The Annals of Statistics
1997, Vol. 25, No. 1, 77–104

TIME SERIES REGRESSION WITH LONG-RANGE DEPENDENCE[1]

BY P. M. ROBINSON AND F. J. HIDALGO

London School of Economics

Dedicated to the memory of E. J. Hannan

A central limit theorem is established for time series regression estimates which include generalized least squares, in the presence of long-range dependence in both errors and stochastic regressors. The setting and results differ significantly from earlier work on regression with long-range-dependent errors. Spectral singularities are permitted at any frequency. When sufficiently strong spectral singularities in the error and a regressor coincide at the same frequency, least squares need no longer be $n^{1/2}$-consistent, where n is the sample size. However, we show that our class of estimates is $n^{1/2}$-consistent and asymptotically normal. In the generalized least squares case, we show that efficient estimation is still possible when the error autocorrelation is known only up to finitely many parameters. We include a Monte Carlo study of finite-sample performance and provide an extension to nonlinear least squares.

1. Introduction. This paper derives central limit theorems for estimates of the slope coefficient vector β in the multiple linear regression.

$$y_t = \alpha + \beta' x_t + u_t, \qquad t = 1, 2, \dots, \tag{1.1}$$

where both the K-dimensional column vector of regressors x_t and the unobservable scalar error u_t are permitted to exhibit long-range dependence, α is an unknown intercept and the prime denotes transposition. Two widely used methods of estimating β are ordinary least squares and generalized least squares. Both are known to be asymptotically normal under a wide variety of regularity conditions. However, it was pointed out by Robinson (1994a) that, when x_t and u_t collectively exhibit long-range dependence of a sufficiently high order, the least squares estimate is not asymptotically normal. We show that a class of weighted least squares estimates, which includes generalized least squares as a special case, is asymptotically normal under rather general forms of long-range dependence in both x_t and u_t.

Given observations y_t, x_t, $t = 1, \dots, n$, consider estimates of the following type [indexed by the function $\phi(\lambda)$, which is real-valued, even, integrable and

Received February 1995; revised March 1996.

[1] Research supported by ESRC Grant R000235892.

AMS 1991 *subject classifications*. Primary 62M10, 60G18; secondary 62F12, 62J05, 62J02.

Key words and phrases. Long-range dependence, linear regression, generalized least squares, nonlinear regression.

periodic of period 2π]:

$$\hat{\beta}_\phi = \hat{A}_\phi^{-1}\hat{b}_\phi, \tag{1.2}$$

where, with Σ_t denoting $\Sigma_{t=1}^n$,

$$\begin{aligned}
\hat{A}_\varphi &= \frac{1}{2\pi}\int_{-\pi}^{\pi} I_x(\lambda)\varphi(\lambda)\,d\lambda,\\
\hat{b}_\varphi &= \frac{1}{2\pi}\int_{-\pi}^{\pi} I_{xy}(\lambda)\varphi(\lambda)\,d\lambda,\\
I_x(\lambda) &= w_x(\lambda)w_x(-\lambda)',\\
I_{xy}(\lambda) &= w_x(\lambda)w_y(-\lambda),\\
w_x(\lambda) &= \frac{1}{(2\pi n)^{1/2}}\sum_t (x_t-\bar{x})e^{it\lambda},\\
w_y(\lambda) &= \frac{1}{(2\pi n)^{1/2}}\sum_t (y_t-\bar{y})e^{it\lambda},\\
\bar{x} &= \frac{1}{n}\sum_t x_t,\qquad \bar{y} = \frac{1}{n}\sum_t y_t,
\end{aligned}\tag{1.3}$$

and $\hat{A}_\phi$ is nonsingular. Here $\hat{A}_\phi$ and $\hat{b}_\phi$ can equivalently be written as

$$\begin{aligned}
\hat{A}_\phi &= \frac{1}{n}\sum_t\sum_s (x_t-\bar{x})(x_s-\bar{x})'\phi_{t-s},\\
\hat{b}_\phi &= \frac{1}{n}\sum_t\sum_s (x_t-\bar{x})(y_s-\bar{y})\phi_{t-s},
\end{aligned}\tag{1.4}$$

where

$$\phi_j = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\phi(\lambda)\cos j\lambda\,d\lambda. \tag{1.5}$$

