^{2 d} + \frac{m^{2 d}}{a^{2 n^{1 + 2 d}}} \right\} \to 0.$$

This condition is needed to obtain the consistency of $\hat{d}$ as in Robinson (1997) and it is identical to his Assumption A11. Also, the conditions on $m$ given in C7
guarantee that the estimator of \( h_1 (e^{i \lambda}) \), \( \hat{h}_1 (e^{i \lambda}) \), is consistent, see Hidalgo and Yajima (2002).

**Theorem 3.1.** Assuming C1 – C7, under \( H_0 \cup H_1 \), as \( n \to \infty \)

\[
\text{Prob} \left\{ v_n \left( \tilde{n}^{1/2} \rho^{-1/2} \left( \theta; \tilde{d} \right) T_{\tau, \epsilon} \xi_n \right) \leq x \right\} \overset{P}{\to} \exp \left( -2e^{-x} \right), \quad c = 1, 2, \text{ for } x > 0,
\]

where \( v_n \) and \( \xi_n \) are as defined in Theorem 2.4.

4. PROOFS OF THE MAIN RESULTS

4.1. Proof of Proposition 2.1. By definition, \( \tilde{n}^{1/2} \text{Cov} (\hat{r}_+ (q_1), \hat{r}_- (q_1), \hat{r}_+ (q_2) - \hat{r}_- (q_2)) \)

is

\[
\tilde{n}^{1/2} \text{Cov} (\hat{r}_+ (q_1), \hat{r}_+ (q_2)) + \tilde{n}^{1/2} \text{Cov} (\hat{r}_- (q_1), \hat{r}_- (q_2))
\]

\[
- \tilde{n}^{1/2} \text{Cov} (\hat{r}_+ (q_1), \hat{r}_- (q_2)) - \tilde{n}^{1/2} \text{Cov} (\hat{r}_- (q_1), \hat{r}_+ (q_2)).
\]

We begin with part (a), that is when \( 0 < d < \frac{1}{2} \). Noting that \( E (u_t u_0) = \int_{- \pi}^{\pi} e^{it \lambda} f (\lambda) d \lambda \), the first term of (4.1) is

\[
\frac{1}{\tilde{n}^{1/2} + 2d} \sum_{i=1}^{q_1 + q_2 + \tilde{n}} E (u_t u_s) K_{+, t - q_1} K_{+, s - q_2}
\]

\[
= \frac{1}{\tilde{n}^{1/2} + 2d} \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} E (u_t + q_1 u_{s + q_1}) K_{+, t} K_{+, s + q_1 - q_2}
\]

\[
= \frac{1}{\tilde{n}^{1/2} + 2d} \int_{- \pi}^{\pi} f (\lambda) \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} K_{+, t} K_{+, s + q_1 - q_2} e^{i(t-s) \lambda} d \lambda
\]

by stationarity of the errors \( u_t \) and where we have abbreviated \( K_{+(t/\tilde{n})} \) by \( K_{+, t} \).

Proceeding as in Robinson (1997, p.2061), the difference between the right side of the last displayed equation and

\[
\frac{H}{\tilde{n}^{1/2} + 2d} \int_{- \pi}^{\pi} g (\lambda) \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} K_{+, t} K_{+, s + q_1 - q_2} e^{i(t-s) \lambda} d \lambda
\]

is bounded in absolute value, because the triangle and Cauchy-Schwarz inequalities, by

\[
\sup_{|\lambda| < \epsilon} \left| h (\lambda) - \frac{H}{\tilde{n}^{1/2} + 2d} \int_{- \pi}^{\pi} g (\lambda) \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} K_{+, t} e^{i \lambda} \right|^2 d \lambda
\]

\[
+ \frac{2H}{\tilde{n}^{1/2} + 2d} \int_{- \pi}^{\pi} g (\lambda) \left| \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} K_{+, t} e^{i \lambda} \right|^2 d \lambda
\]

\[
+ \frac{2}{\tilde{n}^{1/2} + 2d} \int_{- \pi}^{\pi} f (\lambda) \left| \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} K_{+, t} e^{i \lambda} \right|^2 d \lambda,
\]

for any arbitrarily small \( \epsilon > 0 \) and noting that

\[
\left| \sum_{i=1}^{q_1 + q_2 - q_1 + q_1} K_{+, s + q_1 - q_2} e^{i \lambda} \right|^2 = \left| \sum_{s=1}^{q_1 + q_2 - q_1 + q_1} K_{+, s} e^{i \lambda} \right|^2.
\]
The second term of (4.3) is $o(1)$ as we now show. Summation by parts and bounded differentiability of $K_+$ imply that
\[
\left| \sum_{t=1}^{\tilde{n}} K_{+,t} e^{i \lambda t} \right| \leq \sum_{t=1}^{\tilde{n}-1} |K_{+,t} - K_{+,t+1}| + |K_{+,\tilde{n}}| \left| \sum_{t=1}^{\tilde{n}} e^{i \lambda t} \right| \leq D \lambda^{-1},
\]
so that the second term of (4.3) is bounded by
\[
D \varepsilon^{-2} \tilde{n}^{-1 - 2d} \int_{\varepsilon}^{\infty} g(\lambda) d\lambda = o(1).
\]
The third term follows similarly. Finally, the first term of (4.3) is $o(1)$ due to the continuity of $h(\lambda)$ by C3 and because $\varepsilon$ is arbitrary. Next, we show that we can replace (4.2) by
\[
(4.4) \quad \frac{H \theta(d)}{n_1^{1+2d}} \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} \sum_{\bar{t}s} \sum_{n=2q_1 - 1}^{q_1} |t - s|^{2d-1} K_{+,\bar{t}s} K_{+,s+q_1 - q_2},
\]
where $|q|_c = \max \{1, |q|\}$.

Using the convention that $\sum_{t=c}^{e} \equiv 0$ if $e < c$, the triangle inequality implies that the difference between (4.2) and (4.4) is in absolute value bounded by
\[
\frac{D}{n_1^{1+2d}} \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} \| \gamma_{t-s} \| \left| \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} \| \gamma_{t-s} \| \right| |K_{+,t} K_{+,s+q_1 - q_2}|
\]
\[
+ \frac{D}{n_1^{1+2d}} \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} \| \gamma_{t-s} - \theta(d) \| |t - s|^{2d-1} |K_{+,\bar{t}s} K_{+,s+q_1 - q_2}|
\]
\[
\leq D \tilde{n}^{2d}\alpha^{-1} + \frac{D}{n_1^{1+2d}} \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} \| \xi_{t-s} \| |t - s|^{2d-1} |K_{+,\bar{t}s} K_{+,s+q_1 - q_2}|
\]
\[
= o(1)
\]
choosing $\alpha < 1$ and by Toeplitz Lemma because $\xi_t = o(1)$ as $t \to \infty$ by C3. So, to complete the proof of part (a), it suffices to show that the difference between (4.4) and the right side of (2.3) converges to zero. This difference, except the constant $H \theta(d)$, is
\[
\frac{1}{n_1^{1+2d}} \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} |t - s|^{2d-1} K_{+,\bar{t}s} K_{+,s+q_1 - q_2}
\]
\[
- \int_0^1 \int_b^c |v - w|^{2d-1} K_+ (v) K_+ (w-b) dv dw
\]
\[
= \sum_{t=1}^{\tilde{n}} \sum_{s=2q_1 - 1}^{q_1} \int_0^t \int_b^c |t - s|^{2d-1} K_{+,\bar{t}s} K_{+,s+q_1 - q_2} - |v - w|^{2d-1} K_+ (v) K_+ (w-b) dv dw
\]
\[
+ \frac{1}{n_1^{1+2d}} \sum_{t=1}^{\tilde{n}} K_{+,\bar{t}s} - \int_0^1 \int_b^c |v - w|^{2d-1} K_+ (v) K_+ (w-b) dv dw
\]
\[
= O\left( \tilde{n}^{-2d} \right) + D \frac{1}{n_2^{2d}} \sum_{t=1}^{\tilde{n}-1} |t - s|^{2d-1} = O\left( \tilde{n}^{-2d} + \tilde{n}^{2d-1} \right),
\]
because $\tilde{n}^{-1} \sum_{t=1}^{\tilde{n}} K_{+,\bar{t}s} = O(1)$ and $|K_+(v)| \leq D$.

