The Annals of Statistics
1997, Vol. 25, No. 4, 1725–1739

NOTE ON CONVERGENCE RATES OF SEMIPARAMETRIC ESTIMATORS OF DEPENDENCE INDEX

BY PETER HALL, HIRA L. KOUL AND BERWIN A. TURLACH

Australian National University

Considerable recent attention has been devoted to semiparametric estimation of the dependence index, or the Hurst constant, using methods based on information in either frequency or time domains. Convergence rates of estimators in the frequency domain have been derived, and in the present paper we obtain them for estimators in the time domain. It is shown that the latter can have superior performance for moderate-range time series, but are inferior in the context of long-range dependence.

1. Introduction. A discrete-time stationary stochastic process $\{X_i\}$ is said to have dependence index $\theta > 0$ if its autocovariance at lag i, $\gamma(i) = \operatorname{cov}(X_j, X_{i+j})$, is regularly varying at infinity with exponent θ:

$$\gamma(i) = i^{-\theta}L(i) \quad \text{where } L \text{ is slowly varying at infinity.} \tag{1.1}$$

[When $\theta < 1$, this parameter is often called the self-similarity index. For the theory of regular and slow variation, see, e.g., Bingham, Goldie and Teugels (1987). It is not assumed that L is positive; the case where L is eventually negative is permitted by our methods and theory.] Alternative definitions of θ are available via Abelian–Tauberian theorems, which ask that the spectral density f be regularly varying at the origin with exponent $\theta - 1$: $f(t) = t^{\theta-1}M(t)$ where M is slowly varying at the origin. The case where $\theta \le 1$, commonly termed long-range dependence, is of particular interest, not least because there the rate of convergence of the sample mean depends critically on θ. A review article by Beran (1992) contains many examples of long-range dependent processes, indicating their practical importance.

In order to circumvent parametric assumptions about L, considerable recent interest has focussed on semiparametric estimation of θ. Major contributions in this direction have been made by Cheng and Robinson (1994), Delgado and Robinson (1994) and Robinson (1994a, b, c). Approaches based on the averaged periodogram [Robinson (1994a)] and the approximate distribution of the periodogram [Robinson (1995a, b)] have been explored relatively thoroughly, with consistency and upper bounds to convergence rates derived.

Received August 1995; revised October 1996.
AMS 1991 *subject classifications*. Primary 62M10; secondary 62G05.
Key words and phrases. Autocovariance, Hurst constant, long-memory time series, long-range dependence, semiparametric inference, stationary process, Gaussian process, short-range dependence.

On the other hand, techniques that operate in the time domain [Delgado and Robinson (1994), Robinson (1994c)] have received relatively little attention. In this note we derive convergence rates for estimators defined by the latter approach, showing that in the case of medium-range dependence they can be superior to frequency-domain methods, but are inferior under conditions of long-range dependence. These conclusions are supported by a simulation study.

Estimation of θ under parametric assumptions has been treated by Yajima (1985, 1988), Fox and Taqqu (1986) and Dahlhaus (1989). Whittle estimation for a non-Gaussian, long-memory process has been considered by Giraitis and Surgailis (1990), in the case where the process is linear with white noise innovations and square-summable coefficients. Related work on estimation under conditions of long-range dependence includes that of Geweke and Porter-Hudak (1983).

2. Main theoretical results.

2.1. *Definition of estimators*. Put $\overline{X} = n^{-1}\sum_{j \le n} X_j$, the usual sample mean. Define the sample autocovariances by

$$\hat{\gamma}(i) = (n-i)^{-1} \sum_{j=1}^{n-i} \left(X_j - \overline{X}\right)\left(X_{i+j} - \overline{X}\right)$$

and the design points in the regression problem by $x_i = \log i$. Consider regressing $\log|\hat{\gamma}(i)|$ on x_i for i in the range $m_1 + 1 \le i \le m_1 + m_2$, where m_1, m_2 increase with n and are both of smaller size than n. In the sense of first-order asymptotic theory, it is optimal to select both m_j's to be of the same size, and so we assume that there exists $m = m(n)$ with the properties

$$\begin{aligned} &0 < \liminf_{n \to \infty} (\log m / \log n) \le \limsup_{n \to \infty} (\log m / \log n) < 1, \\ &1 \le m_1 + 1 \le m_1 + m_2 \le n \quad \text{and} \quad m_j / m \to a_j \end{aligned} \tag{2.1}$$

where $0 < a_1, a_2 < \infty$.

Put

$$\bar{x}_n = m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} x_i \quad \text{and} \quad y_{ni} = x_i - \bar{x}_n.$$

In this notation, our estimator of θ is

$$\hat{\theta} = -\left\{ \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \log|\hat{\gamma}(i)| \right\} \Bigg/ \left(\sum_{i=m_1+1}^{m_1+m_2} y_{ni}^2 \right).$$

This is essentially the estimator proposed by Robinson (1994c), the main difference being that our theoretical analysis points to advantages of not taking $m_2 = n - m_1 - 1$ as suggested by Robinson.

An alternative proposal of Robinson (1994c) is based on minimizing the sum of squares

$$\sum_{i=m_1+1}^{m_1+m_2} \left\{ \gamma(i) - c i^{-\theta} \right\}^2$$

with respect to both c and θ. Eliminating c, one finds that the resulting estimator $\tilde{\theta}$ is a solution of the equation

$$\Big\{\sum \hat{\gamma}(i) i^{-\tilde{\theta}}\Big\}\Big(\sum i^{-2\tilde{\theta}} \log i\Big) - \Big\{\sum \hat{\gamma}(i) i^{-\tilde{\theta}} \log i\Big\}\Big(\sum i^{-2\tilde{\theta}}\Big) = 0,$$

where each sum is over $m_1 + 1 \le i \le m_1 + m_2$. The convergence rate of $\tilde{\theta}$ is similar to that of $\hat{\theta}$, but simpler to describe, and so we shall discuss it only briefly.