In case $\phi(\lambda) \equiv 1$, so that $\phi_0 = 1/2\pi$ and $\phi_j = 0$ for $j \neq 0$, $\hat{\beta}_\phi = \hat{\beta}_1$, the least squares estimate,

$$\hat{\beta}_1 = \left\{\sum_t (x_t-\bar{x})x_t'\right\}^{-1}\sum_t (x_t-\bar{x})y_t. \tag{1.6}$$

We assume throughout that u_t is covariance stationary, having mean that is (without loss of generality) 0, and absolutely continuous spectral distribution function, so that it has spectral density, denoted $f(\lambda)$, satisfying

$$\gamma_j =_{\mathrm{def}} E(u_1u_{1+j}) = \int_{-\pi}^{\pi} f(\lambda)\cos j\lambda\,d\lambda,\qquad j = 0,1,\dots. \tag{1.7}$$

Suppose $f(\lambda) > 0$, $-\pi < \lambda \le \pi$. Then taking $\phi(\lambda) = f(\lambda)^{-1}$ gives a generalized least squares estimate $\hat{\beta}_{f^{-1}}$. In case f is not known up to scale, but only up to a finite-dimensional vector of parameters θ, so that $f(\lambda) = f(\lambda;\theta)$ is a given function of λ and θ, a feasible generalized least squares estimate is $\hat{\beta}_{\hat{f}^{-1}}$, obtained by taking $\phi(\lambda) = f^{-1}(\lambda;\hat{\theta})$, where $\hat{\theta}$ is an estimate of θ.

Now suppose that x_t is a stochastic sequence, independent of the u_t sequence, and covariance stationary with autocovariance matrix $\Gamma_j = E(x_1 - Ex_1)(x_{1+j} - Ex_1)'$. Then

$$\Gamma_j = \int_{-\pi}^{\pi} e^{ij\lambda}\, dF(\lambda), \tag{1.8}$$

where the matrix F has Hermitian nonnegative definite increments and is uniquely defined by the requirement that it is continuous from the right. Under suitable conditions we then have the central limit theorem (clt)

$$n^{1/2}\left(\hat{\beta}_\phi - \beta\right) \to_d N\left(0, \Sigma_\phi^{-1}\Sigma_\psi\Sigma_\phi^{-1}\right), \tag{1.9}$$

where $\psi(\lambda) = \phi^2(\lambda)f(\lambda)$ and

$$\Sigma_\chi = \frac{1}{2\pi}\int_{-\pi}^{\pi} \chi(\lambda)\, dF(\lambda) \tag{1.10}$$

for $\chi(\lambda)$ such that Σ_χ is finite and nonsingular. In particular,

$$n^{1/2}\left(\hat{\beta}_1 - \beta\right) \to_d N\left(0, (2\pi)^2\Gamma_0^{-1}\Sigma_f\Gamma_0^{-1}\right), \tag{1.11}$$

$$n^{1/2}\left(\hat{\beta}_{f^{-1}} - \beta\right) \to_d N\left(0, \Sigma_{f^{-1}}^{-1}\right), \tag{1.12}$$

$$n^{1/2}\left(\hat{\beta}_{\hat{f}^{-1}} - \beta\right) \to_d N\left(0, \Sigma_{f^{-1}}^{-1}\right). \tag{1.13}$$

