The Annals of Statistics
1997, Vol. 25, No. 5, 2054–2083

LARGE-SAMPLE INFERENCE FOR NONPARAMETRIC REGRESSION WITH DEPENDENT ERRORS[1]

BY P. M. ROBINSON

London School of Economics

A central limit theorem is given for certain weighted partial sums of a covariance stationary process, assuming it is linear in martingale differences, but without any restriction on its spectrum. We apply the result to kernel nonparametric fixed-design regression, giving a single central limit theorem which indicates how error spectral behavior at only zero frequency influences the asymptotic distribution and covers long-range, short-range and negative dependence. We show how the regression estimates can be Studentized in the absence of previous knowledge of which form of dependence pertains, and show also that a simpler Studentization is possible when long-range dependence can be taken for granted.

1. Introduction. This paper justifies approximate normal inference on fixed design nonparametric regression in the presence of dependent observations. The dependence structures covered are unusually diverse, because the stationary errors can exhibit dependence of short-range, long-range, or negative type. Also, unusually for the time series regression literature, we give a single central limit theorem which simultaneously covers all three cases. The limiting covariance structure of the estimates depends on the nature of the dependence through only a self-similarity parameter, as well as a scale factor, and we indicate how to validly Studentize the regression estimates by estimating these parameters, without prejudging whether there is short-range, long-range, or negative dependence. The Studentization is based on residuals from the regression model, but we also show that when long-range dependence can be taken for granted, the raw data can be used for this purpose. The paper clarifies the feature of serial correlation which is really relevant, namely the behavior of the spectral density at only zero frequency. In fact, the same point applies in other problems, to Gasser–Müller regression estimates as well as the kernel ones we study, to certain wavelet regression problems, and to versions of parametric regression; Lemmas 1 and 2 below can be checked to provide analogous central limit theorems in these problems.

Received March 1995; revised March 1997.

[1]Research supported by ESRC Grant R000235892.

AMS 1991 *subject classifications*. Primary 62G07, 60G18; secondary 62G20.

Key words and phrases. Central limit theorem, nonparametric regression, autocorrelation, long range dependence.

We consider the model

$$y_t = r\left(\frac{t}{n}\right) + u_t, \qquad t = 1, 2, \ldots, \tag{1.1}$$

where y_t is observed for $t = 1, \ldots, n$, and u_t is an unobservable error. As always in this model, y_t must be viewed as a triangular array (as can be u_t, if so desired) as n increases, but we suppress reference to n here. To estimate $r(x)$, $x \in (0,1)$, we consider

$$\hat{r}(x) = \frac{1}{nb} \sum_t k\left(\frac{nx - t}{nb}\right) y_t, \tag{1.2}$$

where b is a positive bandwidth number, k is a kernel function such that

$$\int_{-\infty}^{\infty} k(v)\, dv = 1 \tag{1.3}$$

and $\sum_t$ will always denote a sum over t from 1 through n. Taking for granted covariance stationarity of u_t, implying the mean and autocovariances are time invariant, then $Eu_1 = 0$ with no loss of generality, and we denote by $\gamma_j = Eu_1 u_{1+j}$ the lag-j autocovariance of u_t. We introduce the following assumption.

ASSUMPTION 1. The process u_t is covariance stationary with absolutely continuous spectral distribution function, its spectral density, $f(\lambda)$, defined by $\gamma_j = \int_{-\pi}^{\pi} f(\lambda)\cos(j\lambda)\, d\lambda$, being of form

$$f(\lambda) = g(\lambda)h(\lambda), \qquad -\pi < \lambda \le \pi, \tag{1.4}$$

where

(i) g is an even nonnegative function that is continuous and positive at $\lambda = 0$, and we denote $G = g(0)$;

(ii) h is an integrable function such that

$$\begin{aligned} \delta_j &= \int_{-\pi}^{\pi} h(\lambda)\cos(j\lambda)\, d\lambda \\ &\sim \theta(H) j^{2H-2} \quad \text{as } j \to \infty, \qquad H \in (0,1) \setminus \{\tfrac{1}{2}\}, \\ &= 2\pi\Delta_{j0}, \qquad H = \tfrac{1}{2}, \end{aligned} \tag{1.5}$$

where Δ_{ab} is the Kronecker delta, $\theta(H) = 2\Gamma(2 - 2H)\cos\{\pi(1 - H)\}$ and also

$$h(0) = 0, \qquad 0 < H < \tfrac{1}{2}. \tag{1.6}$$

The cases $H = \frac{1}{2}$ and $0 < H < \frac{1}{2}$ referred to here are termed, respectively, "short-range dependence" and "negative dependence," the complementary case, $\frac{1}{2} < H < 1$, on $(0,1)$ being "long-range dependence." Then $\theta(H)$ is negative for $0 < H < 1/2$ and positive for $1/2 < H < 1$, so that δ_j is, respectively, eventually negative and positive. For $H = 1/2$ we have prescribed δ_j to thus interpolate between these other two cases, but there is no loss of generality relative to assuming only that $h(\lambda)$ is continuous and positive at $\lambda = 0$. We formally identify the case $H = 1/2$ with a constant $h(\lambda)$ instead of taking $\delta_j \sim \theta(1/2)j^{-1}$ so as to avoid the discontinuity in convergence rates resulting from the latter specification [see Hall and Hart (1990)], thereby to simplify Studentization (see Section 4) and also to explicitly include short-range dependence. In practice, the scale factor G is unknown, and the reason for incorporating the factor $\theta(H)$ in (1.5) is so that Assumption 1 corresponds approximately to a simple local parameterization in the frequency domain,

$$f(\lambda) \sim G\lambda^{1-2H} \quad \text{as } \lambda \to 0+, \tag{1.7}$$

for all $H \in (0, 1)$. Assumption 1(i) and

$$h(\lambda) \sim \lambda^{1-2H} \quad \text{as } \lambda \to 0+, \tag{1.8}$$

together imply (1.7). For $H \in (0, \frac{1}{2})$, the Fourier series of $h(\lambda)$ converges absolutely and (1.6) implies

$$\sum_{j=-\infty}^{\infty} \delta_j = 0, \tag{1.9}$$

and thence (1.8) from Theorem III-31 of Yong (1974). For $H = \frac{1}{2}$, (1.8) obviously holds. For $H \in (\frac{1}{2}, 1)$, (1.8) is equivalent to (1.5) if the δ_j are quasimonotonically convergent to zero, that is, there exists $C < \infty$ such that $\delta_{j+1} \le \delta_j(1 + C/j)$ for all sufficiently large j [Yong (1974), Theorem III-14].

Whatever the value of $H \in (0, 1)$, Assumption 1 effectively imposes no restrictions on f away from zero frequency, apart from integrability implied by covariance stationarity: it can be infinite or zero at any other frequencies, for example. This contrasts with assumptions made in previous work on central limit theory for nonparametric regression. The case $H = 1/2$ was in effect addressed by Roussas, Tran and Ioannides (1992), Csörgő and Mielniczuk (1995a), and Tran, Roussas, Yakowitz and Truong Van (1996). They modelled u_t as a strongly mixing process with at least summable mixing numbers, a nonlinear function, with Hermite number m, of a Gaussian process with autocovariances whose mth absolute powers are summable, and a linear process with absolutely summable weights (cf. Assumption 2), respectively. All these assumptions imply that f is bounded (and indeed satisfies a Lipschitz continuity condition of degree greater than $1/2$) on $(-\pi, \pi]$. Csörgő and Mielniczuk (1995b, c) addressed the case $H \in (1/2, 1)$. Their assumption on γ_j is more general in that they allow for a slowly varying factor in the right-hand side of (1.5): we could incorporate this, but as Cs̈orgő and Mielniczuk's (1995b, c) work indicates, it will then arise in the norming for asymptotic normality unless it satisfies very restrictive conditions. We stress Studentization later in the paper, in which practical circumstances it seems unlikely that the applied worker would incorporate any such factor. Ignoring this aspect, Assumption 1 is more general than Csörgő and Mielniczuk's (1995b, c), because they effectively take $g(\lambda) \equiv G$; the consequent requirement that

$$\gamma_j \sim G\theta(H) j^{2H-2} \quad \text{as } j \to \infty, \tag{1.10}$$

rules out such models covered by Assumption 1 as

$$f(\lambda) \propto |1 - e^{i\lambda}|^{1-2H} |1 - 2\cos \omega e^{i\lambda} + e^{2i\lambda}|^{1/2-J} \tag{1.11}$$

for $0 < \omega \leq \pi$ and $0 < H \leq J$ if $J > 1/2$, so there is a singularity around frequency $\lambda = \omega$, of magnitude that at least matches that (if $H > 1/2$) at $\lambda = 0$. (The second factor in (1.11) has been discussed in detail by Gray, Zhang and Woodward, 1989.) At about the same time and independently of our work, Deo (1997) has also considered the case $H \in (1/2, 1)$, but requires that $f(\lambda)$ is positive and continuous at all $\lambda \in (-\pi, \pi) \setminus \{0\}$.

The fact that spectral behavior matters only at frequency 0 is familiar from theory for partial sums of weakly autocorrelated series, corresponding to the case $H = 1/2$. In particular, $V(n^{-1/2} \sum_t u_t) \to 2\pi f(0)$ if f is continuous at $\lambda = 0$, where V denotes the variance operator. For $H \in (1/2, 1)$, however, (1.10) has been stressed (again with a slowly varying factor) in theory for partial sums and other statistics, following Taqqu (1975). By arguing as in the proof of Lemma 3 below, we can partially extend Lemma 3.1 of Taqqu (1975) and Theorem 2.1 of Robinson (1993) [who took, respectively, $\frac{1}{2} < H < 1$ and $0 < H < \frac{1}{2}$ in (1.10) and allowed for a slowly varying factor] to obtain under Assumption 1,

$$V\left(n^{-H} \sum_t u_t\right) \to G\theta(H)/H(2H-1), \qquad H \in (0,1), \tag{1.12}$$

taking $\sin 0/0 = 1$. Likewise, under Assumption 1 we can justify the formula in Yajima (1988) for the asymptotic covariance matrix of the least squares estimate in polynomial time series regression with errors u_t, which he obtained under the assumption that g is everywhere continuous.

The following section gives some central limit results for weighted partial sums of a covariance stationary process that is linear in martingale differences with weights that are only square summable and is relevant to Assumption 1 for all $H \in (0,1)$. The results can apply to various problems, but in Section 3 we check them in the case of estimate (1.2) of (1.1). Section 4 justifies the same normal approximation for suitably Studentized estimates. Section 5 contains an empirical application to the series of annual minimum levels of the Nile River.

2. Central limit theorem for weighted sums of linearly dependent variates.

This section considers central limit theory for weighted partial sums of a sequence u_t satisfying the following.

ASSUMPTION 2.

$$u_t = \sum_{j=-\infty}^{\infty} \alpha_j \varepsilon_{t-j}, \qquad \sum_{j=-\infty}^{\infty} \alpha_j^2 < \infty, \tag{2.1}$$

where the ε_t^2 are uniformly integrable and

$$E(\varepsilon_t|F_{t-1}) = 0, \qquad E(\varepsilon_t^2|F_{t-1}) = 1 \quad \text{a.s. } t = 0, \pm 1, \dots,$$

where F_t is the σ-field of events generated by $\{\varepsilon_s,\ s \le t\}$.

For a triangular array $\{w_{tn},\ t = 1, \dots, n;\ n = 1, 2, \dots\}$, write $S = \sum_t w_{tn} u_t$. The following result is related to ones of Eicker (1967), Ibragimov and Linnik (1971) and Hannan (1979).