2.2. *Conditions on strength of dependence.* We assume that $\{X_i\}$ is a stationary Gaussian process, and ask that for all $\varepsilon > 0$ and some α, $\varepsilon_1 > 0$,

$$L(i)/L(m) = 1 + O(m^{\varepsilon - \alpha}) \quad \text{uniformly in } |i - m| \le \varepsilon_1 m. \tag{2.2}$$

Condition (2.2) is an analogue for the autocovariance of a part of Robinson's (1994a) assumption (A′), imposed on the spectral density. Condition (2.1) is the analogue of Robinson's (B′). Results for certain non-Gaussian processes are available in a longer version of this paper [Hall, Koul and Turlach (1995)].

In Theorem 2.2, where we investigate the conciseness of bounds to convergence rates, we strengthen (2.2) to: for some $\alpha > 0$ and for $c_1 \ne 0$ (either positive or negative) and $-\infty < c_2 < \infty$,

$$L(i) = c_1(1 + c_2 i^{-\alpha}) + o(i^{-\alpha}) \quad \text{as } i \to \infty. \tag{2.3}$$

2.3. *Convergence rates.* We devote most of our attention to the estimator $\hat{\theta}$, treating $\tilde{\theta}$ in Remark 2.4. Our first result, the main contribution of this paper, describes upper bounds to convergence rates under mild regularity conditions. Define $\beta = \beta(\theta) \equiv (2\theta)^{-1}$ if $\theta \ge 1/2$ and $\beta \equiv 1$ if $\theta < 1/2$.

THEOREM 2.1. *Assume* (1.1) *and* (2.1), *and that the process* $\{X_i\}$ *is stationary and Gaussian. Then, provided* $m \to \infty$ *so slowly that* $m = O(n^{\beta - \varepsilon})$ *for some* $\varepsilon > 0$, *the estimator* $\hat{\theta}$ *is weakly consistent for* θ. *If in addition* (2.2) *holds, then a convergence rate is provided by the following result: for each* $\varepsilon > 0$,

$$\hat{\theta} - \theta = O_p(n^{\varepsilon}) \begin{cases} (m^{-\alpha} + m^{\theta - 1/2} n^{-1/2} + m^{2\theta} n^{-1}), & \text{if } \theta > 1, \\ (m^{-\alpha} + m^{1/2} n^{-1/2} + m^{2\theta} n^{-1}), & \text{if } 1/2 < \theta \le 1, \\ (m^{-\alpha} + m^{\theta} n^{-\theta}), & \text{if } \theta \le 1/2. \end{cases} \tag{2.4}$$

Observe that the condition on the rate of increase of m becomes more stringent as θ increases.

To show that the convergence rate implied by (2.4) is close to best possible for $\hat{\theta}$, and not just a crude upper bound, we shall describe the sizes and origins of the terms in (2.4) with more precision, assuming more restrictive conditions. This work will demonstrate that in many circumstances Theorem 2.1 describes near-optimal bounds to performance. We need a little additional notation. Let a_1 and a_2 be as in (2.1) and c_1 and c_2 as in (2.3) and define

$$\mu \equiv a_2^{-1} \int_{a_1}^{a_1+a_2} \log y\, dy, \qquad c_3 \equiv -c_2 \int_{a_1}^{a_1+a_2} (\log x - \mu) x^{-\alpha}\, dx,$$

$$c_4 \equiv \int_{a_1}^{a_1+a_2} (\log x - \mu)^2\, dx, \qquad c_5 \equiv c_1^{-1} \int_{a_1}^{a_1+a_2} (\log x - \mu) x^{\theta}\, dx;$$

$$\xi_1 = \xi_1(m, n) \equiv \begin{cases} m^{2\theta-1}/n, & \text{if } \theta > 1, \\ m(\log m)/n, & \text{if } \theta = 1, \\ m/n, & \text{if } 1/2 < \theta < 1, \\ m\{\log(n/m)\}/n, & \text{if } \theta = 1/2, \\ (m/n)^{2\theta}, & \text{if } \theta < 1/2; \end{cases}$$

$$\xi_2 = \xi_2(m, n) \equiv \begin{cases} m^{2\theta}/n, & \text{if } \theta > 1/2, \\ m(\log n)/n, & \text{if } \theta = 1/2, \\ (m/n)^{2\theta}, & \text{if } \theta < 1/2; \end{cases}$$

$$\xi_3 = \xi_3(n) \equiv \begin{cases} n^{-1}, & \text{if } \theta > 1, \\ n^{-1} \log n, & \text{if } \theta = 1, \\ n^{-\theta}, & \text{if } \theta < 1. \end{cases}$$

Put $c_6 \equiv c_3/c_4$ and $c_7 \equiv c_5/c_4$ and let $c_8 > 0$ and c_9 denote constants depending on γ, a_1 and a_2. In Section 2.4 we shall note that in general, each of these constants is nonzero.