Some conditions for (1.9) and (1.11)–(1.13) have already been laid down in the literature. In particular, the case when f is at least bounded,

$$\sup_\lambda f(\lambda) < \infty, \tag{1.14}$$

has effectively been covered in a large literature, some relatively complete results appearing in Hannan (1979). There is also interest in cases where (1.14) does not hold, when we can say that u_t has long-range dependence: $f(\lambda)$ has a singularity at one or more λ, such as $\lambda = 0$. Here the relevant literature is much less extensive, and it has stressed the least squares estimate in case of nonstochastic x_t satisfying what have come to be called "Grenander's conditions." Eicker (1967) gave a clt under the assumption that

$$u_t = \sum \tau_j \varepsilon_{t-j}, \qquad \sum \tau_j^2 < \infty, \tag{1.15}$$

where $\sum$ will always denote a sum over $0, \pm 1, \dots$ of an obvious index, and the ε_t are independent with finite variance, and Hannan (1979) relaxed the latter assumption to square-integrable martingale differences. The square-summability condition in (1.15) is equivalent only to covariance stationarity of u_t given the other conditions (i.e., to integrability of f). Such stronger conditions on the τ_j as absolute summability would rule out long-range dependence, and imply (1.14). Mention must also be made of important early, contributions in the simple location model special case. Ibragimov and Linnik (1971), Theorem 18.6.5, considered this under essentially the same conditions on u_t as Eicker (1967), while Taqqu (1975) initiated a new avenue of research by assuming u_t is a nonlinear function of a Gaussian process.

The work of Yajima (1988, 1991) contained some central limit theory (under conditions on cumulants of all orders) but stressed other aspects of regression with nonstochastic regressors and long-range-dependent errors. He assumed that

$$f(\lambda) = f^*(\lambda)|\lambda|^{-2d}, \qquad -\pi < \lambda \leq \pi, 0 < d < \tfrac{1}{2}, \tag{1.16}$$

where f^* is a continuous function, and showed the importance to the properties of least squares of the behavior near $\lambda = 0$ of the spectral distribution function, denoted $M(\lambda)$, of the generalized harmonic analysis of his nonstochastic x_t: the estimate of the ith element of β has rate of convergence depending on d if the ith diagonal element of $M(\lambda)$ has positive increment at $\lambda = 0$, and the same rate as under (1.14) otherwise. Yajima (1988, 1991) also obtained formulas for the asymptotic covariance matrix of least squares and generalized least squares, and conditions under which they have equal asymptotic efficiency. Dahlhaus (1995) proved asymptotic normality of generalized least squares estimates, and approximations thereto, for certain forms of nonstochastic regressor such that M has mass at zero frequency and under a condition similar to (1.16). Künsch, Beran and Hampel (1993) discussed the effect of long-range-dependent errors on standard independence-based inference rules in the context of certain experimental designs. Koul (1992), Koul and Mukherjee (1993) and Giraitis, Koul and Surgailis (1994) considered the asymptotic properties of various robust estimates of β.

Our setup can be compared to those of Yajima (1991) under his condition (1.16) in that, due to our mean correction, F effectively corresponds to his M in case an element of M has a jump at frequency 0. Here Yajima (1991) imposed conditions that are not innocuous. In particular, he required [case (i) of his Theorem 2.1] that, in our notation, Σ_f be finite. Yajima (1991) included also situations in which Σ_f diverges due to a jump in M at frequency 0. Robinson (1994a) indicated how Σ_f can diverge in our present context. Suppose that a certain diagonal element of $F(\lambda)$ is absolutely continuous with derivative that behaves like $C|\lambda|^{-2c}$ in a neighborhood of $\lambda = 0$, for $c \in (0, \frac{1}{2})$ and $C > 0$. Then $\Sigma_f < \infty$ if and only if $c + d < \frac{1}{2}$, which is not true if there is collectively sufficient long-range dependence in u_t and an x_t element. This does not seem implausible in a situation in which long-range dependence is suspected in u_t, because of empirical evidence of the peakedness, near zero frequency, of spectra of observable series. In this circumstance it appears that (1.11) does not hold; $\hat{\beta}_1$ has a rate of convergence slower than $n^{1/2}$, and need not be asymptotically normal. Note that $\Sigma_f = \infty$ also if sufficiently strong singularities in the spectral densities of x_t and u_t coincide at any nonzero frequency. Note also that the obvious stochastic analog of condition (23) of Yajima (1991) (employed in his clt for least squares) does not hold even in case of white noise x_t. Our focus on stochastic x_t reflects practical situations frequently encountered (e.g., in oceanography and econometrics) where the response of one series to others is of interest, yet it is natural to regard all as stochastically generated [see, e.g., Hamon and Hannan (1963)].