Proceeding similarly as with the first term of (4.1), its second term is
\[
H \theta(d) \int_{-1}^{0} \int_{b-1}^{b} |v - w|^{2d-1} K_- (v) K_- (w-b) dv dw + o(1)
\]
and likewise the third and fourth terms of (4.1) are

\[ H\theta(d) \int_0^1 \int_{b-1}^b |v-w|^{2d-1} K_+(v) K_-(w-b) \, dv \, dw + o(1) \]

\[ H\theta(d) \int_{-1}^0 \int_{b-1}^b |v-w|^{2d-1} K_-(v) K_+(w-b) \, dv \, dw + o(1), \]

respectively, which completes the proof for the case \( 0 < d < \frac{1}{2} \).

Next, part (b). First, proceeding as with \( d > 0 \), it is easy to see that the first term of (4.1) can be replaced by

\[ \frac{H}{2\pi n} \int_{-\pi}^{\pi} \sum_{t=1}^{\tilde{n}} \sum_{s=\tilde{n}+q_1+1}^{\tilde{n}+q_2} K_{+,t} K_{+,s+q_1-q_2} e^{i(t-s)\lambda} d\lambda = \frac{H}{\tilde{n}} \sum_{t=q_2-q_1+1}^{\tilde{n}} K_{+,t} K_{+,t+q_1-q_2}. \]

From here the result follows by standard kernel arguments, so it is omitted.

Finally part (c), that is when \( d < 0 \). The proof proceeds using the same arguments as with part (a) and those in (3.15) of Robinson (1997).

4.2. Proof of Proposition 2.2. We shall begin with the case \( d > 0 \). Denoting by \( \text{Cov}(b) \) expression (4.1), we have from Proposition 2.1 that

\[ \text{Cov}(b) = H\theta(d) \int_0^1 \int_b^{b+1} |v-w|^{2d-1} K_+(v) K_+(w-b) \, dv \, dw - H\theta(d) \int_{-1}^0 \int_b^{b+1} |v-w|^{2d-1} K_-(v) K_+(w-b) \, dv \, dw \]

\[ + H\theta(d) \int_{-1}^0 \int_{b-1}^b |v-w|^{2d-1} K_-(v) K_-(w-b) \, dv \, dw \]

(4.5)

\[ - H\theta(d) \int_0^1 \int_{b-1}^b |v-w|^{2d-1} K_+(v) K_-(w-b) \, dv \, dw + o(1). \]

After noting that for \( \ell > 0 \)

\[ \ell^{2d-1} = \frac{2}{\pi} \Gamma(2d) \cos(\ell\pi) \int_0^\infty \lambda^{-2d} \cos(\ell\lambda) \, d\lambda \]

(4.6)

for \( 0 < d < \frac{1}{2} \), for finite \( b \) we have that the first term on the right of (4.5) is

\[ \int_{-\infty}^{\infty} |\lambda|^{-2d} \cos(|\lambda| b) \left| \int_0^1 K_+(v) e^{i\lambda v} \, dv \right|^2 d\lambda \]

(4.7)
as we now show. Replacing (4.6) into the first term on the right of (4.5), we have that the latter term is

$$\int_{-\infty}^{\infty} |\lambda|^{-2d} \int_{-1}^{1} (e^{i|v|\lambda} + e^{-i|v|\lambda}) d\lambda) K_+(v) K_+(w-b) dv dw$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \int_{0}^{b+1} \left( e^{i(v-w)\lambda} + e^{-i(v-w)\lambda} \right) K_+(v) K_+(w-b) dv dw d\lambda$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \left( \int_{0}^{1} \left( e^{i(v-w)\lambda} e^{-i\lambda} + e^{-i(v-w)\lambda} e^{i\lambda} \right) K_+(v) K_+(w) dv \right) d\lambda$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \cos(|\lambda| b) \left( \int_{0}^{1} K_+(v) e^{i\lambda v} dv \right)^2 d\lambda,$$

that is (4.7), where in the first equality we have used the reflection formula of the gamma function and in the second equality the change of variables $w - b = w'$.

Likewise the third term on the right of (4.5) is

$$\int_{-\infty}^{\infty} |\lambda|^{-2d} \cos(|\lambda| b) \left( \int_{0}^{1} K_+(v) e^{i\lambda v} dv \right)^2 d\lambda$$

(4.8)

because by C5, $K_+(v) = K_-(v)$.

Now, the second term on the right of (4.5), proceeding similarly as with the first term, e.g. using (4.6), is

$$\int_{-1}^{1} \int_{0}^{b+1} \left( \int_{0}^{\infty} \lambda^{-2d} \left( e^{i|v|\lambda} + e^{-i|v|\lambda} \right) d\lambda \right) K_-(v) K_+(w-b) dv dw$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \left( \int_{0}^{1} \left( e^{i(v-w)\lambda} + e^{-i(v-w)\lambda} \right) K_-(v) K_+(w-b) dv \right) d\lambda$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \left( \int_{0}^{1} \left( e^{-i(v+w)\lambda} e^{-i\lambda} + e^{i(v+w)\lambda} e^{i\lambda} \right) K_+(v) K_+(w) dv \right) d\lambda$$

where we have employed that C5 implies that $K_+(v) = K_-(v)$ and in the second equality the change of variables $w - b = w'$ and $v = -v'$. By the same arguments, the fourth term on the right of (4.5) is

$$\frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \int_{0}^{1} \left( e^{i(v+w)\lambda} e^{-i\lambda} + e^{-i(v+w)\lambda} e^{i\lambda} \right) K_+(v) K_+(w) dv dw d\lambda,$$

so that the contribution due to the second and fourth terms on the right of (4.5) is

$$-\frac{1}{2} \int_{-\infty}^{\infty} |\lambda|^{-2d} \int_{0}^{1} \left( e^{-i(v+w)\lambda} + e^{i(v+w)\lambda} \right) (e^{-i\lambda} + e^{i\lambda}) K_+(v) K_+(w) dv dw d\lambda$$

$$= -\int_{-\infty}^{\infty} |\lambda|^{-2d} \cos(|\lambda| b) \int_{0}^{1} \left( e^{-i(v+w)\lambda} + e^{i(v+w)\lambda} \right) K_+(v) K_+(w) dv dw d\lambda.$$

Hence, gathering (4.7), (4.8) and the right side of the last displayed equation, we conclude that

$$Cov (b) = \int_{-\infty}^{\infty} |\lambda|^{-2d} \cos(|\lambda| b) \left( \int_{0}^{1} K_+(v) e^{i\lambda v} - e^{-i\lambda v} dv \right)^2 d\lambda,$$
which implies that

$$\text{Cov} (b) - \text{Cov} (0) = \int_{-\infty}^{\infty} |\lambda|^{-2d} \{ \cos (|\lambda| b) - 1 \} \left| \int_{0}^{1} K_+ (v) (e^{i\lambda v} - e^{-i\lambda v}) \, dv \right|^2 \, d\lambda.$$  

Clearly, because \( \left| \int_{0}^{1} K_+ (v) (e^{i\lambda v} - e^{-i\lambda v}) \, dv \right|^2 \) is nonnegative, different than a constant and \( |\cos x| \leq 1 \), we obtain that \( \text{Cov} (b) - \text{Cov} (0) < 0 \) for \( |b| > 0 \).