THEOREM 2.2. *Assume* (1.1), (2.1) *and* (2.3), *that* $\xi_2 n^{\varepsilon} \to 0$ *for some* $\varepsilon > 0$, *and that the process* $\{X_i\}$ *is stationary and Gaussian. If* $\theta = 1/2$ *or* 1, *strengthen* (2.1) *by asking as well that* $\log m/\log n$ *converges to a proper limit*, l *say. Then for each* $C > 0$ *there exist random variables* S_1, S_2 *and* S_3, *and nonrandom quantities* $t_1(n)$ *and* $t_2(n)$, *such that*

$$\hat{\theta} - \theta = t_1(n) + t_2(n)\Big\{(\bar{X} - EX)^2 + o_p(\xi_3)\Big\} + S_1 + S_2 + S_3 \tag{2.5}$$

with probability at least $1 - O(n^{-C})$, $E(S_1) = 0$, *and*

$$t_1(n) = c_6 m^{-\alpha} + o(m^{-\alpha}), \tag{2.6}$$

$$t_2(n) = c_7 m^{\theta} + o(m^{\theta}), \tag{2.7}$$

$$\operatorname{var} S_1 = c_8 \xi_1 + o(\xi_1), \tag{2.8}$$

$$E(|S_2|) = O(\xi_2) \quad \textit{and} \quad E(S_2) = c_9 \xi_2 + o(\xi_2), \tag{2.9}$$

$$E(|S_3|) = O\big(\xi_2^{3/2} \log n\big). \tag{2.10}$$

(*When* $\theta = 1/2$ *or* 1, *the constant* c_8 *depends on* l *as well as on* γ, a_1 *and* a_2.)

Substituting results (2.6)–(2.10) into (2.5), and noting that $(\bar{X} - EX)^2$ is of precise size ξ_3, we deduce an immediate corollary of Theorem 2.2:

$$
(2.11) \quad \hat{\theta} - \theta = O_p(1) \begin{cases} (m^{-\alpha} + m^{\theta - 1/2} n^{-1/2} + m^{2\theta} n^{-1}), & \text{if } \theta > 1, \\ \{m^{-\alpha} + m^{1/2} n^{-1/2} (\log n)^{1/2} \\ \qquad + m^2 n^{-1}\}, & \text{if } \theta = 1, \\ (m^{-\alpha} + m^{1/2} n^{-1/2} + m^{2\theta} n^{-1}), & \text{if } 1/2 < \theta < 1, \\ \{m^{-\alpha} + m^{1/2} n^{-1/2} (\log n)^{1/2}\}, & \text{if } \theta = 1/2, \\ (m^{-\alpha} + m^{\theta} n^{-\theta}), & \text{if } \theta < 1/2. \end{cases}
$$

This result is directly comparable with (2.4), the main difference being that the factor n^{ε} in the latter formula may be dropped in (2.11), because the slowly varying function L is now assumed to be asymptotic to a constant. The explicitly defined sources of the various contributions to the right-hand side of (2.11) indicate that, under the more restrictive conditions of Theorem 2.2, the right-hand side of (2.4) cannot be improved beyond replacing n^{ε} by 1.

2.4. *Discussion.*

REMARK 2.1 (Optimal rates and values of m). In view of the very close relationship between (2.4) and (2.11) it suffices to confine attention to the latter; optimal m's, and their associated convergence rates, differ only by a factor of order $n^{\pm \varepsilon}$, for arbitrarily small $\varepsilon > 0$, in the more general context of Theorem 2.1.

Of the five cases $\theta > 1$, $\theta = 1$, $\frac{1}{2} < \theta < 1$, $\theta = \frac{1}{2}$ and $\theta < \frac{1}{2}$, the second and fourth differ from their neighbors only in logarithmic factors. Therefore, we shall treat only the first, third and fifth of these cases. There the optimal rates, derived by minimizing over m, are respectively:

$$
\text{when } \theta > 1: \quad \begin{cases} n^{-\alpha/(\alpha + 2\theta)}, & \text{if } 0 < \alpha \le 1, \\ n^{-\alpha/(2\alpha + 2\theta - 1)}, & \text{if } \alpha > 1; \end{cases}
$$

$$
\text{when } \tfrac{1}{2} < \theta < 1: \quad \begin{cases} n^{-\alpha/(\alpha + 2\theta)}, & \text{if } \alpha \le 2\theta - 1, \\ n^{-\alpha/(2\alpha + 1)}, & \text{if } \alpha > 2\theta - 1, \end{cases}
$$

$$
\text{when } \theta < \tfrac{1}{2}: \quad n^{-\alpha\theta/(\alpha + \theta)}.
$$

They are achieved with values of m that are of the following respective sizes:

$$
\text{when } \theta > 1: \quad \begin{cases} n^{1/(\alpha + 2\theta)}, & \text{if } 0 < \alpha \le 1, \\ n^{1/(2\alpha + 2\theta - 1)}, & \text{if } \alpha > 1; \end{cases}
$$

$$
\text{when } \tfrac{1}{2} < \theta < 1: \quad \begin{cases} n^{1/(\alpha + 2\theta)}, & \text{if } \alpha \le 2\theta - 1, \\ n^{1/(2\alpha + 1)}, & \text{if } \alpha > 2\theta - 1, \end{cases}
$$

$$
\text{when } \theta < \tfrac{1}{2}: \quad n^{\theta/(\alpha + \theta)}.
$$

The relationship to the parametric case may be seen by allowing α to increase without bound. Then, the convergence rate tends to $n^{-1/2}$ when

$\theta > 1/2$, and to $n^{-\theta}$ for $\theta < 1/2$. The optimal value of m converges to 1 as α increases, and in fact if L is known up to a finite number of parameters and $\theta > 1/2$, then θ may be estimated root-n consistently from a finite number of $\hat{\gamma}(i)$'s, for *fixed* values of i.