LEMMA 1. *Let Assumption* 2 *hold and*

$$\sum_{j=-\infty}^{\infty} v_{jn}^2 = 1 \quad \textit{for all } n, \tag{2.2}$$

$$\lim_{n\to\infty} \max_{-\infty<j<\infty} |v_{jn}| = 0, \tag{2.3}$$

where $v_{jn} = \sum_t w_{tn} \alpha_{t-j}$. *Then*

$$S \to_d N(0,1) \quad \textit{as } n \to \infty. \tag{2.4}$$

PROOF. For any $N \ge 1$, $S = S_{-\infty}^{-N-1} + S_{-N}^{N} + S_{N+1}^{\infty}$, where $S_p^q = \sum_{j=p}^q v_{jn}\varepsilon_j$. Thus

$$V(S_{-\infty}^{-N-1}) + V(S_{N+1}^{\infty}) = \sum_{|j|>N} v_{jn}^2 \le \left(\sum_t |w_{tn}|\right)^2 \left(\sum_{j>N} \alpha_j^2 + \sum_{j<n-N} \alpha_j^2\right). \tag{2.5}$$

The squared factor depends on n only and in view of (2.1) we can choose $N = N_n$ as a function of n such that (2.5) $\to 0$ as $n \to \infty$. For such N write $S_{-N}^{N} = \sum_{t=1}^{2N+1} x_{tn}$ where $x_{tn} = v_{t-N-1,n}\varepsilon_{t-N-1}$. Thus $E(x_{tn}|F_{t-N-2}) = 0$, $E(x_{tn}^2|F_{t-N-2}) = v_{t-N-1,n}^2$ a.s., by Assumption 2. It follows from (2.2) and from Corollary (3.8) of McLeish (1974) that $S_{-N}^{N} \to_d N(0,1)$ as $n \to \infty$ if, for all $\eta > 0$,

$$\lim_{n\to\infty} \sum_{t=1}^{2N+1} E\big(x_{tn}^2 I(x_{tn}^2 > \eta)\big) = 0.$$

By (2.2) the sum on the left-hand side is bounded by $\max_t E(\varepsilon_t^2 I(\varepsilon_t^2 > \eta/\max_j v_{jn}^2)) \to 0$ as $n \to \infty$ by uniform integrability and (2.3). □

Condition (2.2) is merely a normalization. Sufficient conditions for (2.3) are given in the following lemma. Write

$$\sigma_n^2 = V\left(\sum_t u_t\right) = \sum_{t=-\infty}^{\infty} (\alpha_{t-1} + \cdots + \alpha_{t-n})^2,$$

and introduce the difference operator $\triangle$, such that $\triangle w_{tn} = w_{tn} - w_{t+1,n}$.

LEMMA 2. *Let Assumption* 2 *hold. If either* (i) *there exists a positive-valued sequence* $a = a_n$ *such that, as* $n \to \infty$,

$$\left(\sum_t w_{tn}^2 \sum_{|j|>a} \alpha_j^2 \right)^{1/2} + \max_{1 \le t \le n} |w_{tn}| \sum_{|j| \le a} |\alpha_j| \to 0, \tag{2.6}$$

or if (ii) *for* $I < \infty$, $w_{tn} = \sum_{i=1}^{I} w_{itn}$ *and for all* $i = 1, \dots, I$ *there exist sequences* $p_i = p_{in}$ *and* $q_i = q_{in}$ *such that* $1 \le p_i < q_i \le n$ *for all* n, *and, as* $n \to \infty$,

$$\begin{aligned} &\sum_{t=1}^{p_i - 1} w_{itn}^2 + \sum_{t=q_i+1}^{n} w_{itn}^2 + \sum_{t=p_i}^{q_i} |\triangle w_{itn}| \left(\sigma_{t-p_i+1}^{1/2} + 1 \right) \\ &\qquad + |w_{iq_in}| \left(\sigma_{q_i - p_i + 1}^{1/2} + 1 \right) + \sigma_n^{-1} \to 0, \end{aligned} \tag{2.7}$$

then (2.3) *holds.*

PROOF. (i) It is easily seen that $\max_j |v_{jn}|$ is bounded by the left-hand side of (2.6).

(ii) We check (2.7) for an arbitrary i, dropping the i subscript in w_{itn}, p_i, q_i. We can easily account for the contribution from the summands for $t < p$ and $t > q$ in v_{jn}, and then by summation by parts,

$$\max_j \left| \sum_{t=p}^{q} w_{tn} \alpha_{t-j} \right| \le \sum_{t=p}^{q-1} |\triangle w_{tn}| \max_j |A_{p-j}^{t-j}| + |w_{qn}| \max_j |A_{p-j}^{q-j}|, \tag{2.8}$$

where $A_s^t = \sum_{j=s}^{t} \alpha_j$. As in Ibragimov and Linnik (1971), we have the identity

$$\left(A_{p-i}^{t-i} \right)^2 - \left(A_{p+1-i}^{t+1-i} \right)^2 = 2 A_{p+1-i}^{t+1-i} (\alpha_{p-i} - \alpha_{t+1-i}) + (\alpha_{p-i} - \alpha_{t+1-i})^2.$$

Summing over $i = h + 1, \dots, j$, we obtain

$$\left(A_{p-j}^{t-j} \right)^2 - \left(A_{p-h}^{t-h} \right)^2 \le 4 \sigma_{t-p+1} \left(\sum_{i=-\infty}^{\infty} \alpha_i^2 \right)^{1/2} + 4 \sum_{i=-\infty}^{\infty} \alpha_i^2, \tag{2.9}$$

because $\sigma_{t-p+1}^2 = \sum_{-\infty}^{\infty} (A_{p-i}^{t-i})^2$. The last relation indicates that we can choose h sufficiently negative, as a function of $t - p$, such that $|A_{p-h}^{t-h}| < K(\sigma_{t-p+1}^{1/2} + 1)$, say, for all $t - p \ge 0$, where K is a generic positive constant. It follows from (2.9) that $\max_j |A_{p-j}^{t-j}| \le K(\sigma_{t-p+1}^{1/2} + 1)$, and then (ii) is established by reference to (2.8). □

To illustrate the usefulness of both conditions (i) and (ii) in case Assumption 1 is also imposed, take the simple case of least squares polynomial time series regression, mentioned in Section 1. We have $w_{tn} = \phi_s(t)/n^{s+H}$, where ϕ_s is an sth degree polynomial. For $H \ge 1/2$ we easily check (i). For $0 < H < 1/2$, we can check (i) if we also assume $\alpha_j = o(j^{H-1})$, as is readily seen. Although this does not entail absolute summability of the α_j, it does, for example, imply $\gamma_j = o(j^{2H-1})$, and thus typically rule out the possibility that $f(\lambda) \sim |\lambda - \omega|^{1-2H'}$ as $\lambda \to \omega$, for some $\omega \neq 0 \pmod{2\pi}$, unless $H' < H + 1/2$; a spectrum can be zero at $\lambda = 0$ but elsewhere unbounded. However, the polynomial structure of ϕ_s and (1.12) enables us to check (ii) when $0 < H < 1/2$ without any assumption on the α_j besides the square summability in (2.1), which is merely equivalent to finite variance of u_t.

3. Central limit theorem for nonparametric regression estimates. Now consider the estimate (1.2) for $r(x)$ based on the model (1.1). We impose first a condition on the kernel k.

ASSUMPTION 3. Let $k(v)$ be even, eventually monotone nonincreasing in $|v|$, differentiable with derivative $k'(v)$, satisfying (1.3) and

$$k(v) = O\Big((1+v^2)^{-1}\Big), \qquad k'(v) = O\Big((1+|v|^{1+\eta})^{-1}\Big) \quad \text{for some } \eta > 0. \tag{3.1}$$

It would be possible to establish Theorem 1 under somewhat milder conditions on k, whose strength decreases as H increases. However, we prefer the simpler condition above, which we motivate by a worker willing to contemplate an unknown H that is anywhere in $(0, 1)$, and thus wishing to choose k accordingly. Kernels used in practice (including typical higher-order kernels) are eventually monotonically decreasing. We have avoided compact support assumptions on k, imposing instead tail conditions. The differentiability condition does strictly exclude kernels such as the uniform, but such kernels, which are smooth almost everywhere, could be covered by a slight modification of our proofs. The following lemma estimates the covariance structure of $\hat{r}$.

LEMMA 3. *Let* (1.1) *and Assumptions* 1 *and* 3 *hold, and let*

$$(nb)^{-1} + n^{1-2H}b^{3-2H} \to 0 \quad \text{as } n \to \infty. \tag{3.2}$$

Then for all $x, y \in (0, 1)$

$$(nb)^{2-2H}\operatorname{cov}\{\hat{r}(x), \hat{r}(y)\} \to G\rho(H)\Delta_{xy} \quad \text{as } n \to \infty, \tag{3.3}$$

where

$$\rho(H) = \begin{cases} \theta(H)\iint_{-\infty}^{\infty} k(v)\{k(w) - k(v)\}|v-w|^{2H-2}\,dv\,dw, & 0 < H < \frac{1}{2}, \\ 2\pi\int_{-\infty}^{\infty} k(v)^2\,dv, & H = \frac{1}{2}, \\ \theta(H)\iint_{-\infty}^{\infty} k(v)k(w)|v-w|^{2H-2}\,dv\,dw, & \frac{1}{2} < H < 1. \end{cases}$$

PROOF. For future use we observe that Assumption 3 implies

$$\sum_t |k_{xt}|^a = O(nb), \qquad a \geq 1, \tag{3.4}$$

where $k_{xt} = k((nx - t)/nb)$, so that k effectively behaves like a compactly supported kernel. The left-hand side of (3.3) is

$$(nb)^{-2H} \int_{-\pi}^{\pi} f(\lambda) \hat{k}_x(\lambda) \hat{k}_y(-\lambda)\, d\lambda, \tag{3.5}$$

where $\hat{k}_x(\lambda) = \sum_t k_{xt} e^{it\lambda}$. The difference between (3.5) and G times

$$(nb)^{-2H} \int_{-\pi}^{\pi} h(\lambda) \hat{k}_x(\lambda) \hat{k}_y(-\lambda)\, d\lambda \tag{3.6}$$

is bounded in absolute value, due to the triangle and Schwarz inequalities, by

$$(nb)^{-2H} \max_{|\lambda| < \varepsilon} |g(\lambda) - G| \times \left\{ \int_{-\pi}^{\pi} h(\lambda) \left| \hat{k}_x(\lambda) \right|^2 d\lambda \int_{-\pi}^{\pi} h(\lambda) \left| \hat{k}_y(\lambda) \right|^2 d\lambda \right\}^{1/2}$$

$$+ 2(nb)^{-2H} G \left\{ \int_{\varepsilon}^{\pi} h(\lambda) \left| \hat{k}_x(\lambda) \right|^2 d\lambda \int_{\varepsilon}^{\pi} h(\lambda) \left| \hat{k}_y(\lambda) \right|^2 d\lambda \right\}^{1/2} \tag{3.7}$$

$$+ 2(nb)^{-2H} \left\{ \int_{\varepsilon}^{\pi} f(\lambda) \left| \hat{k}_x(\lambda) \right|^2 d\lambda \int_{\varepsilon}^{\pi} f(\lambda) \left| \hat{k}_y(\lambda) \right|^2 d\lambda \right\}^{1/2} \tag{3.8}$$

for $\varepsilon \in (0, \pi)$, where h is nonnegative because

$$f(\lambda) = \left| \sum_{j=-\infty}^{\infty} \alpha_j e^{ij\lambda} \right|^2 \Big/ 2\pi \tag{3.9}$$

is, and because of Assumption 1. Consider (3.8). By summation by parts

$$\int_{\varepsilon}^{\pi} f(\lambda) \left| \hat{k}_x(\lambda) \right|^2 d\lambda = \int_{\varepsilon}^{\pi} f(\lambda) \left| \sum_{t=1}^{n-1} \triangle k_{xt} D_t(\lambda) + k_{xn} D_n(\lambda) \right|^2 d\lambda, \tag{3.10}$$

where $D_t(\lambda) = \sum_{s=1}^{t} e^{is\lambda}$. Because [see Zygmund (1977), page 51]

$$|D_t(\lambda)| \leq \frac{2}{\lambda}, \qquad 0 < \lambda \leq \pi \quad \text{for all } t \geq 1, \tag{3.11}$$

it follows that (3.10) is bounded by $4\gamma_0\{(\sum_{t=1}^{n-1} |\triangle k_{xt}|)^2 + k_{xn}^2\}/\varepsilon^2$. Choose $M > 0$ such that $k(v)$ is monotone nonincreasing in $|v|$ for $|v| \geq M$. For n sufficiently large $nx - nbM \geq 2$ and $n(1 - x) - nbM \geq 2$, and then

$$\sum_{t=1}^{[nx - nbM] - 1} |\triangle k_{xt}| + \sum_{t=[nx + nbM]+1}^{n-1} |\triangle k_{xt}| \leq 4|k(M)|.$$