Moreover, as \( b \to 0 \), that is for points \( x_{q_1} \) and \( x_{q_2} \) such that \( |q_1 - q_2| = o(n) \), \( |\text{Cov} (b) - \text{Cov} (0)| = O (b^\alpha) + o(b^\alpha) \), for some \( 0 < \alpha \leq 2 \), so that we can conclude that

$$\text{Corr} (b) = \frac{\text{Cov} (b)}{\text{Cov} (0)} = 1 - Db^\alpha + o(b^\alpha); \quad b \to 0$$

for some \( 0 < \alpha \leq 2 \). On the other hand, as \( b \to \infty \), it is easily shown that \( \text{Corr} (b) \log b = o(1) \) from the definition of \( \text{Cov} (b) \) given in (4.5) because \( 0 < d < 1/2 \). Note that the latter agrees with Robinson’s (1997) Proposition 2, since for \( |q_2 - q_1| > zn \), with \( z > 0 \), we have that \( b := b(q_1, q_2) \geq za^{-1} \) and \( a \to 0 \) by C6. This concludes the proof for \( d > 0 \).

Next, when \( d = 0 \). In this case

$$\text{Cov} (b) - \text{Cov} (0) = \frac{1}{2} \int_{0}^{1} K_+ (w) \{ K_+ (w - b) - K_+ (w) \} \, dw$$

so that

$$\text{Corr} (b) = \frac{\int_{0}^{1} K_+ (w) \{ K_+ (w - b) - K_+ (w) \} \, dw}{\int_{0}^{1} K_+^2 (w) \, dw}.$$  

However, as \( b \to 0 \), by Bickel and Rosenblatt’s (1973) Theorem B.1 and C3, the right side of the last displayed equation satisfies

$$\text{Corr} (b) = \begin{cases} \frac{1}{2} K_+ (1) + o(|b|) & \text{if} \ K_+ (1) \neq 0 \\ 1 - \frac{K_+^2}{2} & \text{otherwise}. \end{cases}$$

(Recall that by C5, \( K_+ (0) = 0 \).)

Finally, we need to examine the behaviour when \( d < 0 \), which follows by the same arguments as when \( 0 < d < \frac{1}{2} \), so it is omitted.

### 4.3. Proof of Proposition 2.3.

Proceeding in the same way of Robinson’s (1997), Theorem 1 and Lemma 1, it suffices to show that

$$(4.9) \quad E \left( \hat{\tau}_+ (q) - \hat{\tau}_- (q) \right) = \begin{cases} o(a^\tau) & \text{if} \ 0 < \tau \leq 1 \\ o \left( a^\tau + n^{-1} \right) & \text{if} \ 1 < \tau \leq 2. \end{cases}$$

Observe that by uniform integrability of \( u_i^2 \) and that Propositions 2.1 and 2.2 imply that the covariance of \( \hat{\tau}_+ (q) - \hat{\tau}_- (q) \) at two points \( q_1 \) and \( q_2 \) such that \( |q_1 - q_2| \geq nz > 0 \) converges to zero, the covariance of the asymptotic distribution of the estimators is zero.

Now under \( H_0 \), by standard kernel manipulations,

$$E \left( \hat{\tau}_+ (q) - \tau (q) \right) = \int_{0}^{1} K_+ (v) (r (av + q/n) - r (q/n)) \, dv$$

$$= \int_{0}^{1} K_+ (v) Q (av + q/n) \, dv + \begin{cases} o(a^\tau) & \text{if} \ 0 < \tau \leq 1 \\ o \left( a^\tau + n^{-1} \right) & \text{if} \ 1 < \tau \leq 2. \end{cases}$$
by C4. Similarly,

\[ E(\tilde{\tau}_-(q) - r(q)) = \int_1^0 K_-(v)Q(av + q/n)\,dv + \begin{cases} o(a^\tau) & \text{if } 0 < \tau \leq 1 \\ o(a^\tau + n^{-1}) & \text{if } 1 < \tau \leq 2 \end{cases}, \]

so that, after an obvious change of variables, the left side of (4.9) is

\[ \int_0^1 K_+(v)\{Q(av + q/n) - Q(-av + q/n)\}\,dv + \begin{cases} o(a^\tau) & \text{if } 0 < \tau \leq 1 \\ o(a^\tau + n^{-1}) & \text{if } 1 < \tau \leq 2 \end{cases} = \begin{cases} o(a^\tau) & \text{if } 0 < \tau \leq 1 \\ o(a^\tau + n^{-1}) & \text{if } 1 < \tau \leq 2. \end{cases} \]

by continuity of \( Q \) and that the first moment of \( K_+(x) \) is zero by C5. \( \square \)

4.4. **Proof of Theorem 2.4.** Noting that because the asymptotic distributions of \( \max_q \) and \( \min_q \) are independent, see for instance Theorem 1.8.3 of Leadbetter (1983), and that the (asymptotic) distributions of \( \sup_i X_i \) and \( \inf_i -X_i \) are the same, it suffices to show that

(4.10)

\[ \Pr \left\{ \varepsilon_n \left( \sup_{\tilde{n} < q < \tilde{n} - \tilde{n}} \tilde{n}^{1-d} \rho^{1-q} (0; d) (\tilde{\tau}_+(q) - \tilde{\tau}_-(q)) - \xi_n \right) \leq x \right\} \rightarrow \exp(-x^2), \quad x > 0. \]

To that end, we will show that \( \tilde{n}^{1-d} (\tilde{\tau}_+(q) - \tilde{\tau}_-(q)) \) converges to a Gaussian process \( G(u) \) in \( \mathbb{D}[0, \infty) \), and whose correlation structure satisfies conditions (v) and (vi) of Theorem A1 of Bickel and Rosenblatt (1973) for some \( \alpha > 0 \). See also Pickands's (1969) equations (1.2) and (2.1). Hence, the limiting distribution in (4.10) holds by Theorem 1 of Bickel and Rosenblatt (1973), see also Theorem 12.3.5 of Leadbetter et al. (1983).

From Proposition 2.3 and Cr\'amer-Wold device, we obtain that the finite dimensional distributions converge to those of a Gaussian process \( G(u) \). Next, by Proposition 2.2, the correlation structure of \( G(u) \) is

(4.11)

\[ \text{Corr}(b) = \frac{\text{Cov}(b)}{\text{Cov}(0)} = 1 - Db^\alpha + o(b^\alpha); \quad b \to 0, \]

for some \( 0 < \alpha \leq 2 \), whereas \( \text{Corr}(b) = o(\log^{-1} b) \) as \( b \to \infty \) as it is easily shown from the definition of \( \text{Cov}(b) \) in (4.5) and because \( d < 1/2 \). So, \( \text{Corr}(b) \) satisfies the conditions in Bickel and Rosenblatt (1973) or Pickands (1969).

So, to complete the proof we need to show that the process \( \tilde{n}^{1-d} (\tilde{\tau}_+(q) - \tilde{\tau}_-(q)) \) is tight. To that end, denote

\[ X_{+n}(\tilde{q}) = \frac{1}{\tilde{n}^{1-d}} \sum_{t=1}^{\tilde{n}} u_t \tilde{K}_+ \left( \frac{t}{\tilde{n}} - \tilde{q} \right), \quad \tilde{q} = 1/\tilde{n}, 2/\tilde{n}, ..., [a]^{-1}, \]

where we have abbreviated \( K_+ (\tilde{n}^{-1}t - \tilde{q}) - EK_+ (\tilde{n}^{-1}t - \tilde{q}) \) by \( \tilde{K}_+(\tilde{n}^{-1}t - \tilde{q}) \). So, \( X_{+n}(\tilde{q}) \) is a process in \( \mathbb{D} [0, [a]^{-1}] \) equipped with Skorohod’s metric, where we extend \( \mathbb{D} [0, [a]^{-1}] \) to \( \mathbb{D}[0, \infty) \) by writing \( X_{+n}(\infty) = X_{+n}([a]^{-1}) \). By Pollard (1981, Ch.V), we need to show tightness in \( \mathbb{D}[0, D] \) for any finite \( D > 0 \). Note that

\[ X_{+n}(\tilde{q}) = X_{+n}(\tilde{q}) + X_{-n}(\tilde{q}) \]

where

\[ X_{-n}(\tilde{q}) = \frac{1}{\tilde{n}^{1-d}} \sum_{t=1}^{\tilde{n}} u_t \tilde{K}_- \left( \frac{t}{\tilde{n}} - \tilde{q} \right), \quad \tilde{q} = 1/\tilde{n}, 2/\tilde{n}, ..., [a]^{-1}. \]
Observe that by Proposition 2.2, the process $X_{+,n}(\bar{q})$ has independent increments and is stationary, that is for $\bar{q} \in [a_1, b_1]$ and $\tilde{q} \in [a_2, b_2]$, $X_{+,n}(\tilde{q})$ is (asymptotically) independent with the same finite dimensional distributions.