REMARK 2.2 [Values of constants in (2.6)–(2.9)]. We discuss the constants only in general terms here, since concise identification of their values in each of the five cases addressed by (2.11) requires more space than can be justified. The constant c_9 appearing in (2.9) may be written as $c_9 = -c_{10}c_{11}$, where $c_{10} > 0$ and

$$c_{11} \equiv \int_{a_1}^{a_1+a_2} (\log x - \mu) x^{2\theta} \, dx. \tag{2.12}$$

The values of c_3, c_5 and c_{11} may sometimes be rendered equal to zero by judicious choice of $a \equiv (a_1, a_2)$, for a given value of (α, θ), although different a's are required for c_3, c_5 and c_{11}. In general, particularly since both α and θ are unknown, the constants c_3, c_5 and c_{11} would only vanish if α were estimated as a prelude to estimating θ and a preliminary estimator of θ were obtained, and if this information were employed to select a_1 and a_2.

Therefore, in general we may suppose c_3, c_5 and c_{11} are all nonzero, in which case c_6, c_7 and c_9 in (2.6), (2.7) and (2.9) are nonzero. The constant c_8 in (2.8) is strictly positive and finite. Therefore, each of the terms on the right-hand side of (2.5), except the remainder S_3, introduces a quantity whose size is accurately described by equations (2.6)–(2.9). These correspond to successive terms on the right-hand sides of (2.4) and (2.11). Therefore, the convergence rates given in Remark 2.1 provide a good description of the optimal performance of $\hat{\theta}$ in many settings.

REMARK 2.3 (Size of m_1 and m_2). In Section 2.1 we asserted that it is optimal to choose m_1 and m_2 to be of the same order. Asymptotic theory in the contrary case may be derived much as in Section 4. Without the assumption, contributions from bias and the $(\bar{X} - EX)^2$ and S_2 terms are of larger order than expressed by, for example, (2.11). These properties are apparent from careful analysis of the constants appearing in (2.5)–(2.10). Indeed, if we permit m_2/m to diverge to infinity, as could occur if we did not trim out very high lags, then the absolute values of the constants c_7 and c_9 diverge to $+\infty$. In particular this means that in (2.11), the term in $m^{\theta}n^{-\theta}$ (for $\theta < 1/2$) and the term in $m^{2\theta}n^{-1}$ (for $\theta > 1/2$) should be replaced by quantities of larger order. Furthermore, if we allow m_1/m to converge to zero then $|c_6| \to \infty$, so that the $m^{-\alpha}$ term in (2.11) should be replaced by one of larger order, for all θ. Since the bias term determines the optimal rate for each value of θ, m_1/m_2 should not converge to zero.

REMARK 2.4 (The estimator $\tilde{\theta}$). The rate of convergence of the estimator $\tilde{\theta}$ may be derived similarly, and shown to be similar to that of $\hat{\theta}$. Indeed, note that, defining $\Delta = \tilde{\theta} - \theta$, we have $i^{-k\tilde{\theta}} = i^{-k\theta}\ \{1 - k\Delta \log i + \frac{1}{2}(k\Delta \log i)^2 + \cdots\}$, whence

$$\begin{aligned}
&\left\{\sum \hat{\gamma}(i) i^{-\theta}\right\}\left(\sum i^{-2\theta} \log i\right) - \left\{\sum \hat{\gamma}(i) i^{-\theta} \log i\right\}\left(\sum i^{-2\theta}\right) \\
&\quad = \Delta\Big[2\left\{\sum \hat{\gamma}(i) i^{-\theta}\right\}\left\{\sum i^{-2\theta} (\log i)^2\right\} - \left\{\sum \hat{\gamma}(i) i^{-\theta} (\log i)^2\right\}\left(\sum i^{-2\theta}\right) \\
&\qquad\qquad - \left\{\sum \hat{\gamma}(i) i^{-\theta} \log i\right\}\left(\sum i^{-2\theta} \log i\right)\Big] + \cdots,
\end{aligned}$$

where each sum is over $m_1 + 1 \le i \le m_1 + m_2$, and "$\cdots$" denotes terms in Δ^r for $r \ge 2$. After dividing both sides by $m^{4\theta+2}$, we may show that the left-hand side may be decomposed into a sum of three terms, which to first order are constant multiples of $m^{-\alpha}$, $m^{\theta}(\bar{X} - EX)^2$ and S_1, plus a negligible remainder. [Terms involving $\log n$ cancel, up to and including the order of the aforementioned quantities, and there is no counterpart of S_2. The argument for expanding the left-hand side is similar to that which leads to (2.5).] Similarly, the right-hand side is asymptotic to a constant multiple of Δ, since terms in $\Delta \log n$ and $\Delta(\log n)^2$ cancel, up to and including the order of $|\Delta|$. Arguing thus we may prove that the convergence rate of $\tilde{\theta}$ to θ is that of $m^{-\alpha} + m^{\theta}(\bar{X} - EX)^2 + |S_1|$ to 0. Therefore, in those cases where S_2 is not a determining factor of the rate of convergence of $\hat{\theta}$ [e.g., when $\theta < \frac{1}{2}$, or when $\theta > \frac{1}{2}$ and $\alpha > \min(1, 2\theta - 1)$], the rates of convergence of $\hat{\theta}$ and $\tilde{\theta}$ to θ are identical. However, when S_2 determines the convergence rate of $\hat{\theta}$ (e.g., when $\theta > \frac{1}{2}$ and $\alpha < \min(1, 2\theta - 1)$), the rate of convergence of $\tilde{\theta}$ is superior.