On the other hand, bounded differentiability of k implies

$$\sum_{t=[nx-nbM]}^{[nx+nbM]} |\triangle k_{xt}| \leq KM.$$

It follows that (3.10) $= O(\varepsilon^{-2})$ so the contribution of (3.8) is negligible because $nb \to \infty$ and $H > 0$. The term (3.7) can be handled in the same way. Because ε is arbitrary and g is continuous at $\lambda = 0$, it clearly remains to be shown that

$$(3.6) \to \rho(H)\Delta_{xy} \quad \text{as } n \to \infty, \tag{3.12}$$

because $\sup_H \rho(H) < \infty$. Now (3.6) is

$$(nb)^{-2H} \sum_s \sum_t \delta_{t-s} k_{xt} k_{ys}. \tag{3.13}$$

First assume $x > y$, and put $z = (x - y)/4$. Write $A(x) = \{t\colon 1 \leq t \leq n, |t - nx| \leq nz\}$ and $B(x) = \{t\colon 1 \leq t \leq n, |t - nx| > nz\}$, and note that $|t - nx| \leq nz$ and $|s - ny| \leq nz$ implies $|s - t| \geq 2nz$. Thus because of $\delta_j = O(j^{2H-2})$ and (3.4),

$$\left| \frac{1}{(nb)^{2H}} \sum_{t \in A(x)} \sum_{s \in A(y)} \delta_{s-t} k_{xt} k_{ys} \right| \leq \frac{Kn^{2H-2}}{(nb)^{2H}} \sum_t |k_{xt}| \sum_s |k_{ys}| = O(b^{2-2H}) \to 0.$$

On the other hand, for $0 < H < 1/2$, we use (1.9) to write

$$\begin{aligned} &\frac{1}{(nb)^{2H}} \sum_{t \in A(x)} \sum_{s \in B(y)} \delta_{s-t} k_{xt} k_{ys} \\ &\quad = \frac{1}{(nb)^{2H}} \sum_{s \in B(y)} k_{ys} \sum_{t \in A(x)} \delta_{s-t}(k_{xt} - k_{xs}) \\ &\qquad - \frac{1}{(nb)^{2H}} \sum_{s \in B(y)} k_{ys} k_{xs} \sum_{t \in B(x)} \delta_{s-t}. \end{aligned} \tag{3.14}$$

The first term on the right-hand side is bounded in absolute value by

$$\begin{aligned} \frac{K}{(nb)^{2H+1}} \sum_{s \in B(y)} \left(\frac{nb}{ny - s} \right)^2 \sum_{|s-t| \leq n} |s - t|^{2H-1} &= O\left(\frac{(nb)^2}{(nb)^{2H+1}} \frac{n^{2H}}{n} \right) \\ &= O(b^{1-2H}). \end{aligned}$$

The second term is bounded by

$$\frac{K}{(nb)^{2H}} \sum_{s \in B(y)} \left(\frac{nb}{ny - s} \right)^2 |k_{xs}| \sum_{t=-\infty}^{\infty} |\delta_{s-t}| = O(n^{1-2H} b^{3-2H}),$$

using (3.4). Thus in view of (3.2) it follows that (3.14) tends to 0. For $1/2 \le H < 1$,

$$\left| \frac{1}{(nb)^{2H}} \sum_{t \in A(x)} \sum_{s \in B(y)} \delta_{s-t} k_{xt} k_{ys} \right| \le \frac{K}{(nb)^{2H}} \sum_{s \in B(y)} \left(\frac{nb}{ny - s} \right)^2 \sum_t |\delta_{s-t}| = O(b^{2-2H}),$$

which also tends to 0. Replacing the double summation on the left of (3.14) by $\sum_{t \in B(x)} \sum_{s \in A(y)}$ gives the same result. Finally

$$\left| \frac{1}{(nb)^{2H}} \sum_{t \in B(x)} \sum_{s \in B(y)} \delta_{s-t} k_{xt} k_{ys} \right| \le \frac{K}{(nb)^{2H}} \sum_{t \in B(x)} \left(\frac{nb}{ny - t} \right)^2 \sum_{s \in B(y)} \left(\frac{nb}{ny - s} \right)^2 |\delta_{s-t}|.$$

For $0 < H < \frac{1}{2}$, this is $O(n^{1-2H} b^{4-2H}) \to 0$, and for $\frac{1}{2} \le H < 1$ it is $O(b^{4-2H}) \to 0$. Thus the proof of (3.12) for $x > y$, and thus $x < y$, is completed. Now assume $x = y$. First let $0 < H < \frac{1}{2}$. In view of (1.9), (3.13) can then be written

$$\frac{1}{(nb)^{2H}} \sum_t \sum_s \delta_{s-t} k_{xt} (k_{xs} - k_{xt}) - \frac{1}{(nb)^{2H}} \sum_t k_{xt}^2 \left(\sum_{s \le 0} \delta_{s-t} + \sum_{s > n} \delta_{s-t} \right). \tag{3.15}$$

The second term is

$$O\left((nb)^{-2H} \sum_t \left\{ t^{2H-1} + (n - t + 1)^{2H-1} \right\} k_{xt}^2 \right). \tag{3.16}$$

For $M \in (0, x/2b)$

$$\frac{1}{(nb)^{2H}} \sum_t t^{2H-1} k_{xt}^2 \le \frac{1}{(nb)^{2H}} \max_{1 \le t \le nbM} k_{xt}^2 \sum_{t=1}^{[nbM]} t^{2H-1} + \frac{(nbM)^{2H-1}}{(nb)^{2H}} \sum_{t=[nbM]+1}^{n} k_{xt}^2.$$

By (3.4) the second term on the right is $O(M^{2H-1})$ as $n \to \infty$, and can be made arbitrarily small on making M large. For any M the first term approaches 0 as $n \to \infty$ because $(nx - nbM)/nb \to \infty$ and $k(v) \to 0$ as $v \to \infty$. The other part of (3.16) can be treated in the same way, so that (3.16) is $o(1)$. For $\varepsilon > 0$ put $\delta_j' = \theta(H)|j|^{2H-2}$ if $0 < |j| < nb/\varepsilon$, and $\delta_j' = \delta_j$ otherwise.

Then the first part of (3.15) differs from

$$\frac{1}{(nb)^{2H}} \sum_t \sum_s \delta'_{t-s} k_{xt}(k_{xs} - k_{xt}) \tag{3.17}$$

by

$$O\left(\frac{1}{(nb)^{2H+1}} \sum_t |k_{xt}| \sum_{0<|t-s|\le nb/\varepsilon} \zeta_{s-t} |s-t|^{2H-1} \right)$$
$$= O\left(\frac{1}{(nb)^{2H}} \sum_{j=1}^{[nb/\varepsilon]} \zeta_j j^{2H-1} \right) \to 0,$$

where $\zeta_j \to 0$ as $j \to \infty$ and we use the Toeplitz lemma. Then (3.17) differs from $\rho(H)$ by $\int\int_{-\infty}^{\infty} f_n(v,w)\,dv\,dw$, where

$$f_n(v,w) = \frac{\delta'_{t-s}}{(nb)^{2H-2}} k_{xt}(k_{xs} - k_{xt}) - \theta(H)|v-w|^{2H-2} k(v)\{k(w) - k(v)\}$$

for

$$\frac{nx-t-1}{nb} < v \le \frac{nx-t}{nb} \quad \text{and} \quad \frac{nx-s-1}{nb} < w \le \frac{nx-s}{nb}, \qquad s,t = 1,\dots,n,$$

and

$$f_n(v,w) = -\theta(H)|v-w|^{2H-2} k(v)\{k(w) - k(v)\}$$

for

$$v \le \frac{x-1}{n} - \frac{1}{nb} \quad \text{or} \quad v > \frac{x}{n} - \frac{1}{nb} \quad \text{or}$$
$$w \le \frac{x-1}{b} - \frac{1}{nb} \quad \text{or} \quad w > \frac{x}{b} - \frac{1}{nb}.$$

For almost all v,

$$\lim_{n\to\infty} \max_{t:\, 0 \le ((nx-t)/nb) - v < (1/nb)} |k_{xt} - k(v)| = 0$$

from Assumption 3 and (3.2), while from Assumption 1, for all $v \ne w$,

$$\lim_{n\to\infty} \max_{s,t:\, |((s-t)/nb)+w-v| \le 1/(nb)} \left| (nb)^{2-2H} \delta'_{t-s} - \theta(H)|v-w|^{2H-2} \right| = 0,$$

so that, for all fixed M, $f_n(v,w) \to 0$ as $n \to \infty$ for almost all v, w such that $|v-w| \le M$. For all sufficiently large n, $|f_n(v,w)| \le K|v-w|^{2H-1}$ from Assumption 3, and so

$$\left| \int\int_{|v-w|\ge M} f_n(v,w)\,dv\,dw \right| \le \frac{K}{(nb)^{2H}} \sum_t |k_{xt}| \sum_{|s-t|>nbM/2} |\delta'_{s-t}|$$
$$+ K \int\int_{|v-w|>M} |k(v)|\,|v-w|^{2H-2}\,dv\,dw,$$

and this is $O(M^{2H-1}) \to 0$ as $M \to \infty$. The proof is completed for $0 < H < \frac{1}{2}$. When $H = \frac{1}{2}$, (3.13) with $x = y$ is 2π times

$$\begin{aligned}\frac{1}{nb}\sum_t k_{xt}^2 &= \int_{-\infty}^{\infty} k(v)^2\,dv + \sum_t \int_{(nx-t-1)/nb}^{(nx-t)/nb} \left\{k_{xt}^2 - k(v)^2\right\} dv \\ &\quad - \int_{x/b}^{\infty} k(v)^2\,dv - \int_{-\infty}^{(x-1)/b} k(v)^2\,dv \\ &= \int_{-\infty}^{\infty} k(v)^2\,dv + O\left(\frac{1}{nb} + b^3\right)\end{aligned}$$

by straightforward use of Assumption 3. For $\frac{1}{2} < H < 1$, the proof of (3.12) is omitted because it is similar to, and simpler than, that already given for $0 < H < \frac{1}{2}$, and the same type of result has been obtained previously, albeit under somewhat different conditions [see Hall and Hart (1990); Csörgő and Mielniczuk (1995b, c)]. □

In case $H \geq \frac{1}{2}$, (3.2) entails only $nb \to \infty$ and $b \to 0$. To estimate the bias of $\hat{r}$ we impose the following.

ASSUMPTION 4. Either $r(x)$ satisfies a Lipschitz condition of degree τ, $0 < \tau \leq 1$, or $r(x)$ is differentiable with derivative satisfying a Lipschitz condition of degree $\tau - 1$, $1 < \tau \leq 2$.

The following lemma is standard and the proof is omitted.

LEMMA 4. *Under* (1.1) *with* $Eu_t = 0$, $t = 1, 2, \ldots,$ *and Assumptions* 3 *and* 4, *for all* $x \in (0,1)$

$$\begin{aligned}E\{\hat{r}(x) - r(x)\} &= O(b^\tau), && 0 < \tau \leq 1\\ &= O(b^\tau + n^{-1}), && 1 < \tau \leq 2.\end{aligned}$$

In order that the bias be small enough to permit centering at $r(x)$ in the central limit theorem, we impose the following.

ASSUMPTION 5. For the same τ as in Assumption 4,

$$(nb)^{-1} + n^{1-H}b^{1-H+\tau} \to 0 \quad \text{as } n \to \infty. \tag{3.18}$$

For given τ, the strength of Assumption 5 decreases in H, so that a given b requires less smoothness in r when $H > \frac{1}{2}$ compared to the usual case $H = \frac{1}{2}$, but more when $H < \frac{1}{2}$.

THEOREM 1. *Let* (1.1) *and Assumptions* 1–5 *hold. Then for any distinct* x_i, $i = 1, \ldots, I$, *in* $(0,1)$, *the* $(nb)^{1-H}\{\hat{r}(x_i) - r(x_i)\}$, $i = 1, \ldots, I$, *converge to independent* $N(0, G\rho(H))$ *variates as* $n \to \infty$.