By Billingsley’s (1968) Theorem 15.6, it suffices to show the moment condition

$$E \left( |X_{+,n}(\bar{q}) - X_{+,n}(\bar{q})|^\beta |X_{+,n}(\bar{q}) - X_{+,n}(\tilde{q})|^\beta \right) \leq D |\bar{q} - \bar{q}|^{(1+\delta)/2} |\bar{q} - \tilde{q}|^{(1+\delta)/2}$$

for some $\delta, \beta > 0$ and where $0 < \bar{q}_1 < \tilde{q} < \bar{q}_2 \leq D$. Observe that we can consider only the situation for which $\bar{q}_1 < \tilde{q}_2 - \bar{q}_1$, since otherwise we have that

$$(X_{+,n}(\bar{q}_2) - X_{+,n}(\bar{q}_1)) (X_{+,n}(\tilde{q}) - X_{+,n}(\bar{q}_1))$$

is zero. This is the case because either $\bar{q}_1$ and $\tilde{q}$ lie in the same interval $j, j/n$ or else $\bar{q}$ and $\tilde{q}$ do; in either of these cases $X_{+,n}(\bar{q}_2) - X_{+,n}(\bar{q}_1)) (X_{+,n}(\tilde{q}) - X_{+,n}(\bar{q}_1))$ is zero. Because for any $0 < a < b < c \leq D$, $|c - b| |b - a| \leq |c - a|^2$, by Cauchy-Schwarz inequality, the last displayed inequality holds true if

$$E |X_{+,n}(\bar{q}_2) - X_{+,n}(\bar{q}_1)|^{2\beta} \leq D |\bar{q}_2 - \bar{q}_1|^{1+\delta}.$$ 

It suffices to consider $|\bar{q}_2 - \bar{q}_1| < 1$, the case $|\bar{q}_2 - \bar{q}_1| \geq 1$ is trivial since the left side of (4.12) is bounded.

Let us consider $d > 0$ first. By definition, $X_{+,n}(\bar{q}_2) - X_{+,n}(\bar{q}_1)$ is

$$\frac{1}{n^{d+\beta}} \sum_{t=\bar{q}_2-q_1+1}^{\bar{q}_1} u_{t+\bar{q}_2} \bar{K}_{+,t} + \frac{1}{n^{d+\beta}} \sum_{t=1}^{\bar{q}_1-q_2+1} u_{t+\bar{q}_2} \left( \bar{K}_{+,t} - \bar{K}_{+,t+q_1-q_2} \right)$$

(4.13)

Choose $\beta = 1$ in (4.12). Because $\bar{n}\bar{q} = q$ and

$$\frac{1}{n^{d+2\beta}} \sum_{s=1}^{q_2-q_1} |t - s|^{2d-1} \leq D |\bar{q}_2 - \bar{q}_1|^{1+2d},$$

the last term of (4.13) satisfies the inequality (4.12). Similarly, we obtain that

$$E \left( \frac{1}{n^{d+\beta}} \sum_{t=\bar{q}_2-q_1+1}^{\bar{q}_1} u_{t+\bar{q}_2} \bar{K}_{+,t} \right)^2 \leq D |\bar{q}_2 - \bar{q}_1| (1 - (1 - (\bar{q}_2 - \bar{q}_1)))^{2d}$$

because $0 < \bar{q}_2 - \bar{q}_1 < 1$. Finally by continuous differentiability of $K_+ (u)$ for $u \in (0, 1)$, the middle term in (4.13) has a second moment bounded by

$$D (\bar{q}_2 - \bar{q}_1)^2 \frac{1}{n^{d+2\beta}} \sum_{s=1}^{q_2-q_1} |t - s|^{2d-1} \leq D |\bar{q}_2 - \bar{q}_1|^{1+2d}$$

because $0 < d < 1/2$. So, (4.12) holds true choosing $\delta = 2d$ and hence $X_{+,n}(\bar{q})$ is tight. By identical arguments $X_{-,n}(\tilde{q})$ is also tight, which implies that the process $\bar{n}^{-d}(\bar{r}_+ (q) - \bar{r}_- (q))$ is tight, which concludes the proof of the theorem for $0 < d < 1/2$.

Next, when $d = 0$. From the preceding arguments, it is obvious that we only need to show the tightness condition of

$$X_{+,n}(\bar{q}) = \frac{1}{n^{1/2}} \sum_{t=1}^q u_t K_+ \left( \frac{t}{n} \right).$$
In this case the proof is even simpler. Choosing $\beta = 2$ in (4.12), because $u_t$ is a weakly dependent sequence and by well known standard kernel arguments, we have that, say

$$E \left| \frac{1}{\tilde{n}^{1/2}} \sum_{t=q_1+1}^{q_2} u_t K_+ \left( \frac{t}{\tilde{n}} \right) \right|^4 \leq D |\tilde{q}_2 - \tilde{q}_1|^2,$$

so that

$$E |X_{+,\tilde{n}} (\tilde{q}_2) - X_{+,\tilde{n}} (\tilde{q}_1)|^4 \leq D |\tilde{q}_2 - \tilde{q}_1|^2.$$ 

Proceeding similarly, $X_{-,\tilde{n}} (\tilde{q}) = \tilde{n}^{-1/2} \sum_{t=1}^{\tilde{q}} u_t K_- (-\frac{t}{\tilde{n}})$ is also tight, which concludes the proof.

4.5. **Proof of Corollary 2.5.** From the proof of Theorem 2.2, we only need to show that

$$\sup_{\tilde{n} < q < n - \tilde{n}} \tilde{n}^{\frac{1}{2} - d} |E (\widehat{r}_+ (q) - \widehat{r}_- (q))| \rightarrow r \rho (K).$$

But this is the case because by standard kernel manipulations and that by $C6 K_+ (x) = K_- (-x)$, under $H_a$, we obtain that

$$E (\widehat{r}_+ (q) - \widehat{r}_- (q)) = \frac{r}{\tilde{n}^{\frac{1}{2} - d}} \frac{1}{\tilde{n}^{1/2}} \sum_{t=|q-q_0|}^{\tilde{n}} K_+ \left( \frac{t}{\tilde{n}} \right)$$

$$= \frac{r}{\tilde{n}^{\frac{1}{2} - d}} \int_{|q-q_0|/\tilde{n}}^{1} K_+ (x) dx (1 + \mathcal{O} (1)).$$

From here the conclusion is standard because $\sup_{\tilde{n} < q < n - \tilde{n}} \tilde{n}^{1/2} \int_{|q-q_0|/\tilde{n}}^{1} K_+ (x) dx \rightarrow \rho (K)$. Because under $H_a$ given in (2.6), $\nu_n \tilde{n}^{1/2} r_n \left( x^0 \right) \rightarrow r$, following the arguments before (4.10), it suffices to show that

(4.14)

$$\text{Prob} \left\{ \nu_n \left( \sup_{\tilde{n} < q < n - \tilde{n}} (\widehat{r}_+ (q) - \widehat{r}_- (q)) - r \rho (K) - \xi_n \right) \leq x \right\} \rightarrow \exp \left( -e^{-x} \right).$$

But, proceeding as with Theorem 2.4, (4.14) holds true because

(a): The finite dimensional distributions of $\tilde{n}^{1/2} (\widehat{r}_+ (q) - \widehat{r}_- (q)) - r \rho (K)$ converge to those of a Gaussian process with correlation structure $\text{Corr} (b)$.

(b): $\tilde{n}^{1/2} (\widehat{r}_+ (q) - \widehat{r}_- (q))$ is tight proceeding as in Theorem 2.4. 