REMARK 2.5 (Comparison with spectral estimators). Bounds on the rate of convergence of Robinson's (1994a) averaged periodogram estimator may be deduced from Robinson (1994b). Robinson's (1995b) periodogram-based estimator, which may be expected to be superior to that of Robinson (1994a), has a known convergence rate which, when θ is close to 0, is superior to that of both $\hat{\theta}$ and $\tilde{\theta}$. When $\theta > \frac{1}{2}$, the rates for $\hat{\theta}$ and $\tilde{\theta}$ are superior, although by only a logarithmic factor, and so the relative performance will in practice depend on matters such as the spread of the limiting distributions.

REMARK 2.6 (Asymptotic distributions). In the context of Theorem 2.2 the asymptotic distribution of $\hat{\theta}$ may be deduced from that for $\bar{X}$, S_1 or S_2, as follows. Except at the boundaries between the different regimes represented by (2.11), exactly one of $\bar{X}$, S_1 and S_2 will make a first-order contribution to the expansion of $\hat{\theta} - \theta$; which one may be determined from information given in the theorem about their respective sizes. For example, when $\theta < 1/2$ we may write $\hat{\theta} - \theta = t_1(n) + t_2(n)\ (\bar{X} - EX)^2 +$ negligible terms. Then, under our assumption that $\{X_i\}$ is Gaussian, the limiting distribution of $\hat{\theta}$ (suitably normalized) is chi-squared with one degree of freedom. Exact formulas for S_1

and S_2 are given in Section 4. They enable the limiting distribution to be determined when one or other of those variables dominates, using results of Taqqu (1975).

3. Numerical study. We are aware of at least five different approaches to estimating θ, using information in the time or frequency domains. And for each estimator type there are several variants, depending for example on how one estimates γ (in the time domain). The theoretical analysis reported in Section 2 indicates that time domain estimators such as $\hat{\theta}$ and $\tilde{\theta}$ perform similarly; and that they are inferior to frequency domain estimators in the case of very long-range dependence, but can be superior in conditions of medium-range dependence. In the present section we summarize the conclusions of a brief simulation study which verifies this conclusion in the case of the time domain estimator $\hat{\theta}$ and the frequency domain estimator suggested by Robinson (1994a). We shall explore three different variants of the former, and consider the case where $\{X_i\}$ is a stationary Gaussian processes with autocovariance

$$\gamma(i) \equiv (1 + i^2)^{-\theta/2}, \qquad -\infty < i < \infty. \tag{3.1}$$

This autocovariance was chosen in preference to the one considered by Robinson (1994a) because it is well defined for all $\theta > 0$. It satisfies (2.3) with $c_1 = 1$, $c_2 = -\frac{1}{2}\theta$ and $\alpha = 2$.

We simulated extensively for parameter settings in the ranges $\theta = 0.1(0.2)0.9$, 1.0, 1.5, 2.0 and $n = 64$, 128 and 256 and summarize the results in this section. Since $\alpha = 2$, the optimal convergence rate of $\hat{\theta}$, in the form stated in Section 2, is

$$\begin{array}{ll} n^{-2/(3+2\theta)}, & \text{if } \theta > 1, \\ (n^{-1}\log n)^{2/5}, & \text{if } \theta = 1/2 \text{ or } 1, \\ n^{-2/5}, & \text{if } 1/2 < \theta < 1, \\ n^{-2\theta/(2+\theta)}, & \text{if } 0 < \theta < 1/2, \end{array}$$

and is achieved with

$$m = \text{const.} \begin{cases} n^{1/(3+2\theta)}, & \text{if } \theta < 1, \\ (n/\log n)^{1/5}, & \text{if } \theta = 1/2 \text{ or } 1, \\ n^{1/5}, & \text{if } 1/2 < \theta < 1, \\ n^{\theta/(2+\theta)}, & \text{if } 0 < \theta < 1/2. \end{cases}$$

We constructed the estimator $\hat{\gamma}$ using three methods—the approach described in Section 2; a slightly modified form in which the divisor $n - i$ was replaced by n [thus, $\hat{\gamma}(i)$ is effectively replaced by $\tilde{\gamma}(i) \equiv (1 - in^{-1})\hat{\gamma}(i)$] and a jackknifed version in which $\hat{\gamma}$ was replaced by $2\tilde{\gamma} - \frac{1}{2}(\tilde{\gamma}_1 + \tilde{\gamma}_2)$, where $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are covariance estimates based on the first and second halves, respectively, of the observed time series.

TABLE 1
Values of mean squared error (approximated by numerical simulation over 5000 independent samples) for different values of the smoothing pair (m_1, m_2), in the case $(n, \theta) = (128, 1.0)$

	m_2				
m_1	**12**	**13**	**14**	**15**	**16**
0	0.04083	0.03870	0.03781	0.03813	0.03876
1	0.09896	0.09554	0.09433	0.09349	0.09382
2	0.21617	0.20674	0.19862	0.19318	0.18969

The estimator $\tilde{\gamma}$ is sometimes recommended because of relatively low variance, but this advantage was not clearly apparent in our analysis. In fact, $\hat{\gamma}$ and $\tilde{\gamma}$ had practically the same performance with $\tilde{\gamma}$ being slightly better.

Surprisingly, the expected advantage of the jackknifed estimator, lower bias at the expense of higher variance, was only noticeable for small values of θ. For larger θ's the estimators $\hat{\gamma}$ and $\tilde{\gamma}$ showed better performance. For example, the biases and mean squared errors (when tuning parameters were chosen so that the latter were at their minima) were reduced by 18.2% and 22.8%, respectively, for $(n, \theta) = (256, 0.1)$, by biasing $\hat{\theta}$ on the jackknifed form of $\tilde{\gamma}$ rather than $\hat{\gamma}$ or $\tilde{\gamma}$ itself. On the other side, for $(n, \theta) = (256, 0.9)$ we observe an increase of bias and mean square of the same magnitude.