PROOF. From Lemmas 3 and 4 and Assumption 5 we have that $(nb)^{1-H}E\{\hat{r}(x) - r(x)\} \to 0$ and (3.3) holds for all $x, y \in (0,1)$, noting that for all $\tau \in (0,2]$, (3.18) implies (3.2). By the Cramér–Wold device, it then remains to show that for all constants $h_1, \dots, h_I$ that are not all zero,

$$(nb)^{-H}\left\{G\rho(H)\sum_{i=1}^{I} h_i^2\right\}^{-1/2} \sum_t \left(\sum_{i=1}^{I} h_i k_{x_i t}\right) u_t \to_d N(0,1).$$

By Lemma 3, the left-hand side differs by $o_p(1)$ from $\sum_t w_{tn} u_t$, where

$$w_{tn} = (nb)^{-H}\nu_n^{-1}\sum_{i=1}^{I} h_i k_{x_i t},$$

where $\nu_n^2 = (nb)^{-2H}V(\sum_t\{\sum_{i=1}^{I} h_i k_{x_i t}\}u_t)$. Then (2.2) holds and we need to check (2.3). From Assumption 3, the left-hand side of (2.6) is, for $\frac{1}{2} < H < 1$,

$$O\left(\left\{\frac{1}{(nb)^{2H}\nu_n^2}\sum_{i=1}^{I}\sum_t k_{x_i t}^2\right\}^{1/2} + \frac{1}{(nb)^H \nu_n}\right) = O\left(\frac{(nb)^{1/2-H}}{\nu_n}\right)$$

from the proof of Lemma 3 and choosing $a \equiv 1$, say, whereas for $H = \frac{1}{2}$ we can choose a such that $a^{-1} + a/nb \to 0$ as $n \to \infty$ so that the left-hand side of (2.6) is

$$o\left(\left\{\frac{1}{nb\nu_n^2}\sum_{i=1}^{I}\sum_t k_{x_i t}^2\right\}^{1/2}\right) + O\left(\frac{a}{nb\nu_n}\right) = o(\nu_n^{-1}).$$

For $0 < H < \frac{1}{2}$ we consider (ii) of Lemma 2. Note first that $\sigma_n^2 \sim n^{2H}G\theta(H)/H(2H-1) \to \infty$ [see (1.12)]. For each i, introduce an increasing integer-valued sequence $z_i = z_{in} < \min([nx_i], n - [nx_i])$, and put $p_i = [nx_i - z_i]$, $q_i = [nx_i + z_i]$. Then

$$\begin{aligned}\sum_{t=1}^{p_i - 1} w_{itn}^2 &\le \frac{K}{(nb)^{2H}\nu_n^2} h_i^2 \sum_{t=1}^{p_i-1}\left(\frac{nb}{nx_i - t}\right)^4 \\ &\le \frac{K(nb)^{4-2H}}{\nu_n^2}\sum_{j=[nx_i - p_i]}^{\infty} j^{-4} \\ &= O\left(\frac{(nb)^{4-2H}}{\nu_n^2 z_i^3}\right),\end{aligned}$$

with the same bound for $\sum_{t=q_i+1}^{n} w_{itn}^2$. It follows from Assumption 3 that

$$\sum_{t=p_i}^{q_i} |\triangle w_{itn}| = O\left(\frac{(nb)^{-H}}{\nu_n}\right).$$

For $t \le q_i$, $\sigma_{t-p_i+1} = O(\sigma_{q_i - p_i + 1}) = O(z_i^H)$ as $n \to \infty$, so

$$\sum_{t=p}^{q} |\triangle w_{itn}|\left(\sigma_{t-p_i+1}^{1/2} + 1\right) = O\left(\frac{(z_i^{1/2}/nb)^H}{\nu_n}\right).$$

Clearly $|w_{iq_in}|(\sigma_{q_i-p_i+1}^{1/2}+1)=O((z_i^{1/2}/nb)^H/\nu_n)$ also. Taking $z_i \sim (nb)^{(4-H)/3}$, say (which is $o(n)$ under Assumption 5), it follows that the left-hand side of (2.7) is $o(\nu_n^{-1})$. Now Lemma 3 implies that $\nu_n \to G\rho(H)$ and so application of Lemma 2, for any H, requires $\rho(H) > 0$, because $G > 0$ is assumed. Clearly $\rho(\frac{1}{2}) > 0$ from (1.3). It is not immediately obvious that $\rho(H) > 0$ for $\frac{1}{2} < H < 1$ because we have not assumed that k is nonnegative. However, noting that for $v > 0$,

$$v^{\beta-1} = \frac{2}{\pi}\Gamma(\beta)\cos\left(\frac{\beta\pi}{2}\right)\int_0^\infty \lambda^{-\beta}\cos(\lambda v)\,d\lambda, \qquad 0 < \beta < 1,$$

and replacing v by $|v-w|$ and β by $2H-1$, we have after rearrangement and use of the reflection formula for the gamma function

$$\rho(H) = \int_{-\infty}^{\infty}|\lambda|^{1-2H}\left|\int_{-\infty}^{\infty}k(v)e^{iv\lambda}\,dv\right|^2 d\lambda, \qquad \tfrac{1}{2} < H < 1;$$

this is positive because (1.3) and Assumption 3 implies that the Fourier transform of k is not almost everywhere zero. To prove that $\rho(H) > 0$ for $0 < H < \frac{1}{2}$, note that

$$\rho(H) = -\frac{\theta(H)}{2}\iint_{-\infty}^{\infty}\{k(v)-k(w)\}^2|v-w|^{2H-2}\,dv\,dw, \qquad 0 < H < \frac{1}{2};$$

this is positive because $\theta(H) < 0$ for $0 < H < \frac{1}{2}$, and (1.3) and Assumption 3 imply that k is not almost everywhere constant. □

A similar end result was achieved by Roussas, Tran and Ioannides (1992), Csörgő and Mielniczuk (1995a), and Tran, Roussas, Yakowitz and Truong Van (1996), in case $H = \frac{1}{2}$, and by Csörgő and Mielniczuk (1995b, c), Deo (1997) in case $\frac{1}{2} < H < 1$. We previously indicated how their assumptions differ from ours in respect of their global implications for f. In other respects, their conditions are sometimes weaker, sometimes stronger than ours; there is substantial scope for trade-offs between the assumptions on u_t, r, k and b.

We can partially extend Theorem 2.2 of Hall and Hart (1990) by assuming that r is twice continuously differentiable with second derivative r'' and impose (3.2) rather than (3.18): an easy extension of Lemma 4 gives

$$E\{\hat{r}(x)-r(x)\}^2 \sim (nb)^{2H-2}G\rho(H) + b^4k_2r''(x)^2/4,$$

where $k_2 = \int v^2k(v)\,dv$, and thence the "optimal bandwidth"

$$\hat{b} = \left\{(2-2H)G\rho(H)/k_2r''(x)^2\right\}^{1/(6-2H)}n^{(H-1)/(3-H)}.$$

The optimal rate has exponent which tends to $-1/3$ as $H \downarrow 0$, and increases in H to 0 as $H \uparrow 1$. Ray and Tsay (1996) discuss a practical procedure for the choice of b.

4. Studentization. In practice G and H will be unknown and estimates will have to be inserted in the approximate variance formula implied by

Theorem 1. In order to ensure that such studentization does not affect the limiting distribution we first present a lemma.

LEMMA 5. *Under Assumption* 3, $\rho(H)$ *is continuous on* $(0,1)$.

PROOF. Write $\rho(H)$, for $H \neq 1/2$, as

$$\rho(H) = \theta(H)\int_{-\infty}^{\infty} \kappa(v)|v|^{2H-2}\,dv,$$

where

$$\kappa(v) = \int_{-\infty}^{\infty} \{k(v+w) - k(w)\}k(w)\,dw, \qquad 0 < H < \tfrac{1}{2},$$
$$= \int_{-\infty}^{\infty} k(v+w)k(w)\,dw, \qquad \tfrac{1}{2} < H < 1.$$

Clearly $\theta(H)$ is continuous on $(0,1)$. Evenness of k implies evenness of κ. Boundedness and integrability of k imply boundedness of κ. For $v > 0$,

$$|\kappa(v)v^{2H-2}| \le Kv^{2H-2}\{vI(0 \le v \le 1) + I(v > 1)\}, \qquad 0 < H < \tfrac{1}{2}, \tag{4.1}$$

$$|\kappa(v)v^{2H-2}| \le Kv^{2H-2}I(0 \le v \le 1) + \int |k(v+w)k(w)|\,dw I(v>1), \qquad \tfrac{1}{2} < H < 1, \tag{4.2}$$

where bounded differentiability of k is used in (4.1) and the right-hand sides of (4.1) and (4.2) are both integrable in view of the respective values of H concerned and the integrability of k. Because v^{2H-1} is continuous in H for $v > 0$, the lemma is proved for $H \neq 1/2$ by dominated convergence. To prove continuity at $H = 1/2$, note first that for $H \neq 1/2$,

$$\theta(H)\int_0^{\infty} \kappa(v)v^{2H-2}\,dv = \theta(H)\left[\frac{\kappa(v)v^{2H-1}}{2H-1}\right]_0^{\infty} - \frac{\theta(H)}{2H-1}\int_0^{\infty} \kappa'(v)v^{2H-1}\,dv, \tag{4.3}$$

where $\kappa'(v) = \int_{-\infty}^{\infty} k'(v+w)k(w)\,dw$ satisfies

$$|\kappa'(v)| \le \int_{-\infty}^{-v/2} |k'(v+w)k(w)|\,dw + \int_{-v/2}^{\infty} |k'(v+w)k(w)|\,dw$$
$$= O\Big((1+v^2)^{-1} + (1+|v|^{1+\eta})^{-1}\Big).$$

Thus $|\kappa'(v)v^{\xi}|$ is bounded by the integrable function $v^{\xi}/(1+v^{1+\eta})$, when $|\xi| \le \eta/2$, say. Thus by dominated convergence, as $H \to \frac{1}{2}$,

$$\int_0^{\infty} \kappa'(v)v^{2H-1}\,dv \to \int_0^{\infty} \kappa'(v)\,dv = -\int_{-\infty}^{\infty} k^2(v)\,dv.$$

Because

$$\cos\{\pi(1-H)\}/(2H-1) \to \pi/2 \quad \text{as } H \to \tfrac{1}{2}, \tag{4.4}$$

it follows that the last term of (4.3) tends to $\rho(1/2)/2$. The proof is completed on noting that the first term on the right-hand side of (4.3) is zero, using (4.4) and also (4.1) in case of $H \uparrow 1/2$ and

$$|\kappa(v)| \le \int_{-v/2}^{\infty} |k(v+w)k(w)|\,dw + \int_{-\infty}^{-v/2} |k(v+w)k(w)|\,dw = O(v^{-2})$$

as $v \to \infty$ in case $H \downarrow 1/2$. □

Now suppose we have estimates $\hat{G}$, $\hat{H}$ satisfying the following assumption.

ASSUMPTION 6.

$$\hat{G} \to_p G, \qquad (\log nb)(\hat{H} - H) \to_p 0 \quad \text{as } n \to \infty. \tag{4.5}$$

The following theorem is a simple application of Theorem 1, Lemma 5 and Slutsky's lemma, and the inequality $|(nb)^{\hat{H}-H} - 1| \le |\hat{H} - H| \log nb (\exp\{|\hat{H} - H| \log nb\})$ for $nb > 1$.

THEOREM 2. *Let* (1.1) *and Assumptions* 1–6 *hold. Then for any distinct* x_i, $i = 1, \dots, I$, *in* $(0,1)$ *the* $\{\hat{G}\rho(\hat{H})\}^{-1/2}(nb)^{1-\hat{H}}\{\hat{r}(x_i) - r(x_i)\}$, $i = 1, \dots, I$, *converge to independent* $N(0,1)$ *variables as* $n \to \infty$.