4.6. **Proof of Theorem 2.6.** Observing that for any sequence of random variables, $X_1, \ldots, X_n$, $\Pr \{ \max_{i=1}^{n} X_i > x \} \geq \Pr \{ \max_{i=k}^{n} X_i > x \}$, it suffices to show that there exists $\tilde{n} < q < n - \tilde{n}$ such that

$$\Pr \{ \nu_n (\widehat{r}_+ (q) - \widehat{r}_- (q) - \xi_n) > x \} \rightarrow 1$$

for all $x > 0$. Choose $q = q_0$, where $x_{q_0}$ is the closest point $x_q$ to $x^0$. Proceeding as in the proof of Proposition 2.3, we have that

$$\tilde{n}^{\frac{1}{2} - d} |\widehat{r}_+ (q_0) - \widehat{r}_- (q_0) - r| = \mathcal{O}_p (1).$$

So, $|\widehat{r}_+ (q_0) - \widehat{r}_- (q_0)| = \mathcal{O}_p \left( \tilde{n}^{d - \frac{1}{2}} \right) + r (1 + \mathcal{O} (1))$ so that

$$\nu_n \left( \tilde{n}^{\frac{1}{2} - d} |\widehat{r}_+ (q_0) - \widehat{r}_- (q_0)| - \xi_n \right) \rightarrow \infty$$

since by $C5$, $\tilde{n}^{\frac{1}{2} - d} \xi_n^2 = C \tilde{n}^d \log^{-1} n \rightarrow \infty$. From here the conclusion follows by standard arguments.
4.7. Proof of Theorem 3.1. The strategy is similar to that employed to show (4.10). First, we only need to show that for $c = 1, 2$,
\begin{equation}
\Pr \left\{ v_n \sup_{\tilde{a} < q < \tilde{a} + \tilde{b}} \tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q) - \tilde{a}^{\frac{1}{2}} (q) - \xi_n \right) \leq x | \tilde{y} \right\} \rightarrow \exp \left( -e^x \right),
\end{equation}
for $x > 0$. To that end, we will show that $\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q) - \tilde{a}^{\frac{1}{2}} (q) \right)$ converges, in bootstrap sense, to the Gaussian process $G (q)$ in $[0, \infty)$, and whose correlation structure is that given in (4.11). From here the arguments given in Theorem 2.4 imply that (4.15), and thus the theorem, holds true. First, by Lemmas 5.2 and 5.3, we have that
\begin{equation}
\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q_1) - \tilde{a}^{\frac{1}{2}} (q_1) \right) \rightarrow Cov \left( b \right).
\end{equation}
So,
\begin{equation}
\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q_1) - \tilde{a}^{\frac{1}{2}} (q_1) \right) \rightarrow Cov \left( b \right),
\end{equation}
holds if
\begin{equation}
\tilde{a}^{-\frac{1}{2}} b - 1 = o_p (1),
\end{equation}
which is the case, since Taylor’s expansion implies that the left side is bounded in absolute value by $|\tilde{a} - 1| \log n = o_p (1)$.

So, we have shown that the second moments of $\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q) - \tilde{a}^{\frac{1}{2}} (q) \right)$ converges in probability to those of $\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q) - \tilde{a}^{\frac{1}{2}} (q) \right)$. Since for $c = 1, 2$, $E^{\ast} u^{\ast} \tilde{a} = 0$ by construction, we obtain that, conditional on the data, the first moment of $\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q) - \tilde{a}^{\frac{1}{2}} (q) \right)$ converges to zero in probability. Next, we shall show that the finite dimensional distributions converge in probability to those of $G (q)$. However, the latter follows by Lemmas 5.2 to 5.4. Observe that Lemmas 5.2 and 5.3 imply that the second moments of $\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q_1) - \tilde{a}^{\frac{1}{2}} (q_1) \right)$ converges in probability to $Cov \left( b \right)$. On the other hand, Lemma 5.4 implies that Lindeberg’s condition holds in probability, so that we have that
\begin{equation}
\tilde{a}^{-\frac{1}{2}} \tilde{a} \left( \tilde{a}^{\frac{1}{2}} (q) - \tilde{a}^{\frac{1}{2}} (q) \right) \rightarrow \mathcal{N} (0, 1).
\end{equation}

To finish the proof of the theorem, it remains to prove the tightness condition to complete the proof. To that end, denote
\begin{equation}
X_{\tilde{a}, n} (q) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} u^{\ast} \tilde{K} \left( \frac{L}{n} - \tilde{q} \right), \quad \tilde{q} = \frac{1}{\tilde{n}}, 2/\tilde{n}, ..., [n]^{-1}.
\end{equation}
So, $X_{\tilde{a}, n} (q)$ is a process in $[0, D]$ equipped with Skorohod’s metric. Extend $[0, D]$ to $[0, \infty)$ by writing $X_{\tilde{a}, n} (\infty) = X_{\tilde{a}, n} \left( [a]^{-1} \right)$. By Pollard (1981, Ch.V), we need to show tightness in $[0, D]$ for any finite $D > 0$. To that end, by Billingsley’s (1968) Theorem 15.6, it suffices to show the moment condition
\begin{equation}
E \left[ \left( X_{\tilde{a}, n} (\tilde{q}) - X_{\tilde{a}, n} (\tilde{q}) \right)^{\beta} \right] \leq \left( \tilde{D} H_n (\tilde{q}, \tilde{q}) \right)^{2} \left( \tilde{q} - \tilde{q} \right)^{(1+\delta)/2} \left( \tilde{q} - \tilde{q} \right)^{(1+\delta)/2}
\end{equation}
for some $\delta, \beta > 0$ with $H_n (\tilde{q}, \tilde{q})$ being bounded in probability. Observe that we can consider only the situation for which $\tilde{n} < \tilde{q} \tilde{n}$, since otherwise we have that $(X_{\tilde{a}, n} (\tilde{q}) - X_{\tilde{a}, n} (\tilde{q})) (X_{\tilde{a}, n} (\tilde{q}) - X_{\tilde{a}, n} (\tilde{q}))$ is zero. This is the case because either $\tilde{q}$ and $\tilde{q}$ lie in the same interval $(j - 1)/n, j/n$ or else $\tilde{q}$ and $\tilde{q}$ do; in either of these cases $(X_{\tilde{a}, n} (\tilde{q}) - X_{\tilde{a}, n} (\tilde{q})) (X_{\tilde{a}, n} (\tilde{q}) - X_{\tilde{a}, n} (\tilde{q}))$ is zero.
Because for any $0 \leq a < b < c \leq D$,\(|c - b|\ |b - a| \leq |c - a|^2\), by the Cauchy-Schwarz inequality, a sufficient condition for the last displayed inequality is

\[
(4.16) \quad E^* \left| X^*_{+, n} (\tilde{q}_2) - X^*_{+, n} (\tilde{q}_1) \right|^{2\beta} \leq DH_n (\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{1+(\delta)} ,
\]

which is the case as we now show. It suffices to consider $|\tilde{q}_2 - \tilde{q}_1| \geq 1$ is trivial since the left side of (4.16) is bounded in probability.