For the sake of definiteness, the results reported below are based on $\tilde{\gamma}$. All reported mean squared errors for $\tilde{\gamma}$ are divided by four to make them directly comparable with the mean squared error achieved by Robinson's (1994a) estimator.

As a rule, Robinson's (1994a) estimator performed better than $\hat{\theta}$ for $\theta < 1/2$, whereas $\hat{\theta}$ was superior for $\theta > 1/2$. Robinson's estimator is appropriate only for $\theta < 1$, and its performance deteriorates as the boundary $\theta = 1$ is approached from below. For this reason we do not provide details of its performance for $\theta \geq 1$. Unexpectedly, there was a tendency for the performance of Robinson's estimator to decline with increasing sample size, for given settings of the smoothing parameters. We were unable to find the cause of this problem.

Table 1 presents mean squared errors as functions of the smoothing parameters m_1 and m_2 in the case $n = 128$ and $\theta = 1.0$. This gives an idea of the values at which optimality was achieved in general settings. Table 2 presents minimum mean squared errors of both $\hat{\theta}$ and the estimator of Robinson (1994a), as functions of sample size. These results indicate the trends noted above.

4. Outline proofs of Theorems 2.1 and 2.2. [In the case of Theorem 2.1 we derive only the convergence rate (2.4).] We may suppose without

TABLE 2

Values of minimum mean squared error multiplied by 100 (*approximated by numerical simulation over* 5000 *independent samples*) *as a function of sample size, for the estimator introduced by* Robinson (1994a) *and for* $\hat{\theta}$, *respectively**

	$\theta = 0.1$			$\theta = 0.3$		
n	64	128	256	64	128	256
Robinson's	0.065	0.0378	0.0274	0.86	0.4	0.279
$\tilde{\gamma}$	5.304	2.167	0.915	2.064	1.409	0.409
	$\theta = 0.5$			$\theta = 0.7$		
n	64	128	256	64	128	256
Robinson's	1.68	1.116	0.834	3.317	2.52	1.69
$\tilde{\gamma}$	1.306	1.392	0.518	2.892	2.043	1.075
	$\theta = 0.9$			$\theta = 1.0$		
n	64	128	256	64	128	256
Robinson's	5.85	4.6	3.172			
$\tilde{\gamma}$	4.178	3.209	1.995	5.110	3.781	2.582
	$\theta = 1.5$			$\theta = 2.0$		
n	64	128	256	64	128	256
$\tilde{\gamma}$	12.646	8.794	6.643	26.126	18.86	14.13

*In the case of Robinson's estimator, we took his tuning parameter q to equal $\frac{1}{2}$ in each case, and optimized over his smoothing parameter m.

loss of generality that $E(X) = 0$. Define

$$\xi_4 = \xi_4(n) \equiv \begin{cases} n^{-1}, & \text{if } \theta > 1/2, \\ n^{-1}\log n, & \text{if } \theta = 1/2, \\ n^{-2\theta}, & \text{if } \theta < 1/2. \end{cases}$$

In this notation, it may be shown by direct calculation of moments that for any integer $k \geq 1$,

$$E\{\hat{\gamma}(i) - \gamma(i)\}^{2k} = O\left(E\left[n^{-1}\sum_{j=1}^{n-i}\{X_j X_{i+j} - \gamma(i)\}\right]^{2k} + E(\bar{X}^{2k})\right) = O(\xi_4^k).$$

Hence, from Markov's inequality we have that for each $C > 0$ there exists $C_1 > 0$ such that

$$\begin{aligned} P\{n^{-C_1} < |\hat{\gamma}(i)| < n^{C_1} \quad \text{for all } i \in [m_1+1, m_1+m_2]\} \\ = 1 - O(n^{-C}). \end{aligned} \tag{4.1}$$

Write $\hat{\gamma}(i) = \gamma(i)\{1 + \Delta(i)\}$, where $\Delta(i) = \{\hat{\gamma}(i) - \gamma(i)\}/\gamma(i)$. Defining

$$s \equiv m_2^{-1}\sum_{i=m_1+1}^{m_1+m_2} y_{ni}^2 \to a_2^{-1}c_4,$$

$$t_1(n) \equiv -s^{-1} m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \log|L(i)/L(m)|,$$

$$V_1 \equiv -(sm_2)^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \log\{1 + \Delta(i)\} I\{|\Delta(i)| \le \tfrac{1}{2}\},$$

$$V_2 \equiv -(sm_2)^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \log\{|1 + \Delta(i)|\} I\{|\Delta(i)| > \tfrac{1}{2}\},$$

we have

$$\hat{\theta} - \theta = t_1(n) + V_1 + V_2. \tag{4.2}$$

Let $k = 2$ or 3, and note that

$$\begin{aligned} \Bigg| \sum_{i=m_1+1}^{m_1+m_2} & y_{ni} \log\{1 + \Delta(i)\} I\{|\Delta(i)| \le \tfrac{1}{2}\} \\ & - \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \sum_{j=1}^{k-1} j^{-1}(-1)^{j+1} \Delta(i)^j I\{|\Delta(i)| \le \tfrac{1}{2}\} \Bigg| \\ & \le 2k^{-1} \sum_{i=m_1+1}^{m_1+m_2} |y_{ni}|\, |\Delta(i)|^k, \end{aligned}$$

$$\begin{aligned} \Bigg| \sum_{i=m_1+1}^{m_1+m_2} & y_{ni} \sum_{j=1}^{k-1} j^{-1}(-1)^{j+1} \Delta(i)^j I\{|\Delta(i)| \le \tfrac{1}{2}\} \Bigg| \\ & \le (k-1) 2^{k-1} \sum_{i=m_1+1}^{m_1+m_2} |y_{ni}|\, |\Delta(i)|^k. \end{aligned}$$