We now discuss estimation of H and G. The mildness of Assumption 6 indicates little incentive for basing estimates on a full parametric model for $f(\lambda)$ across $(-\pi, \pi]$, such as a fractionally integrated autoregressive moving average (FARIMA) model, which would typically lead to estimates that are $\sqrt{n}$-consistent if the autoregressive and moving average orders are correctly specified [Fox and Taqqu (1986)], but inconsistent otherwise. As the discussion of Section 1 suggests, it is more appropriate to base estimates on Assumption 1 or (1.7). Several such estimates (based on observable u_t) have been justified as less than $\sqrt{n}$-consistent but to have convergence rates which can satisfy Assumption 6. The estimate suggested by Künsch (1987) seems, on the basis of results of Robinson (1995b), to have desirable large sample properties. To adapt this, we suggest as proxies for the u_t:

$$\hat{u}_t = \left(y_t - (nc)^{-1} \sum_s \ell_{t-s} y_s\right) i_t, \tag{4.6}$$

where $i_t = I(nc < t \le n - [nc])$, $\ell_{t-s} = \ell((t-s)/nc)$, $\ell(v)$ is a kernel function [not necessarily identical to $k(v)$] satisfying

$$\ell(v) = 0 \quad \text{for } |v| > 1, \qquad \int_{-1}^{1} \ell(v)\,dv = 1, \tag{4.7}$$

and c is a positive bandwidth number (not necessarily equal to b). Without taking the $\hat{u}_t$ to be zero close to the boundaries, and the compact support assumption on ℓ, an additional term would arise in the proof of Theorem 3 below which rules out the existence of a suitable c sequence when $0 < H < 1/5$, and entails a narrower band of acceptable c's for other H. The fact that problems can arise with standard kernel estimates near boundaries is familiar since the work of Rice (1984a). For sequences p_t, q_t, define

$$w_p(\lambda) = \frac{1}{(2\pi n)^{1/2}} \sum_t p_t e^{it\lambda}, \qquad I_{pq}(\lambda) = w_p(\lambda) w_q(-\lambda). \tag{4.8}$$

For $\lambda_j = 2\pi j/n$ and integer $m \in [1, n/2)$, let

$$\begin{aligned} \tilde{G}(\psi) &= \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2\psi-1} I_{\hat{u}\hat{u}}(\lambda_j), \\ \tilde{R}(\psi) &= \log \tilde{G}(\psi) - (2\psi - 1) \sum_{j=1}^{m} \log \lambda_j, \end{aligned} \tag{4.9}$$

and define $\hat{G}$ and $\hat{H}$ as

$$\hat{G} = \tilde{G}(\hat{H}), \qquad \hat{H} = \underset{\psi \in [\Delta_1, \Delta_2]}{\arg\min}\, \tilde{R}(\psi), \tag{4.10}$$

for $0 < \Delta_1 < \Delta_2 < 1$. In particular, one can take $\Delta_1 = 1 - \Delta_2 = \varepsilon$ for some small positive ε if the admissible region includes negative-dependent and short-range dependent H values, as well as long-range dependent ones.

We introduce the following further assumptions.

ASSUMPTION 7. For finite constants μ_3 and μ_4,

$$E(\varepsilon_t^3 | F_{t-1}) = \mu_3 \quad \text{a.s.}; \qquad E(\varepsilon_t^4 | F_{t-1}) = \mu_4, \qquad t = 0, \pm 1, \dots .$$

ASSUMPTION 8. In a neighborhood $(0, \delta)$ of the origin, $\alpha(\lambda) = \sum_{j=-\infty}^{\infty} \alpha_j e^{ij\lambda}$ is differentiable and $(d/d\lambda)\alpha(\lambda) = O(|\alpha(\lambda)|/\lambda)$ as $\lambda \to 0+$.

ASSUMPTION 9. For some $\beta \in (0, 2]$,

$$f(\lambda) \sim G\lambda^{1-2H}(1 + O(\lambda^{\beta})) \quad \text{as } \lambda \to 0+, \tag{4.11}$$

where $G > 0$ and $H \in [\Delta_1, \Delta_2]$.

ASSUMPTION 10. Satisfying (4.7), ℓ is even and boundedly differentiable.

ASSUMPTION 11. As $n \to \infty$,

$$(\log n)^4 \left\{ \left(\frac{m}{n} \right)^{4\beta} + \frac{(\log m)^2}{m} + \frac{1}{cm} + c + \frac{c^4 n^{2-2H}}{m^{1-2H}} + \frac{m^{2H-1}}{c^2 n^{2H}} \right\} \to 0.$$

Assumptions 7–9 are taken from Robinson (1995b), where we note that $E(\varepsilon_t^4) = \mu_4$ is insufficiently assumed, instead of $E(\varepsilon_t^4 | F_{t-1}) = \mu_4$. Both Assumptions 8 and 9 hold (with $\beta = 2$) in case of FARIMA processes; these latter also satisfy Assumption 1, which the earlier discussion partially related to (4.11).

THEOREM 3. *Let* (1.1) *and Assumptions* 2, 4 *with* $\tau = 2$, *and* 7–11 *hold, and let* $\hat{H}$ *and* $\hat{G}$ *be given by* (4.6), (4.9) *and* (4.10). *Then* (4.5) *follows.*

The proof of this theorem is extremely technical, and is relegated to the Appendix.

To interpret the joint impact of Assumptions 5 and 11 when Theorems 2 and 3 are combined, suppose that $b \sim n^{-\eta}$, $c \sim n^{-\zeta}$, $m \sim n^{\rho}$. Then Assumption 5 for $\tau = 2$ and Assumption 11 hold when

$$\beta > 0, \qquad \frac{1-H}{3-H} < \eta < 1, \qquad \zeta < \rho < 1,$$

$$\tfrac{1}{2}(1-H) + \tfrac{1}{4}(2H-1)\rho < \zeta < H - \tfrac{1}{2}(2H-1)\rho.$$

Because H is unknown, it is useful to deduce conditions that hold for all $H \in (0,1)$:

$$\beta > 0, \qquad \tfrac{1}{3} \le \eta < 1, \qquad \tfrac{2}{3} \le \rho < 1, \qquad \tfrac{1}{2}\left(1 - \tfrac{1}{2}\rho\right) \le \zeta \le \tfrac{1}{2}\rho,$$

implying that $\zeta = 1/3$ when $\rho = 2/3$ and we cannot choose ζ outside $(1/4, 1/2)$ for any $\rho \in [2/3, 1)$.

The residual computation (4.6) is rather heavy, and also requires choice of c and ℓ. In the context of independent u_t, Rice (1984b) indicated that the error variance can be estimated without computation of nonparametric residuals, but by a differencing of the raw data, while Müller and Stadtmüller (1987, 1988) considered extensions to more general models. Perhaps more surprisingly, we now show that when long-range dependence, $H > 1/2$, can be taken for granted, it is possible to satisfy (4.5) even when $\hat{G}$ and $\hat{H}$ are computed from the raw data y_t; that is, we take

$$\hat{u}_t = y_t, \qquad 1 \le t \le n. \tag{4.12}$$

ASSUMPTION 12. For $H \ge \Delta_1 > 1/2$ and some $\delta > 0$, as $n \to \infty$

$$(\log n)^4 \left(\frac{m}{n} \right)^{4\beta} + \frac{n^{2-2H}}{m^{1-2\max(\delta, H-\Delta_1)}} \to 0.$$

THEOREM 4. *For* $H \in (\frac{1}{2}, 1)$, *let* (1.1) *and Assumptions* 2, 4 *with* $\tau = 2$, 7–9, *and* 12 *hold, and let* $\hat{H}$ *and* $\hat{G}$ *be given by* (4.9), (4.10) *with* $\Delta_1 > 1/2$, *and* (4.12). *Then* (4.5) *follows.*

Again the proof is left to the Appendix. Taking $b \sim n^{-\eta}$, $m \sim n^{\rho}$ as before, we now require $(2-2H)/\{1-2\max(\delta, H-\Delta_1)\} < \rho < 1$. While such ρ exist whenever $1/2 < \Delta_1 \le H < 1$, the admissible set is very narrow when H is close to 1/2.

5. Empirical application to Nile data. Some of the earliest theoretical development of long-range dependence was prompted by empirical studies of the Nile River data, which has since routinely illustrated new methods of estimating H. These data consist of readings of annual minimum levels at the Roda gorge near Cairo, commencing in the year 622; often only the first 663 observations are employed because missing observations occur after the year 1284 [see Toussoun (1925)]. It was one of the hydrological series examined by Hurst (1951) which led to his recognition of the "Hurst effect" and invention of the R/S statistic. The series provides evidence of long periods of unusually high or low precipitation, named the "Joseph effect" by Mandelbrot and Wallis (1968), who argued that stationary long-range dependent models like (1.7) are appropriate for such data, which are notably cyclic but not periodic. Subsequently, estimates of H for the Nile data were obtained by various methods by such authors as Graf (1983), Beran (1992) and Robinson (1995b).

On the other hand, it has been suggested that the phenomena noted by these authors could also be symptomatic of forms of nonstationarity; see, for example, Klemes (1974), Bhattacharya, Gupta and Waymire (1983) and Teverovsky and Taqqu (1997), who considered the possibility of a time-varying mean, and Beran and Terrin (1996), who considered piecewise stationarity with H varying over time.

Application of our present methods to the Nile series between 622 and 1284 provides evidence of nonstationarity in the mean and also an illustration of the consequences of Studentizing with an estimated H rather than by the conventional method that assumes $H = 1/2$, and of the desirability of allowing for the possibility of negative dependent, as well as long-range dependent, H.

We computed $\hat{r}(x)$ (1.2) for $x = i/30$, $1 \le i \le 29$, with

$$\begin{aligned} k(v) &= \tfrac{1}{2}\{1 + \cos(\pi v)\}, \qquad |v| \le 1, \\ &= 0, \qquad |v| > 1, \end{aligned}$$

which satisfies Assumption 3 [note that $k(v)$ is differentiable at $|v| = 1$, $k'(v)$ tending to 0 as $|v| \uparrow 1$ and being zero for $|v| > 1$]. In the Studentization we took $\ell(v) = k(v)$, so Assumption 10 holds also. We took $b = c$, considering each of the values 0.05, 0.075 and 0.1 [these would all result in estimates $\hat{r}(i/30)$ that would be exactly independent across i if y_t were independent across t, and not merely asymptotically independent as Theorem 1 implies]. In $\hat{H}$ given by (4.10), we took $m = 82$ [one of the values used by Robinson (1995b) for the same data and estimates]. We tried both choices (4.6) and (4.12) of $\hat{u}_t$ and computed for $x = i/30$, $i = 1, \dots, 29$, the $100(1-2\alpha)\%$

pointwise confidence intervals

$$\hat{r}(x) \pm z_{\alpha}\{\hat{G}\rho(\hat{H})\}^{1/2}(nb)^{\hat{H}-1} \tag{5.1}$$

and

$$\tilde{r}(x) \pm z_{\alpha}\{\tilde{G}(\tfrac{1}{2})\rho(\tfrac{1}{2})\}^{1/2}(nb)^{-1/2}, \tag{5.2}$$

for $\alpha = 0.05$ and 0.025, where $P(Z > z_{\alpha}) = \alpha$ for a standard normal variate Z.

In case $\hat{u}_t = y_t$, we have $\hat{H} = 0.905$ [as in Robinson (1995b)] and $\hat{G} = 0.076$. For brevity we display only the results for $b = c = 0.05$ with 90% intervals; see Figure 1. The solid line is the interpolated $\hat{r}(x)$, the broken lines indicate the interpolated (5.1) and the dotted lines the interpolated (5.2); we stress that the latter are not simultaneous bands; these are far wider, replacing $z_{0.05} = 1.645$ by $z_{0.002} = 2.88$. There is clearly evidence of changes in level, though use of the conventional $H = 1/2$ interval (5.2) substantially overstates their significance relative to (5.1). In case $\hat{u}_t$ are the modified residuals (4.6), $\hat{H}$ and $\hat{G}$ vary with c: for $c = 0.1$, $\hat{H} = 0.6928$, $\hat{G} = 0.094$; for $c = 0.075$, $\hat{H} = 0.614$, $\hat{G} = 0.114$; for $c = 0.05$, $\hat{H} = 0.407$, $\hat{G} = 0.169$. These estimates thus vary greatly over the range of smoothing employed, $\hat{H}$ decreasing with c to the extent even that a negative dependent $\hat{H}$ occurs when $c = 0.05$. Figure 2 displays the results in the latter case (with $b = 0.05$ and $\alpha = 0.05$ again). The intervals are far narrower (note the difference in scale from Figure 1) and now the ones for (5.1) are narrower than those for (5.2). The conclusions suggested by the other choices of b, c and α are qualitatively similar, though of course the larger b produce smoother $\hat{r}(x)$. This study highlights the need for developing methods for choosing b and c which respond automatically to the strength of the dependence in u_t.