Let us consider $d > 0$ first. By definition, $X^*_{+, n} (\tilde{q}_2) - X^*_{+, n} (\tilde{q}_1)$ is

\[
\frac{1}{n^{\frac{3}{2}+d}} \sum_{t=-n^{-(q_2-q_1)}+1}^{\tilde{n}} u^*_t + q_2, c \tilde{K}_{+, t} + \frac{1}{n^{\frac{3}{2}+d}} \sum_{t=1}^{\tilde{n}-(q_2-q_1)} u^*_t + q_2, c \left( \tilde{K}_{+, t} - \tilde{K}_{+, t+q_2-q_1} \right)
\]

(4.17)

\[-\frac{1}{n^{\frac{3}{2}+d}} \sum_{t=1}^{\tilde{n}-(q_2-q_1)} u^*_t + q_2, c \tilde{K}_{+, t}.
\]

Choosing $\beta = 1$, because $n \tilde{q} = q$ and

\[
\left| \frac{1}{n^{\frac{3}{2}+2d}} \sum_{t,s=1}^{q_2-q_1} E^* (u^*_t + q_1, c, u^*_s + q_1, c) \tilde{K}_{+, t} \tilde{K}_{+, s} \right| \leq D \frac{1}{n^{\frac{3}{2}+2d}} \sum_{t,s=1}^{q_2-q_1} \left| t - s \right|^{2d-1} 
\]

\[
\leq DH_n (\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{1+2d} 
\]

\[
\leq DH_n (\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{1+(\delta)} ,
\]

because by Lemma 5.2, $|E^* (u^*_t + q_1, c, u^*_s + q_1, c)| = |t - s|^{2d-1} (1 + o_p (1))$ where the $o_p (1)$ is independent of $\tilde{q}_1$ and $\tilde{q}_1$. So, the last term of (4.17) satisfies the inequality (4.16). Similarly, we obtain that

\[
E \left| \frac{1}{n^{\frac{3}{2}+d}} \sum_{t=-\tilde{n}-(q_2-q_1)+1}^{\tilde{n}} u^*_t + q_2, c \tilde{K}_{+, t} \right|^{2} \leq D \frac{1}{n^{\frac{3}{2}+2d}} \sum_{t,s=1}^{q_2-q_1} \left| t - s \right|^{2d-1} 
\]

\[
\leq DH_n (\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{1+(\delta)} 
\]

because $0 < \tilde{q}_2 - \tilde{q}_1 < 1$. Finally, the middle term in (4.17), which by continuous differentiability of $K_{+, u}$ for $u \in (0, 1)$, has a second moment bounded by

\[
D (\tilde{q}_2 - \tilde{q}_1)^2 \frac{1}{n^{\frac{3}{2}+2d}} \sum_{t,s=1}^{q_2-q_1} \left| t - s \right|^{2d-1} \leq D \frac{1}{n^{\frac{3}{2}+2d}} \sum_{t,s=1}^{q_2-q_1} \left| t - s \right|^{2d-1} 
\]

\[
\leq DH_n (\tilde{q}_2, \tilde{q}_1) |\tilde{q}_2 - \tilde{q}_1|^{1+(\delta)} 
\]

because $0 < d < 1/2$. So, (4.16) holds true choosing $\delta = 2d$ and hence $X^*_{+, n} (\tilde{q})$ is tight. On the other hand, proceeding similarly as with $X^*_{+, n} (\tilde{q})$,

\[
X^*_{-, n} (\tilde{q}) = \frac{1}{n^{\frac{3}{2}+d}} \sum_{t=1}^{\tilde{n}} u^*_t \tilde{K}_{-, \left( \frac{t}{n} - \tilde{q} \right)} \quad \tilde{q} = 1/n, 2/n, ... , [a]^{-1},
\]

where we have abbreviated $\tilde{K}_{-} \left( \tilde{n}^{-1} t - \tilde{q} \right) - E \tilde{K}_{-} \left( \tilde{n}^{-1} t - \tilde{q} \right)$ by $\tilde{K}_{-} \left( \tilde{n}^{-1} t - \tilde{q} \right)$ is also tight. So, the process $\tilde{n}^{\frac{3}{2}+d} (\tilde{r}^*_t, c (\tilde{q}) - \tilde{r}^*_t, c (\tilde{q}))$ is tight, which concludes the proof of the theorem for $0 < d < 1/2$, since we have that $X^*_{+, n} (\tilde{q})$ converges, in bootstrap, weakly to a gaussian process in $\mathbb{D} [0, \infty)$ whose correlation structure satisfies

\[(a) \ Corr (b) = 1 - Dh^a + o (b^a) ; \quad b \to 0 \]

\[(b) \ Corr (b) \log b = o (1) ; \quad b \to \infty. \]
Next, when \( d = 0 \). In this case the proof is even simpler because \( u_{t,c}^* \) is a weakly dependent sequence with fourth moments which are bounded in probability and it is well known by standard kernel arguments that, say,

\[
E^\ast \left( \frac{1}{n} \sum_{t=1}^{d} u_{t,c}^* K_{+} \left( \frac{t}{n} \right) \right)^4 \leq DH_n \left( \tilde{q}_2, \tilde{q}_1 \right) |\tilde{q}_2 - \tilde{q}_1|^2
\]

so that

\[
E \left| X_{t,n}^*(\tilde{q}_2) - X_{t,n}^*(\tilde{q}_1) \right|^4 \leq DH_n \left( \tilde{q}_2, \tilde{q}_1 \right) |\tilde{q}_2 - \tilde{q}_1|^2,
\]

which completes the proof for \( d = 0 \). Finally, the case \( -\frac{1}{2} < d < 0 \) follows similarly as that for \( d > 0 \), so the proof is omitted. \( \square \)

5. APPENDIX A

**Lemma 5.1.** Assuming C1 – C6, we have that

\[
\frac{1}{n^{2d}} \sum_{t=0}^{n-1} \left( \hat{\gamma} \left( t; \hat{d} \right) - \gamma (t; d) \right) = o_p (1).
\]

**Proof.** The left side of (5.1) is

\[
\frac{1}{n^{2d}} \sum_{t=0}^{n-1} \left( \hat{\gamma} \left( t; \hat{d} \right) - \gamma (t; d) \right) + \frac{1}{n^{2d}} \sum_{t=p+1}^{n-1} t^{1-2d} \left( \hat{\gamma} \left( t; \hat{d} \right) - \gamma (t; d) \right).
\]

The first term of the last displayed expression is \( o_p (1) \) if \( p \) is chosen such that \( p^{-1} + p/n^{2d} \to 0 \). On the other hand, because

\[
t^{1-2d} \left| \hat{\gamma} \left( t; \hat{d} \right) - \gamma (t; d) \right| \leq D \left| t^{2(\hat{d} - d)} - 1 \right| \leq D 2 |\hat{d} - d| \log t
\]

by Taylor expansion, it implies that the absolute value of the second term of (5.2) is bounded by

\[
D \sup_{p+1 \leq t \leq n-1} |\hat{d} - d| \log t = o_p (1)
\]

following Robinson (1997) as \( p \) and \( n \) increase to infinity and that \( \hat{d} - d = o_p (\log^{-1} n) \). Observe that to show (5.1), we have showed that \( \sup_t t^{1-2d} |\hat{\gamma} \left( t; \hat{d} \right) - \gamma (t; d)| = o_p (1) \). \( \square \)

**Lemma 5.2.** Assuming C1 – C6, we have that for \( c = 1, 2 \)

\[
E^\ast \left( u_{t,c}^* u_{s,c}^* \right) = o_p \left( n^{2d-1} \right).
\]

**Proof.** Because for \( c = 1, 2 \), \( E^\ast (\eta_{j,c}^* \eta_{j,c}^*) = \mathcal{I} (j = \ell) \), we obtain that the left side is

\[
\frac{2\pi}{n} \sum_{j=1}^{n} e^{i(t-s)\lambda_j} \left( \sum_{\ell=-n+1}^{n-1} \hat{\gamma} \left( \ell; \hat{d} \right) e^{-i\ell \lambda_j} \right) \left| \hat{h}_1 \left( e^{-i\lambda_j} \right) \right|^2
\]

\[
= \frac{2\pi}{n} \sum_{j=1}^{n} e^{i(t-s)\lambda_j} \left( \sum_{\ell=-n+1}^{n-1} \hat{\gamma} \left( \ell; \hat{d} \right) e^{-i\ell \lambda_j} \right) h \left( e^{-i\lambda_j} \right) (1 + o (1))
\]

because by Hidalgo and Yajima (2002), \( \sup_j \left( \left| \hat{h}_1 \left( e^{-i\lambda_j} \right) \right|^2 - h \left( e^{-i\lambda_j} \right) \right) = o_p (1) \).
Using the identity $h(e^{-i\lambda}) = (2\pi)^{-1} \sum_{\ell = -\infty}^{\infty} \beta_{\ell} e^{i\lambda \ell}$, where $\beta_{\ell} = \int h(\lambda) e^{i\lambda \ell} d\lambda$, it implies that the dominant term of the right side of the last displayed equation is