Therefore,

$$\begin{aligned} \Bigg| \sum_{i=m_1+1}^{m_1+m_2} & y_{ni} \log\{1 + \Delta(i)\} I\{|\Delta(i)| \le \tfrac{1}{2}\} \\ & - \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \sum_{j=1}^{k-1} j^{-1}(-1)^{j+1} \Delta(i)^j \Bigg| \\ & \le k 2^{k-1} \sum_{i=m_1+1}^{m_1+m_2} |y_{ni}|\, |\Delta(i)|^k. \end{aligned}$$

In view of (4.1) there exists $C_2 > 0$, depending on C_1, such that with probability at least $1 - O(n^{-C})$,

$$|V_2| \le C_2 (\log n) m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} |y_{ni}| I\{|\Delta(i)| > \tfrac{1}{2}\}.$$

Noting that $\sup_n \sup_{m_1+1\le i\le m_1+m_2} |y_{ni}| < \infty$, and combining the results from (4.2) down, we deduce that

$$\hat{\theta} - \theta = t_1(n) - (sm_2)^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni} \sum_{j=1}^{k-1} j^{-1}(-1)^{j+1}\Delta(i)^j + R_{k-1},$$

where, defining

$$\delta_1 \equiv m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} \Big[(\log n) I\{|\Delta(i)| > \tfrac{1}{2}\} + \Delta(i)^2\Big]$$

and

$$\delta_2 \equiv m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} \Big[(\log n) I\{|\Delta(i)| > \tfrac{1}{2}\} + |\Delta(i)|^3\Big],$$

we have for a constant $C_3 > 0$ and with probability at least $1 - O(n^{-C})$,

$$|R_{k-1}| \le C_3 \delta_{k-1} \le C_3(2^k \log n + 1) m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} |\Delta(i)|^k. \tag{4.3}$$

Therefore, defining

$$S_1 \equiv -s^{-1} m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni}(n-i)^{-1} \sum_{j=1}^{n-i} \gamma(i)^{-1}\{X_j X_{i+j} - \gamma(i)\},$$

$$S_2 \equiv -\tfrac{1}{2} s^{-1} m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni}\Delta(i)^2,$$

$$S_4 \equiv \Big\{\bar{X}^2 + o_p(\xi_3)\Big\} t_2(n) \quad \text{and} \quad t_2(n) \equiv s^{-1} m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni}\gamma(i)^{-1},$$

we have

$$\hat{\theta} = \theta + t_1(n) + S_1 + S_4 + R_1 = \theta + t_1(n) + S_1 + S_2 + S_4 + R_2. \tag{4.4}$$

The first identity in (4.4) is used to derive (2.4) in Theorem 2.1, and the second to prove Theorem 2.2. Note that S_3 in (2.5) equals R_2.

Under (2.2),

$$|t_1(n)| = O(m^{\varepsilon-\alpha}), \tag{4.5}$$

for all $\varepsilon > 0$, while under (2.3), $t_1(n) = c_1 m^{-\alpha} + o(m^{-\alpha})$, which is (2.6). Next we determine the size of S_4, observing that under (2.2), $t_2(n) = O(m^{\theta+\varepsilon})$ and so

$$S_4 = O_p\big(m^{\theta+\varepsilon}\xi_3\big); \tag{4.6}$$

and under (2.3),

$$m^{-\theta} s^{-1} m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} y_{ni}\gamma(i) \to c_7,$$

from which follows (2.7).

Our next step is to derive the variance of S_1. We shall prove only (2.8), under the more stringent conditions of Theorem 2.2, it being a little simpler to show that under the assumptions of Theorem 2.1,

$$\operatorname{var} S_1 = O(\xi_1 n^{\varepsilon}) \tag{4.7}$$

for all $\varepsilon > 0$. Put

$$g(i_1, i_2) \equiv \sum_{j_1=1}^{n-i_1} \sum_{j_2=1}^{n-i_2} \{\gamma(j_1 - j_2 + i_1)\gamma(j_1 - j_2 - i_2) + \gamma(j_1 - j_2)\gamma(j_1 - j_2 + i_1 - i_2)\},$$

in which notation,

$$\operatorname{var} S_1 = s^{-2} m_2^{-2} \sum_{i_1=m_1+1}^{m_1+m_2} \sum_{i_2=m_1+1}^{m_1+m_2} y_{ni_1} y_{ni_2} \times \{(n-i_1)(n-i_2)\gamma(i_1)\gamma(i_2)\}^{-1} g(i_1, i_2).$$

Given $\varepsilon > 0$, let $\mathscr{M}$ denote the class of pairs (i_1, i_2) such that $m_1 + 1 \le i_j \le m_1 + m_2$ and $|i_1 - i_2| > \varepsilon m$. By approximating series by integrals, one may derive the following two results. (a) There exists a constant $C_4 = C_4(\gamma) > 0$ such that, uniformly in $(i_1, i_2) \in \mathscr{M}$, $g(i_1, i_2) \sim nC_4 h(i_1, i_2)$ where

$$h(i_1, i_2) \equiv \begin{cases} (|i_1 - i_2| + 1)^{-\theta} + (i_1 + i_2)^{-\theta}, & \text{if } \theta > 1, \\ (|i_1 - i_2| + 1)^{-1} \log(|i_1 - i_2| + 1) & \\ \quad + (i_1 + i_2)^{-1} \log(i_1 + i_2), & \text{if } \theta = 1, \\ (|i_1 - i_2| + 1)^{1-2\theta} + (i_1 + i_2)^{1-2\theta}, & \text{if } 1/2 < \theta < 1, \\ \log\{n/(|i_1 - i_2| + 1)\} + \log\{n/(i_1 + i_2)\}, & \text{if } \theta = 1/2, \\ n^{1-2\theta}, & \text{if } \theta < 1/2. \end{cases}$$