APPENDIX

Proofs of Theorems 3 and 4.

PROOF OF THEOREM 3.

(i) *Plan of proof*. We show first that it suffices to prove that, as $n \to \infty$,

$$\sum_{i=1}^{m-1}\left(\frac{i}{m}\right)^{2(\Delta-H)+1}\frac{1}{i^2}\left|\sum_{j=1}^{i} d_j\right| \to_p 0, \tag{A.1}$$

$$(\log n)^2\sum_{i=1}^{m-1}\left(\frac{i}{m}\right)^{1-2\delta}\frac{1}{i^2}\left|\sum_{j=1}^{i} d_j\right| \to_p 0, \tag{A.2}$$

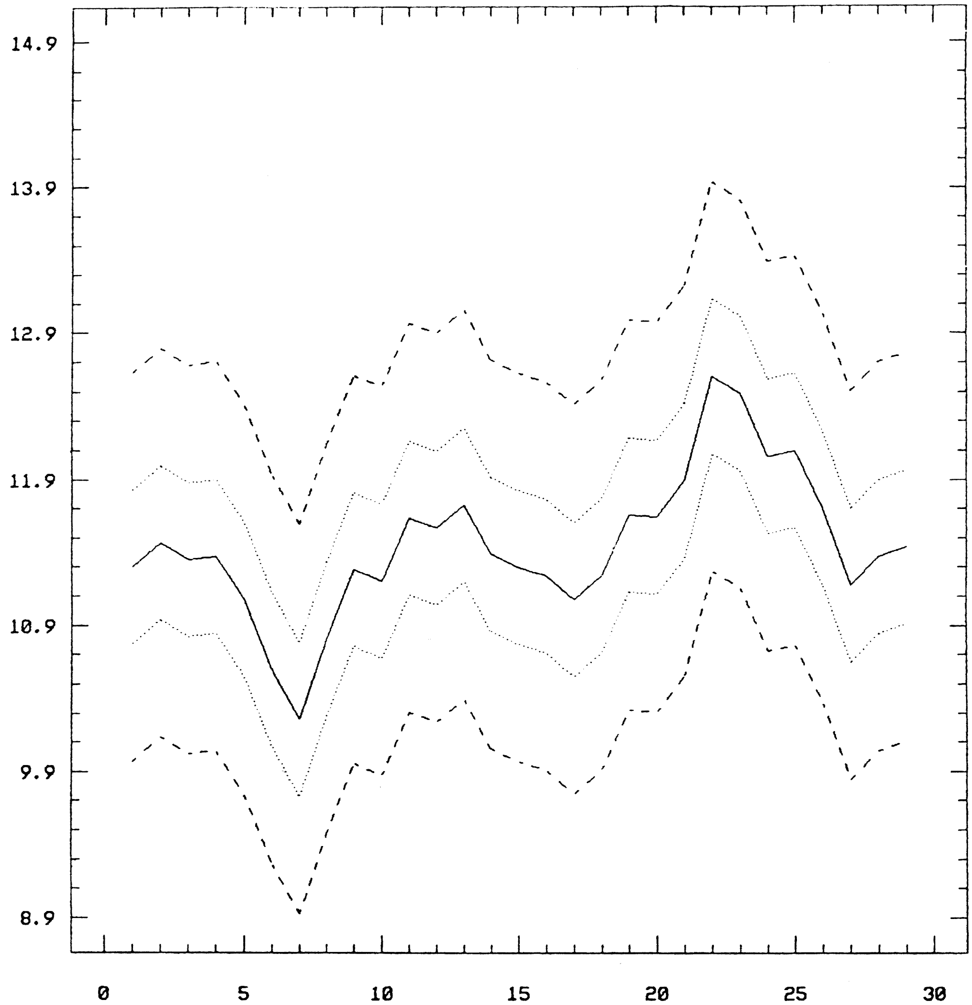

FIG. 1. *Nonparametric regression and 90% interval estimates for Nile data based on $b = c = 0.05$ and* (4.12): $\hat{r}(x)$ —; (5.1) - - - - -; (5.2) ·····.

$$\frac{(\log n)^2}{m} \sum_{j=1}^{m} d_j \to_p 0, \tag{A.3}$$

$$\frac{1}{m} \sum_{j=1}^{m} (a_j - 1) d_j \to_p 0, \tag{A.4}$$

where $\Delta = \Delta_1$ when $H < \frac{1}{2} + \Delta_1$ and $\Delta \in (H - \frac{1}{2}, H]$ otherwise; δ is an arbitrarily small positive number; $d_j = \{I_{\hat{u}\hat{u}}(\lambda_j) - I_{uu}(\lambda_j)\}/g_j$, $g_j = G\lambda_j^{1-2H}$; $a_j = (j/h)^{2(\Delta - H)}$, $1 \le j \le h$; $a_j = (j/h)^{2(\Delta_1 - H)}$, $h < j \le m$, $h =$

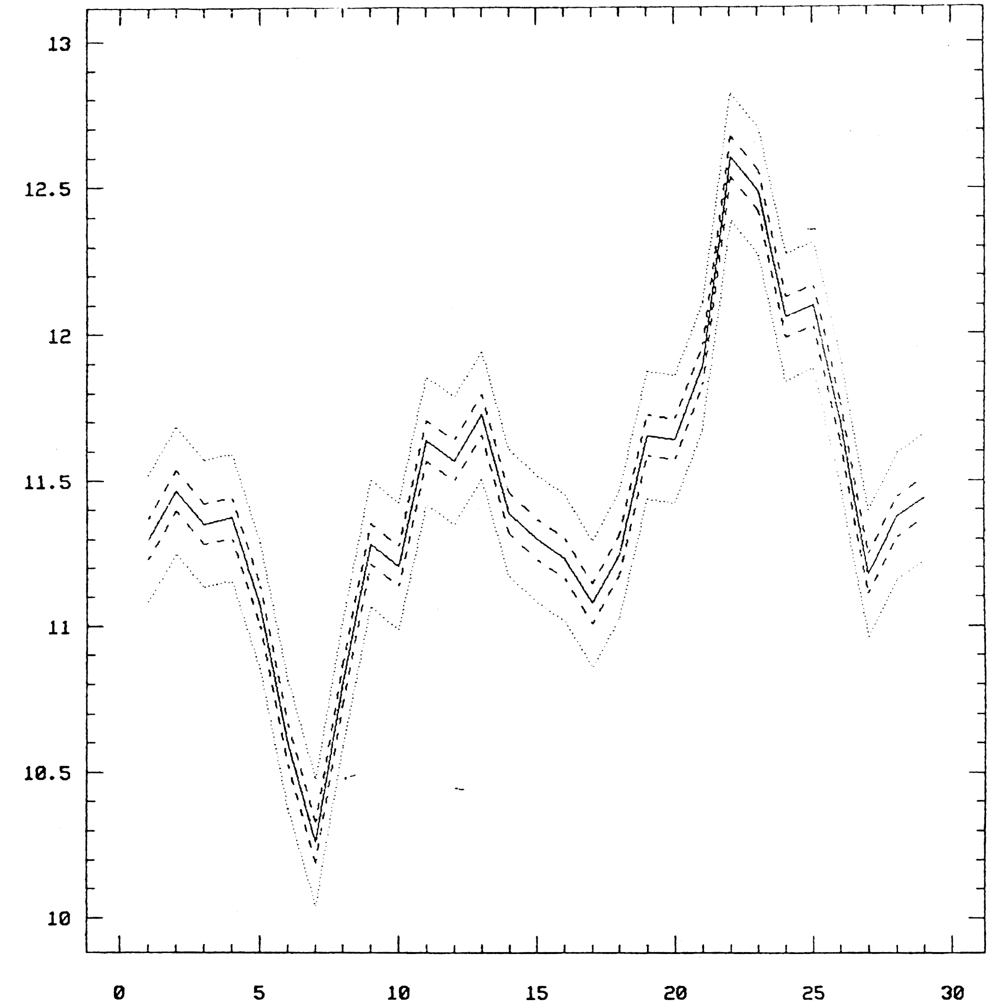

FIG. 2. *Nonparametric regression and* 90% *interval estimates for Nile data based on* $b = c = 0.05$ *and* (4.6): $\hat{r}(x)$ —; (5.1) - - - - -; (5.2) ·····.

$\exp(m^{-1}\sum_{j=1}^{m}\log j)$. Write $\hat{u}_t = u_t + v_t$, where $v_t = \chi_t + \theta_t + \zeta_t + \xi_t$ and $\chi_t = u_t(i_t - 1)$, $\theta_t = -(nc)^{-1}\sum_s \ell_{t-s} u_s i_t$, $\zeta_t = (nc)^{-1}\sum_s (r_t - r_s)\ell_{t-s} i_t$, $\xi_t = r_t\{1 - (nc)^{-1}\sum_s \ell_{t-s}\} i_t$, where $r_t = r(t/n)$. By elementary inequalities

$$\text{(A.5)}\quad |I_{\hat{u}\hat{u}} - I_{uu}| \le 2\{I_{uu}I_{vv}\}^{1/2} + I_{vv}, \qquad I_{vv} \le 4(I_{\chi\chi} + I_{\theta\theta} + I_{\zeta\zeta} + I_{\xi\xi})$$

(suppressing reference to the argument λ), where here and below subscripted I and w are defined as in (4.8). We shall then estimate $I_{vv}(\lambda_j)/g_j$, and thence verify (A.1)–(A.4).

(ii) *Proof of sufficiency of* (A.1)–(A.4). This makes heavy use of the proofs of Theorems 1 and 2 of Robinson (1995b), which in turn use results of Robinson (1995a). We have

$$|\hat{G} - G| \le \left|\tilde{G}(\hat{H}) - \tilde{G}(H)\right| + \left|\tilde{G}(H) - \hat{G}(H)\right| + \left|\hat{G}(H) - G\right|,$$

where $\hat{G}(H) = m^{-1}\sum_{j=1}^{n} \lambda_j^{2H-1} I_{uu}(\lambda_j)$. The last term on the right is $o_p(1)$ from Robinson (1995b), the middle term is $o_p(1)$ by the remainder of the current proof, and the first term is bounded by

$$\max_{1 \le j \le m} \left| \lambda_j^{2(H-\hat{H})} - 1 \right| \tilde{G}(H) = O_p\left(\log n |\hat{H} - H|\right)$$

[cf. the proof of Theorem 5 of Robinson (1994a)], so we shall actually show that $(\log n)(\hat{H} - H) \to_p 0$.