$$\sum_{\ell = -n+1}^{n-1} \hat{\gamma}(\ell; \tilde{d}) \sum_{q} \beta_{t-s+\ell+qv}$$

because $\sum_{\ell} e^{i\lambda \ell} = 1$ if $\ell = 0, n, 2n, \ldots$, and zero otherwise. Now, because $\beta_{\ell} = o\left(|p|^{-2}\right)$, we have that the last expression is

$$\sum_{\ell = -n+1}^{n-1} \hat{\gamma}(\ell; \tilde{d}) \beta_{t-s+\ell} + o\left(n^{-1}\right) \sum_{\ell = -n+1}^{n-1} |\hat{\gamma}(\ell; \tilde{d})|$$

$$= \sum_{\ell = -n+1}^{n-1} \hat{\gamma}(\ell; \tilde{d}) \beta_{t-s+\ell} + o_p(n^{2d-1})$$

because from the proof of Lemma 5.1, we have that $\sup_{\lambda} \ell^{2d-1} |\hat{\gamma}(\ell; \tilde{d}) - \gamma(\ell; d)| = o_p(1)$ and C3 implies that $|\gamma(\ell; d)| = O(\ell^{2d-1})$. Now, the first term on the right of the last displayed equation is

$$(5.4) \quad \sum_{\ell = -n+1}^{n-1} \left(\hat{\gamma}(\ell; \tilde{d}) - \gamma(\ell; d)\right) \beta_{t-s+\ell} + \sum_{\ell = -n+1}^{n-1} \gamma(\ell; d) \beta_{t-s+\ell}.$$ 

The second term of (5.4) is

$$\sum_{\ell = -\infty}^{\infty} \gamma(\ell; d) \beta_{t-s+\ell} + \left\{ \sum_{\ell = n}^{\infty} + \sum_{\ell = -\infty}^{n-1} \right\} \gamma(\ell; d) \beta_{t-s+\ell}$$

$$= \delta_{[t-s]} + o_p\left(|t-s|^{-1}\right)$$

because $\beta_{\ell} = o\left(|p|^{-2}\right)$ and by definition $\delta_{[t-s]} = \sum_{\ell = -\infty}^{\infty} \gamma(\ell; d) \beta_{t-s+\ell}$, see for instance Corollary 3.4.1.1 of Fuller (1996).

So, to complete the proof it suffices to show that the first term of (5.4) is $o_p\left(|t-s|^{2d-1}\right)$. But that term is bounded in absolute value by

$$\left\{ \sum_{|\ell| < 2^{-1}|t-s|} + \sum_{2|t-s| < |\ell| < n} + \sum_{2^{-1}|t-s| < |\ell| < 2|t-s|} \right\} \left|\hat{\gamma}(\ell; \tilde{d}) - \gamma(\ell; d)\right| \left|\beta_{t-s+\ell}\right|.$$ 

The contribution of the first two terms, by Lemma 5.1 and that $\beta_{\ell} = o\left(|p|^{-2}\right)$, are easily shown to be $o_p\left(|t-s|^{-1}\right)$, whereas the contribution due to the third term is $o_p\left(|t-s|^{2d-1}\right)$ because by Lemma 5.1 we have that $\sup_{\lambda} \ell^{2d-1} |\hat{\gamma}(\ell; \tilde{d}) - \gamma(\ell; d)| = o_p(1)$, so that $\sup_{2^{-1}|t-s| < |\ell| < 2|t-s|} |\hat{\gamma}(\ell; \tilde{d}) - \gamma(\ell; d)| = o_p\left(|t-s|^{2d-1}\right)$. This concludes the proof of the lemma.

**Lemma 5.3.** For any $\bar{n} < q_1 \leq q_2 < n - \bar{n}$, as $n \to \infty$,

$$\bar{n}^{1-2d} \text{Var}^* \left( \tilde{\tau}_{x,c}^* (q_1) - \tilde{\tau}_{x,c}^* (q_1) \tilde{\tau}_{x,c}^* (q_2) - \tilde{\tau}_{x,c}^* (q_2) \right)$$

$$\to \rho_+ (b; d) + \rho_- (b; d) - \rho_\pm (b; d) - \rho_{\mp} (b; d),$$

for $c = 1, 2$, where the right side is as defined in Proposition 2.1.
Proof. By definition, $\tilde{n}^{1-2d}\text{Cov}^*(\tilde{r}^*_+, (q_1), \tilde{r}^*_+, (q_2), \tilde{r}^*_+, (q_3))$ is
\begin{equation}
(5.5) \quad \tilde{n}^{1-2d}\text{Cov}^*(\tilde{r}^*_+, (q_1), \tilde{r}^*_+, (q_2)) + \tilde{n}^{1-2d}\text{Cov}^*(\tilde{r}^*_+, (q_1), \tilde{r}^*_+, (q_2)) - \tilde{n}^{1-2d}\text{Cov}^*(\tilde{r}^*_+, (q_1), \tilde{r}^*_+, (q_2)),
\end{equation}

As was done in the proof of Proposition 2.1, we will only examine the first term of (5.5), the other three terms follow similarly. This term is
\[
\frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} E^* (u^*_t, u^*_s) K_{t+q_1, t+q_2} \cdot \frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} E (u^*_t, u^*_s) K_{t+q_1, t+q_2}.
\]
Because the right side is
\[
\frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} \{ E (u^*_t, u^*_s) - \delta_{t-s} \} K_{t+q_1, t+q_2} + \frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} \delta_{t-s} K_{t+q_1, t+q_2},
\]
it only suffices to show that the first term of the last displayed expression converges to zero in probability because by Proposition 2.1, the second term converges to $p_{(b, d)}$.

But by Lemma 5.2, we have that the former term is bounded in absolute value by
\[
o_p (n^{2d-1}) \frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} \sum_{s+q_1+1} |K_{t+q_1, t+q_2}| + o_p (1) \frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} \sum_{s+q_1+1} |t-s|^{2d-1} |K_{t+q_1, t+q_2}|
\]
which is $o_p(1)$ because $\tilde{n}^{-1} \sum_{t=1}^{D} |K_{t+q_1, t+q_2}| \leq D$ and in Proposition 2.1 we have shown that
\[
\frac{1}{\tilde{n}^{1+2d}} \sum_{q_1+\tilde{n}} q_2+\tilde{n} \sum_{s+q_1+1} |t-s|^{2d-1} |K_{t+q_1, t+q_2} \leq D
\]
so that the second term of the last displayed expression is $o_p(1)$.