(Here and below the relation $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.) (b) There exist constants $0 < C_5 < C_6 < \infty$, not depending on n, such that $C_5 g(i_1, i_2) \le h(i_1, i_2) \le C_6 g(i_1, i_2)$ for all $m_1 + 1 \le i_1$, $i_2 \le m_1 + m_2$. Arguing thus,

$$\operatorname{var} S_1 \sim (s a_2 m n)^{-2} \sum_{i_1=m_1+1}^{m_1+m_2} \sum_{i_2=m_1+1}^{m_1+m_2} y_{ni_1} y_{ni_2} (i_1 i_2)^{\theta} g(i_1, i_2) \sim C_4 C_7 \xi_1,$$

where $C_7 > 0$ depends only on a_1, a_2 and θ (and, in the special cases $\theta = 1/2$ and 1, on l). This proves (2.8), with $c_8 = C_4 C_7$.

Next we show that (2.9) holds under the conditions of Theorem 2.2; it may similarly be proved that under the assumptions of Theorem 2.1,

$$E\left\{ \sum_{i=m_1+1}^{m_1+m_2} \Delta(i)^2 \right\} = O(\xi_2 n^{\varepsilon}). \tag{4.8}$$

Note that, uniformly in $m_1 + 1 \le i \le m_1 + m_2$,

$$E\{\Delta(i)^2\} = \{1 + o(1)\} c_1^{-2} i^{2\theta} n^{-2} g(i, i) + O\Big[i^{2\theta}\Big\{n^{-1} g(i,i)^{1/2} (E\bar{X}^4)^{1/2} + E(\bar{X}^4)\Big\}\Big],$$

and $g(i, i) \sim C_8 n^2 \xi_4$, where $C_8 > 0$ depends on γ. Furthermore, $E(\bar{X}^4)$ is of smaller order than this bound multiplied by n^{-2}. Result (2.9) follows on combining these results. The constant c_9 there is, in view of the definition of S_2, given by $c_9 = -\frac{1}{2} c_1^{-2} c_4^{-1} C_8 c_{11}$, where c_{11} is defined at (2.12).

Theorem 2.1 follows from (4.3), the first identity in (4.4), and (4.5)–(4.8). To complete the proof of Theorem 2.2 we must derive (2.10), for which it is sufficient to prove that

$$E(\delta_2) = O\big(m^{3\theta} \xi_4^{3/2} \log n\big). \tag{4.9}$$

By (4.3),

$$\begin{aligned} E(\delta_2) &\le (8 \log n + 1) m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} \big\{E|\Delta(i)|^4\big\}^{3/4}, \\ E|\Delta(i)|^4 &= O\big[i^{4\theta}\{n^{-4} a(i) + E(\bar{X}^8)\}\big], \end{aligned} \tag{4.10}$$

where

$$a(i) \equiv E\left[\sum_{j=1}^{n-i} \{X_j X_{i+j} - \gamma(i)\}\right]^4 = O\big(n^4 \xi_4^2\big),$$

uniformly in $m_1 + 1 \le i \le m_1 + m_2$ for all $\varepsilon > 0$. Similarly, $n^4 E(\bar{X}^8)$ is of smaller order than this. Therefore,

$$m_2^{-1} \sum_{i=m_1+1}^{m_1+m_2} \big\{E|\Delta(i)|^4\big\}^{3/4} = O\big(m^{3\theta} \xi_4^{3/2} \log n\big).$$

Results (4.9) follows from this result and (4.10).

Acknowledgments. The helpful comments of an Associate Editor and three reviewers led to significant improvements in the paper.

REFERENCES

BERAN, J. (1992). Statistical methods for data with long-range dependence (with discussion). *Statist. Sci.* **7** 404–427.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press.

CHENG, B. and ROBINSON, P. M. (1994). Semiparametric estimation from time series with long-range dependence. *J. Econometrics* **64** 335–353.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

DELGADO, M. A. and ROBINSON, P. M. (1994). New methods for the analysis of long-memory time series: application to Spanish inflation. *J. Forecasting* **13** 97–107.

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.

GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation of long-memory time series models. *J. Time Ser. Anal.* **4** 221–238.

GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104.

HALL, P., KOUL, H. L. and TURLACH, B. A. (1995). On estimating the self-similarity index for a long-range dependent process. Research Report No. SRR-32, Centre for Mathematics and Its Applications, Australian National Univ.

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 spectral bandwidth for long range dependence. *Probab. Theory Related Fields* **99** 443–473.

ROBINSON, P. M. (1994c). Time series with strong dependence. In *Advances in Econometrics, Sixth World Congress* (C. A. Sims, ed.) **1** 47–95. Cambridge Univ. Press.

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 time dependence. *Ann. Statist.* **23** 1630–1661.

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

YAJIMA, Y. (1985). On estimation of long-memory time series models. *Austral. J. Statist.* **27** 303–320.

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

P. HALL
B. A. TURLACH
CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA
E-MAIL: halpstat@durras.anu.edu.au

H. L. KOUL
DEPARTMENT OF STATISTICS
AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824