By a standard type of argument for proving consistency of implicitly defined extremum estimates,

$$P\left(\log n |\hat{H} - H| \ge \varepsilon\right) \le P\left(\inf_{\Theta \cap M} \tilde{S}(\psi) \le 0\right), \tag{A.6}$$

where $M = \{\psi\colon \log n |\psi - H| > \varepsilon\}$ for $\varepsilon \in (0, \frac{1}{2}\log n)$, and $\tilde{S}(\psi) = \tilde{R}(\psi) - \tilde{R}(H)$. As in Robinson (1995b), define $\Theta_1 = \{\psi\colon \Delta \le \psi \le \Delta_2\}$ and also $\Theta_2 = \{\psi\colon \Delta_1 \le \psi \le \Delta\}$ when $H \ge \frac{1}{2} + \Delta_1$, and to be empty otherwise. For $\delta \in (\varepsilon/\log n, \frac{1}{2})$, define $N_\delta = \{\psi\colon |\psi - H| < \delta\}$, $\overline{N}_\delta = (-\infty, \infty) \setminus N_\delta$, so that (A.6) is bounded by

$$P\left(\inf_{\Theta_1 \cap \overline{N}_\delta} \tilde{S}(\psi) \le 0\right) + P\left(\inf_{\Theta_1 \cap N_\delta \cap M} \tilde{S}(\psi) \le 0\right) + P\left(\inf_{\Theta_2} \tilde{S}(\psi) \le 0\right). \tag{A.7}$$

Now $\tilde{S}$ is S of Robinson (1995b) with $\hat{G}(\psi) = m^{-1}\sum_{j=1}^{m} \lambda_j^{2\psi-1} I_{uu}(\lambda_j)$ replaced by $\tilde{G}(\psi)$. By the arguments in Theorem 1 of Robinson (1995b), the first two probabilities in (A.7) tend to 0 if

$$\sup_{\Theta_1} \left| \frac{\tilde{G}(\psi) - G(\psi)}{G(\psi)} \right| + (\log n)^2 \sup_{\Theta_1 \cap N_\delta} \left| \frac{\tilde{G}(\psi) - G(\psi)}{G(\psi)} \right| \to_p 0,$$

where $G(\psi) = (G/m)\sum_{j=1}^{m} \lambda_j^{2(\psi-1)}$. By the triangle inequality this is implied if

$$\sup_{\Theta_1} \left| \frac{\hat{G}(\psi) - G(\psi)}{G(\psi)} \right| + (\log n)^2 \sup_{\Theta_1 \cap N_\delta} \left| \frac{\hat{G}(\psi) - G(\psi)}{G(\psi)} \right| \to_p 0, \tag{A.8}$$

$$\sup_{\Theta_1} \left| \frac{\tilde{G}(\psi) - \hat{G}(\psi)}{G(\psi)} \right| + (\log n)^2 \sup_{\Theta_1 \cap N_\delta} \left| \frac{\tilde{G}(\psi) - \hat{G}(\psi)}{G(\psi)} \right| \to_p 0. \tag{A.9}$$

Now (A.8) is proved with only minor modifications in Robinson (1995b), whereas (A.9) is implied if

$$\sup_{\Theta_1} \left| \frac{1}{m} \sum_{j=1}^{m} \left(\frac{j}{m}\right)^{2(\psi-H)} d_j \right| + (\log n)^2 \sup_{\Theta_1 \cap N_\delta} \left| \frac{1}{m} \sum_{j=1}^{m} j^{2(\psi-H)} d_j \right| \to_p 0. \tag{A.10}$$

By summation by parts and the inequalities $2(\psi - H) + 1 \ge 2(\Delta - H) - 1 > 0$ and, for $r > 0$, $|(1 + 1/i)^{2(\psi - H)} - 1| \le 2/i$ on Θ_1, it follows that (A.10) is implied by (A.1)–(A.3). For $H \le \frac{1}{2} + \Delta_1$, (A.1)–(A.3) suffice. For $H > \frac{1}{2} + \Delta_1$, the last probability in (A.7) can be nonzero, but following the argument in Robinson (1995b), it is bounded by

$$P\left(\left|\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\left\{\frac{I_{\hat{u}\hat{u}}(\lambda_j)}{g_j} - 1\right\}\right| \ge 1\right)$$

$$\le P\left(\left|\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\left\{\frac{I_{uu}(\lambda_j)}{g_j} - 1\right\}\right| \ge \frac{1}{2}\right)$$

$$+ P\left(\left|\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)d_j\right| \ge \frac{1}{2}\right).$$

The first of the last two probabilities tends to zero by the proof in Robinson (1995b), so that (A.1)–(A.4) suffice for $H > \frac{1}{2} + \Delta_1$.

(iii) *Estimation of* $I_{vv}(\lambda_j)/g_j$. Bearing in mind the second inequality in (A.5), we first consider $I_{\theta\theta}$. We have

$$EI_{\theta\theta}(\lambda) = (2\pi n^3 c^2)^{-1}\int_{-\pi}^{\pi} f(\mu)\left|J_\lambda(\mu)\right|^2 d\mu,$$

where $J_\lambda(\mu) = \sum_t i_t \sum_s \ell_{t-s} e^{it\lambda - is\mu}$. Now

$$J_\lambda(\mu) = \sum_t i_t \sum_s \ell_{t-s} e^{it\lambda - is\mu} = \left\{\sum_{|t| \le nc} \ell_t e^{it\mu}\right\}\left\{\sum_{nc < t \le n - [nc]} i_t e^{it(\lambda - \mu)}\right\}$$

by Assumption 10 and the definition of i_t, so that

$$|J_\lambda(\mu)| \le Knc\left|D_{n-2[nc]}(\lambda - \mu)\right|, \tag{A.11}$$

where $D_n(\lambda)$ is defined below (3.10). Also

$$\int_{-\pi}^{\pi}|J_\lambda(\mu)|^2 d\mu = 2\pi\sum_s\left|\sum_t i_t \ell_{t-s} e^{it\lambda}\right|^2, \tag{A.12}$$

which is bounded by

$$2\pi\sum_s\left|\sum_{t=1}^{n-1} i_t(\ell_{t-s} - \ell_{t+1-s})\sum_{v=1}^{t} i_v e^{iv\lambda} + \ell_{n-s}\sum_t i_t e^{it\lambda}\right|^2 \le K\frac{n}{\lambda^2}$$

for $0 < \lambda < \pi$ from (3.11), and also by Kn^3c^2 because ℓ has compact support. For $\lambda \in (0, \pi)$, $EI_{\theta\theta}(\lambda)$ is bounded by

$$\frac{1}{2\pi n^3 c^2}\int_{\lambda/2}^{3\lambda/2}|f(\mu) - f(\lambda)|\,|J_\lambda(\mu)|^2 d\mu \tag{A.13}$$

$$+\frac{1}{2\pi n^3c^2}\left\{\int_{3\lambda/2}^{\pi}+\int_{-\pi}^{\lambda/2}\right\}f(\mu)|J_\lambda(\mu)|^2\,d\mu \tag{A.14}$$

$$+\frac{f(\lambda)}{\pi n^3c^2}\int_{-\pi}^{\pi}|J_\lambda(\mu)|^2\,d\mu. \tag{A.15}$$

It follows from both bounds for (A.12) that (A.15) is $O(f(\lambda)\min\{1,(nc\lambda)^{-2}\})$ as $\lambda\to 0+$. We split (A.14) into several components. For all $H\in(0,1)$ and sufficiently small λ, there exists $\varepsilon\in(3\lambda/2,\pi)$ such that $f(\mu)/\mu^{1-H}=O(f(\lambda)/\lambda^{1-H})$, for $\lambda<\mu<\varepsilon$. Thus, as $\lambda\to 0+$, and using (3.11) and (A.11),

$$\begin{aligned}\frac{1}{n^2c^2}\int_{3\lambda/2}^{\varepsilon}f(\mu)|J_\lambda(\mu)|^2\,d\mu&=O\left(\frac{f(\lambda)}{\lambda^{1-H}}\int_{3\lambda/2}^{\pi}\frac{\mu^{1-H}}{(\mu-\lambda)^2}\,d\mu\right)\\&=O\left(\frac{f(\lambda)}{\lambda}\right),\end{aligned} \tag{A.16}$$

$$\frac{1}{n^2c^2}\int_{\varepsilon}^{\pi}f(\mu)|J_\lambda(\mu)|^2\,d\mu=O\left(\int_{-\pi}^{\pi}f(\mu)\,d\mu/(\varepsilon-\lambda)^2\right)=O(\varepsilon^{-2}), \tag{A.17}$$

$$\frac{1}{n^2c^2}\int_{-\lambda/2}^{\lambda/2}f(\mu)|J_\lambda(\mu)|^2\,d\mu=O\left(\int_{-\lambda}^{\lambda}f(\mu)\,d\mu/\lambda^2\right)=O(f(\lambda)/\lambda), \tag{A.18}$$

using (3.11), (A.11) and Assumption 9, while $\int_{-\varepsilon}^{-\lambda/2}$ and $\int_{-\pi}^{-\varepsilon}$ are treated like (A.16) and (A.17), so that (A.14) $=O(f(\lambda)/n\lambda)$. Finally, Assumption 8 and (3.9) imply that

$$\sup_{\lambda/2\le\mu\le 3\lambda/2}\{|f(\lambda)-f(\mu)|/|\lambda-\mu|\}=O(f(\lambda)/\lambda)$$

as $\lambda\to 0+$, so that from (3.11) and (A.11), (A.13) is bounded by

$$\frac{Kf(\lambda)}{n\lambda}\int_{\lambda/2}^{3\lambda/2}|D_{n-2[nc]}(\lambda-\mu)|\,d\mu\le\frac{Kf(\lambda)}{n\lambda}\int_{0}^{\lambda/2}|D_{n-2[nc]}(\mu)|\,d\mu$$

and this is uniformly $O(g_j(1+\log j)/j)$ as $n\to\infty$ when evaluated at $\lambda=\lambda_j$ for $j=1,\dots,m$, from Lemma 5 of Robinson (1994b). It follows that uniformly in $j=1,\dots,m$,

$$I_{\theta\theta}(\lambda_j)=O_p\left(g_j\left\{\min\left(1,\frac{1}{c^2j^2}\right)+\frac{1+\log j}{j}\right\}\right)\quad\text{as } n\to\infty. \tag{A.19}$$

Using the same decomposition of $(-\pi,\pi)$, and by similar arguments to some of those just used, we deduce that

$$\begin{aligned}EI_{\chi\chi}(\lambda)&=\frac{1}{2\pi n}\int_{-\pi}^{\pi}f(\mu)|D_{[nc]}(\lambda-\mu)+D_n(\lambda-\mu)-D_{n-[nc]}(\lambda-\mu)|^2\,d\mu\\&\le\frac{2}{\pi n}\int_{-\pi}^{\pi}f(\mu)|D_{[nc]}(\lambda-\mu)|^2\,d\mu.\end{aligned}$$

Expanding the latter integral as in (A.13)–(A.18) and proceeding similarly we find that, uniformly,

$$I_{\chi\chi}(\lambda_j) = O_p\left(g_j\left\{\frac{1+\log j}{j} + c\right\}\right). \tag{A.20}$$

To deal with $I_{\zeta\zeta}$, we have [see (4.8)]

$$|w_\zeta(\lambda)| \le \frac{Kn^{-3/2}}{c}\sum_t i_t|r_t'|\left|\sum_s\left(\frac{t-s}{n}\right)\ell_{t-s}\right| + \frac{Kn^{-3/2}}{c}\sum_t\sum_s\left(\frac{t-s}{n}\right)^2|\ell_{t-s}|$$

by a two-term Taylor expansion. The last term is trivially $O(n^{1/2}c^2)$. From Assumption 10 the first term on the right is bounded by

$$\begin{aligned}&\frac{Kn^{-3/2}}{c}\sum_t\left|\sum_s\left(\frac{t-s}{n}\right)\ell_{t-s} - nc^2\int_{-1}^1 v\ell(v)\,dv\right|\\ &\quad\le \frac{Kc}{n^{1/2}}\sum_t\left|\sum_s\int_{(t-s)/nc}^{(t+1-s)/nc}\left\{\left(\frac{t-s}{nc}\right)\ell_{t-s} - v\ell(v)\right\}dv\right|\end{aligned}$$

for n large enough. By the mean value theorem this is bounded by

$$\frac{Kc}{n^{1/2}}\sum_t\left\{\frac{1}{(nc)^2}\sum_s|\ell_{t-s}| + \frac{1}{nc}\int_{-1}^1|v|\,dv\right\} = O(n^{-1/2}).$$

Thus

$$I_{\zeta\zeta}(\lambda_j) = O(nc^4 + n^{-1}) \tag{A.21}$$

uniformly. Finally,

$$I_{\xi\xi}(\lambda) \le \frac{K}{n}\left\{\sum_t i_t\left|\sum_s\int_{(t-s)/nc}^{(t+1-s)/nc}\{\ell_{t-s} - \ell(v)\}\,dv\right|\right\}^2 = O\left(\frac{1}{nc^2}\right). \tag{A.22}$$

From (A.19)–(A.22) it follows that for $j = 1,\dots,m$,

$$\frac{I_{vv}(\lambda_j)}{g_j} = O_p\left(\min\left(1, \frac{1}{c^2j^2}\right) + \frac{(1+\log j)}{j} + c + \frac{nc^4}{g_j} + \frac{1}{nc^2g_j}\right) \tag{A.23}$$

uniformly, as $n\to\infty$.