\begin{lemma}
Assuming C1 - C7, for $c = 1, 2$
\begin{equation}
(5.6) \quad \sum_{j=1}^{n} E^* \left( n^{-1} \tilde{n}^{-1-2\tilde{d}} \left| \xi_{+, q} (\lambda_j) \eta_{j,c}^* \right|^2 \cdot \mathcal{T} \left( n^{-1} \tilde{n}^{-1-2\tilde{d}} \left| \xi_{+, q} (\lambda_j) \eta_{j,c}^* \right|^2 > \phi \right) \right) L_{0, \tilde{n}}
\end{equation}
\begin{equation}
(5.7) \quad \sum_{j=1}^{n} E^* \left( n^{-1} \tilde{n}^{-1-2\tilde{d}} \left| \xi_{-, q} (\lambda_j) \eta_{j,c}^* \right|^2 \cdot \mathcal{T} \left( n^{-1} \tilde{n}^{-1-2\tilde{d}} \left| \xi_{-, q} (\lambda_j) \eta_{j,c}^* \right|^2 > \phi \right) \right) L_{0, \tilde{n}}
\end{equation}
for all $\phi > 0$, where $\xi_{+, q} (\lambda_j) = \tilde{k}_{\pm, q} (\lambda_j) \sum_{t=0}^{n-1} \tilde{\gamma} (\tilde{t}, \tilde{d}) e^{-it\lambda_j} \left| \right|_{1/2} \tilde{h}_{\pm, q} (\tilde{t}) e^{-it\lambda_j}$, and
\begin{align*}
\hat{k}_{\pm, q} (\lambda_j) &= \sum_{t=1}^{n} \hat{K}_{\pm} \left( (t-q)/\tilde{n} \right) e^{it\lambda_j},
\end{align*}
for all $\phi > 0$.
Proof. We shall prove only (5.6), the proof for (5.7) is identical. First we show that
\[(5.8) \quad n^{-1} \bar{n}^{-1-2d} \max_{1 \leq j \leq n} \left| \xi^{(i,j)} \right|^2 = O_p (a^{1-2d}) \]
Because \( \sup_j \left| \tilde{h}_1 \left( e^{i\lambda_j} \right) \right|^2 = o_p (1) \) by Hidalgo and Yajima (2002), (5.8)
holds true if
\[
\max_{1 \leq j \leq n} \left( \frac{1}{n \bar{n} 1+2d} h \left( e^{i\lambda_j} \right) \sum_{\ell=n+1}^{\infty} \tilde{\gamma} \left( \ell; \hat{d} \right) e^{-i\lambda_j} \left| \sum_{t=1}^{n} K_+ \left( \frac{t-q}{\bar{n}} \right) e^{it\lambda_j} \right|^2 \right) = O_p (a^{1-2d}) .
\]
Next the left side of the last displayed equation is
\[(5.9) \quad \max_{1 \leq j \leq n} \left( \frac{1}{n \bar{n} 1+2d} h \left( e^{i\lambda_j} \right) \sum_{\ell=-n+1}^{\infty} \gamma \left( \ell; d \right) e^{-i\lambda_j} \left| \sum_{t=1}^{n} K_+ \left( \frac{t-q}{\bar{n}} \right) e^{it\lambda_j} \right|^2 \right) \]
\[+ \left( \frac{1}{n \bar{n} 1+2d} h \left( e^{i\lambda_j} \right) \sum_{\ell=-n+1}^{\infty} \left( \tilde{\gamma} \left( \ell; \hat{d} \right) - \gamma \left( \ell; d \right) \right) e^{-i\lambda_j} \left| \sum_{t=1}^{n} K_+ \left( \frac{t-q}{\bar{n}} \right) e^{it\lambda_j} \right|^2 \right) \].
Now, the contribution of the first term of (5.9) is bounded in modulus by
\[ D \left( \frac{\bar{n}}{n} \right)^{1-2d} \left| \sum_{t=1}^{n} K_+ \left( \frac{t-q}{\bar{n}} \right) e^{it\lambda_j} \right|^2 \left| \sum_{\ell=-n+1}^{\infty} \gamma \left( \ell; d \right) \right| = O_p (a^{1-2d}) \]
because \( \sum_{t=1}^{n} \left| K_+ \left( \frac{t-q}{\bar{n}} \right) \right| = O (\bar{n}) \), \( \sum_{\ell=-n+1}^{\infty} \left| \gamma \left( \ell; d \right) \right| = O (n^{2d}) \), \( \left| \gamma \left( \lambda \right) \right| \leq D \) and \( n^{2(\bar{d}-d)} - 1 = o_p (1) \). On the other hand, the contribution of the second term of (5.9) is bounded in modulus by
\[ \frac{\bar{n}}{n} \left( \frac{1}{\bar{n}^2} \sum_{t=1}^{n} \left| K_+ \left( \frac{t-q}{\bar{n}} \right) \right| \right)^2 \left| \sum_{\ell=-n+1}^{\infty} \left( \tilde{\gamma} \left( \ell; \hat{d} \right) - \gamma \left( \ell; d \right) \right) \right| = O_p (n^{2d}) \]
by the previous arguments and because by Lemma 5.1, \( \sum_{\ell=-n+1}^{\infty} \left| \tilde{\gamma} \left( \ell; \hat{d} \right) - \gamma \left( \ell; d \right) \right| = o_p (n^{2d}) \). So, we conclude that (5.8) holds true, which implies that with probability approaching one the left side of (5.6) is bounded by
\[(5.10) \quad \sum_{j=1}^{n} E^* \left( n^{-1} \bar{n}^{-1-2d} \left| \xi^{(i,j)} \right|^2 \left| \eta_{j,c}^{*} \right|^2 \mathcal{I} \left( \left| \eta_{j,c}^{*} \right|^2 > \phi a^{2d-1} \right) \right) .
\]
However, conditional on the data, \( \eta_{j,1}^{*} \) is a sequence of iid random variables, so for \( i = 1 \) (5.10) is
\[ E^* \left( \left| \eta_{j,1}^{*} \right|^2 \mathcal{I} \left( \left| \eta_{j,1}^{*} \right|^2 > \phi a^{2d-1} \right) \right) \frac{D}{n \bar{n}^{1+2d}} \sum_{j=1}^{n} \left| \xi^{(i,j)} \right|^2 ,
\]
which converge to zero in probability since \( \phi > 0, d < 1/2, a \to 0 \) by C6 and \( \eta_{j,1}^{*} \)
has finite second moments and by Lemmas 5.2 and 5.3, the second factor of the last displayed expression is bounded in probability.
On the other hand, for \( i = 2 \), (5.10) is bounded by
\[ \sup_j E^* \left( \left| \eta_{j,2}^{*} \right|^2 \mathcal{I} \left( \left| \eta_{j,2}^{*} \right|^2 > \phi a^{2d-1} \right) \right) \frac{D}{n \bar{n}^{1+2d}} \sum_{j=1}^{n} \left| \xi^{(i,j)} \right|^2 ,
\]
which converges to zero in probability for all \( \phi > 0 \) as we now show. First, by Lemmas 5.2 and 5.3, the second factor of the last displayed expression is bounded in probability. On the other hand, standard algebra implies that

\[
E^* \left( |\eta^*_{j,2}|^2 I \left( |\eta^*_{j,2}|^2 > \phi a^{2d-1} \right) \right) \leq \frac{Da^2}{\phi^2} E^* |\eta^*_{j,2}|^4
\]

\[
\leq \frac{Da^2}{\phi^2} \left\{ \frac{1}{T} \sum_{t=1}^{T} u^4_t + \left( \frac{1}{T} \sum_{t=1}^{T} u^2_t \right)^2 \right\}.
\]

But, by a well-know argument, see Stout’s (1974) Theorem 3.5.8, \( u_t \) is also ergodic by \( C^2 \), which implies that the right side of the last displayed inequality converges to zero because \( \phi > 0, \ d < 1/2 \) and \( a \to 0 \) by \( C^6 \). So, we conclude that (5.10) \( P \to 0 \) and the proof of the Lemma.

\[\Box\]

6. CONCLUSIONS AND EXTENSIONS

In this paper we have described an alternative bootstrap to the sieve-bootstrap or others widely used with time series data. We have applied the bootstrap to test for smoothness of a nonparametric regression function. We have shown that the bootstrap is valid for data which may exhibit long-memory. In this sense, we have been able to relax the assumption of the data being (strong) mixing, which is commonly assumed to show the validity of the Moving block or subsampling bootstraps as well as the sieve-bootstrap. Since, our test is based on the supremum of the difference between the kernel regression estimates when estimated from the right and left of the point, our results can easily be adapted to construct simultaneous confidence bands in an interval. So, contrary to Bühlmann’s (1998) results, we do not need to shrink to zero the length of the interval as the sample size increases to infinity, albeit extending his work to situations for which the data exhibits long-memory dependence. The latter problem of simultaneous confidence bands has been analyzed by, for instance, Härdle and Bowman (1988), Härdle and Marron (1991) or Eubank and Speckman (1993) among others, extending earlier work by Bickel and Rosenblatt (1973) for the kernel density estimator. Another possible extension is to test for a particular regression function, by comparing the parametric and nonparametric fit using the sup norm.
References


London School of Economics, Houghton Street, London WC2A 2AE, U.K.