(iv) *Verification of* (A.1). By changing the order of summation, the left-hand side of (A.1) is bounded by

$$Km^{2(H-\Delta)-1}\sum_{j=1}^m j^{2(\Delta-H)}|d_j| \quad \text{for } \Delta < H, \tag{A.24}$$

and by

$$K\frac{\log m}{m}\sum_{j=1}^m|d_j| \quad \text{for } \Delta = H. \tag{A.25}$$

Applying (A.5), (A.24) is bounded by

$$Km^{2(H-\Delta)-1}\left[\left\{\sum_{j=1}^{m} j^{2(\Delta-H)}\frac{I_{uu}(\lambda_j)}{g_j}\sum_{j=1}^{m} j^{2(\Delta-H)}\frac{I_{vv}(\lambda_j)}{g_j}\right\}^{1/2} + \sum_{j=1}^{m} j^{2(\Delta-H)}\frac{I_{vv}(\lambda_j)}{g_j}\right]. \tag{A.26}$$

Because $2(\Delta - H) + 1 > 0$,

$$\sum_{j=1}^{m} j^{2(\Delta-H)}\frac{I_{uu}(\lambda_j)}{g_j} = O_p\left(m^{2(\Delta-H)+1}\right) \tag{A.27}$$

by (3.16) of Robinson (1995b). From (A.23), for $s \in [1, m-1]$,

$$\begin{aligned}\sum_{\substack{j=1\\(1)}}^{m} \frac{I_{vv}(\lambda_j)}{g_j} &= O_p\left(s + \frac{1}{c^2}\sum_{j=s+1}^{m} j^{-2} + \sum_{j=1}^{m}\frac{\log j}{j} + cm\right. \\ &\qquad \left. + \left(nc^4 + \frac{1}{nc^2}\right)n^{1-2H}\sum_{j=1}^{m} j^{2H-1}\right) \\ &= O_p\left(\frac{1}{c} + (\log m)^2 + cm + \left(nc^4 + \frac{1}{nc^2}\right)n^{1-2H}m^{2H}\right),\end{aligned} \tag{A.28}$$

on choosing $s \sim c^{-1}$. Thus (A.25) is $O_p(\log m(z_n + z_n^{1/2})) = o_p(1)$ by Assumption 11, where

$$z_n = \frac{1}{cm} + \frac{(\log m)^2}{m} + c + \left(nc^4 + \frac{1}{nc^2}\right)\left(\frac{n}{m}\right)^{1-2H}.$$

Proceeding similarly, for $\Delta \in (H - \frac{1}{2}, H)$

$$\begin{aligned}&\sum_{j=1}^{m} j^{2(\Delta-H)}\frac{I_{vv}(\lambda_j)}{g_j} \\ &\quad = O_p\left(c^{2(H-\Delta)-1} + 1 + cm^{2(\Delta-H)+1} + \left(nc^4 + \frac{1}{nc^2}\right)n^{1-2H}m^{2\Delta}\right)\end{aligned} \tag{A.29}$$

so that (A.24) is $O_p(z_n + z_n^{1/2}) = o_p(1)$.

(v) *Verification of* (A.2). The left-hand side of (A.2) is bounded by

$$\begin{aligned}&K(\log n)^2 m^{2\delta-1}\sum_{j=1}^{m} j^{-2\delta}|d_j| \\ &\leq K(\log n)^2 m^{2\delta-1} \\ &\quad \times\left[\left\{\sum_{j=1}^{m} j^{-4\delta}\frac{I_{uu}(\lambda_j)}{g_j}\sum_{j=1}^{m}\frac{I_{vv}(\lambda_j)}{g_j}\right\}^{1/2} + \sum_{j=1}^{m} j^{-2\delta}\frac{I_{vv}(\lambda_j)}{g_j}\right].\end{aligned} \tag{A.30}$$

Applying (A.27)–(A.29) with $\delta \in (0, \min(\frac{1}{4}, H))$, this is

$$O_p\left((\log n)^2\left(\frac{1}{mc} + \frac{(\log m)^2}{m} + c + \left(nc^4 + \frac{1}{nc^2}\right)\left(\frac{n}{m}\right)^{1-2H}\right)^{1/2}\right) = o_p(1)$$

by Assumption 11.

(vi) *Verification of* (A.3). In view of the bound for (A.25), the left-hand side of (A.3) is clearly $O_p((\log n)^2(z_n + z_n^{1/2})) \to_p 0$ under Assumption 11.

(vii) *Verification of* (A.4). From Robinson (1995b) we have $h \sim m/e$ as $n \to \infty$ and using $a_j = O(1)$ uniformly for $j > h$ and (vi) the left-hand side of (A.4) is, as $n \to \infty$, bounded by

$$\frac{1}{m}\sum_{j=1}^{m} a_j|d_j| + o_p(1) = O_p\left(\frac{1}{m^{2(\Delta-H)+1}}\sum_{j=1}^{m} j^{2(\Delta-H)}|d_j|\right) + o_p(1) = o_p(1)$$

by (i). □

PROOF OF THEOREM 4. It suffices to check (A.1)–(A.3) only, with d_j defined as before but with $I_{\hat{u}\hat{u}}$ defined in terms of (4.12). The first part of (A.5) still holds but now $v_t = r_t$, and so by Assumption 4, (3.11) and (A.11),

$$|w_v(\lambda)| \le \frac{K}{n^{1/2}}\sum_{t=1}^{n-1}|v_t - v_{t+1}||D_t(\lambda)| + \frac{K}{n^{1/2}}|v_n||D_n(\lambda)| \le \frac{K}{n^{1/2}\lambda},$$
$$0 < \lambda < \pi.$$

It follows that $I_{vv}(\lambda_j)/g_j = O(n^{2-2H}/j^{3-2H})$ uniformly. Next, (A.26) is

$$O_p\left(m^{2(H-\Delta)-1}\left[\left\{m^{2(\Delta-H)+1}n^{2-2H}\sum_{j=1}^{m} j^{2\Delta-3}\right\}^{1/2} + n^{2-2H}\sum_{j=1}^{m} j^{2\Delta-3}\right]\right)$$
$$= O_p(m^{(H-\Delta_1)-1/2}n^{1-H} + m^{2(H-\Delta_1)-1}n^{2-2H}) = o_p(1),$$

under Assumption 12, and because $\Delta = \Delta_1$. Likewise (A.25) is

$$O_p\left(\frac{\log m}{m}\left\{(mn^{2-2H})^{1/2} + n^{2-2H}\right\}\right) = O_p\left(\log m\left\{\frac{n^{1-H}}{m^{1/2}} + \frac{n^{2-2H}}{m}\right\}\right) = o_p(1).$$

Thus (A.1) is checked. To check (A.2), we see that (A.30) is, with $0 < \delta < \frac{1}{4}$,

$$O_p\left((\log n)^2 m^{2\delta-1}\left[\{m^{1-4\delta}n^{2-2H}\}^{1/2} + n^{2-2H}\right]\right)$$
$$= O_p\left((\log n)^2\left(\frac{n^{1-H}}{m^{1/2}} + \frac{n^{2-2H}}{m^{1-2\delta}}\right)\right) = o_p(1)$$

under Assumption 12. Finally, (A.3) is

$$O_p\left((\log n)^2\left(\frac{n^{1-H}}{m^{1/2}} + \frac{n^{2-2H}}{m}\right)\right) = o_p(1). \qquad \square$$

Acknowledgments. I am very grateful for two comprehensive referee reports which uncovered numerous defects and led to significant improvements in presentation, and for the comments of an Associate Editor. I thank Luis Gil-Alana for carrying out the computations described in Section 5.

REFERENCES

BERAN, J. (1992). A goodness-of-fit test for time series with long-range dependence. *J. Roy. Statist. Soc. Ser. B* **54** 749–760.

BERAN, J. and TERRIN, N. (1996). Testing for a change in the long-memory parameter. *Biometrika* **83** 627–638.

BHATTACHARYA, R. N., GUPTA, V. K. and WAYMIRE, E. (1983). The Hurst effect under trends. *J. Appl. Probab.* **20** 649–662.

CSÖRGŐ, S. and MIELNICZUK, J. (1995a). Close short-range dependent sums and regression estimation. *Acta Sci. Math. (Szeged)* **60** 177–196.

CSÖRGŐ, S. and MIELNICZUK, J. (1995b). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23** 1000–1014.

CSÖRGŐ, S. and MIELNICZUK, J. (1995c). Distant long-range dependent sums and regression estimation. *Stochastic Process. Appl.* **59** 143–155.

DEO, R. S. (1997). Nonparametric regression with long memory errors. *J. Time Ser. Anal.* To appear.

EICKER, F. (1967). Limit theorems for regressions with unequal and dependent errors. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 59–82. Univ. California Press, Berkeley.

FOX, R. and TAQQU, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GRAF, H.-P. (1983). Long-range correlations and estimation of the self-similarity parameter. Ph.D. dissertation, ETH Zurich.

GRAY, H. L., ZHANG, N.-F. and WOODWARD, W. A. (1989). On generalized fractional processes. *J. Time Ser. Anal.* **10** 233–257.

HALL, P. G. and HART, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351.

HANNAN, E. J. (1979). The central limit theorem for time series regression. *Stochastic Process. Appl.* **9** 281–289.

HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineeering* **16** 770–799.

IBRAGIMOV, I. A. and LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff, Gröningen.

KLEMES, V. (1974). The Hurst phenomenon: a puzzle? *Water Resources Research* **10** 678–688.

KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Proc. First World Congress of the Bernoulli Society* 67–74. VNU Science Press.

MANDELBROT, B. B. and WALLIS, J. R. (1968). Noah, Joseph and operational hydrology. *Water Resources Research* **4** 909–918.

MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628.

MÜLLER, H.-G. and STADMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.

MÜLLER, H.-G. and STADMÜLLER, U. (1988). Detecting dependencies in smooth regression models. *Biometrika* **75** 639–650.

RAY, B. and TSAY, R. (1966). Bandwidth selection for nonparametric regression with long range dependent errors. In *Athens Conference in Applied Probability and Time Series Analysis* (P. M. Robinson and M. Rosenblatt, eds.) **2** 339–351. Springer, New York.

RICE, J. A. (1984a). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13** 893–900.

RICE, J. A. (1984b). Bandwidth choice for nonparametric kernel regression. *Ann. Statist.* **12** 1215–1230.

ROBINSON, P. M. (1993). Highly insignificant *F*-ratios. *Econometrica* **61** 687–696.

ROBINSON, P. M. (1994a). Semiparametric analysis of long memory time series. *Ann. Statist.* **22** 515–539.

ROBINSON, P. M. (1994b). Rates of convergence and optimal bandwidth for long-range dependence. *Probab. Theory Related Fields* **99** 443–473.

ROBINSON, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072.

ROBINSON, P. M. (1995b). Gaussian semiparametric estimation of long-range dependence. *Ann. Statist.* **23** 1630–1661.

ROUSSAS, G. G., TRAN, L. T. and IOANNIDES, O. A. (1992). Fixed design regression for time series: asymptotic normality. *J. Multivariate Anal.* **40** 262–291.

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302.

TEVEROVSKY, V. and TAQQU, M. S. (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *J. Time Ser. Anal.* To appear.

TOUSSOUN, O. (1925). Mémoire sur l'histoire du Nil. *Mémoires de l'Institut d'Egypte* **9**. Cairo.

TRAN, L. T., ROUSSAS, G. G., YAKOWITZ, S. and TRUONG VAN, B. (1996). Fixed-design regression for linear time series. *Ann. Statist.* **24** 975–991.

YAJIMA, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791–807.

YONG, C. H. (1974). *Asymptotic Behaviour of Trigonometric Series*. Chinese Univ. Hong Kong.

ZYGMUND, A. (1977). *Trigonometric Series*. Cambridge Univ. Press.

LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
ENGLAND
E-MAIL: p.m.robinson@lse.ac.uk