The present paper identifies functions $\phi(\lambda)$ which can guarantee the clt (1.9), with its $n^{1/2}$ rate of convergence, under effectively no restrictions on the degree of stationary long-range dependence in x_t and u_t. Such ϕ include f^{-1}, so that (1.12) is included. The efficiency advantages of $\hat{\beta}_{f^{-1}}$ are well known in other circumstances, but the preceding discussion of the pitfalls of $\hat{\beta}_1$ makes $\hat{\beta}_{f^{-1}}$ seem even more attractive than usual. The conditions for the clt's (1.9) and (1.12) are introduced and described in the following section, which, under the same conditions, asserts corresponding results in case $\hat{\beta}_\phi$ is replaced by

$$\tilde{\beta}_\phi = \tilde{A}_\phi^{-1}\tilde{a}_\phi, \tag{1.17}$$

where

$$\tilde{A}_\phi = \frac{1}{n}\sum_{j=1}^{n-1} I_x(\lambda_j)\phi(\lambda_j), \qquad \tilde{a}_\phi = \frac{1}{n}\sum_{j=1}^{n-1} I_{xy}(\lambda_j)\phi(\lambda_j), \tag{1.18}$$

where $\lambda_j = 2\pi j/n$. Note that $\tilde{\beta}_\phi$ will sometimes be preferred on computational grounds, especially as formulas for $\phi(\lambda) = f^{-1}(\lambda)$ [and indeed for many other choices of $\phi(\lambda)$] are often simpler than the corresponding ones for ϕ_j. For $1 \le j \le n-1$, $w_x(\lambda_j)$ and $w_y(\lambda_j)$ are invariant to location shift and so the mean correction in the formulas in (1.3) is vacuous; it is the omission of $j = 0$ (and n) from the sums in (1.18) which permits unknown α and $E(x_t)$. (Note that $\tilde{\beta}_1 = \hat{\beta}_1$.) The clt proofs for both $\hat{\beta}_\phi$ and $\tilde{\beta}_\phi$ are given in Section 3. Section 4 discusses the clt for the case (1.13) for a parametric f. Section 5 includes an extension to nonlinear regression. We report in Section 6 the results of a small Monte Carlo study of finite-sample behavior.

2. Central limit theorems.

We introduce the following conditions.

CONDITION 1.

$$u_t = \sum_0^\infty \tau_j \varepsilon_{t-j}, \qquad \sum_0^\infty \tau_j^2 < \infty,$$

where

$$E(\varepsilon_t | F_{t-1}) = 0, \qquad E(\varepsilon_t^2 | F_{t-1}) = E(\varepsilon_1^2) = \sigma^2 \quad \text{a.s.},$$

F_{t-1} being the σ-field of events generated by $\{\varepsilon_s,\ s \le t-1\}$, and the ε_t^2 being uniformly integrable.

CONDITION 2.

$$\sum_0^\infty \tilde{\phi}_j < \infty \quad \text{where } \tilde{\phi}_a = \max_{j \geq a} |\phi_j|.$$

CONDITION 3. The spectral density $f(\lambda)$ exists for all λ, and

$$\left(\sum_0^n |\gamma_j| + n\tilde{\gamma}_n \right) \left\{ \left(\sum_0^n \tilde{\phi}_j^{1/2} \right)^2 + n\Phi_n \right\} = O(n) \quad \text{as } n \to \infty, \tag{2.1}$$

where $\tilde{\gamma}_a = \max_{j \geq a} |\gamma_j|$, $\Phi_a = \sum_{|j| > a} |\phi_j|$.

CONDITION 4. $\psi(\lambda)$ is a continuous function.

CONDITION 5. Σ_ϕ and Σ_ψ are nonsingular; Σ_ψ is finite.

CONDITION 6. $\{x_t\}, \{\varepsilon_t\}$ are independent sequences.

CONDITION 7. $\{x_t\}$ is fourth-order stationary, and

$$\Gamma_j \to 0 \quad \text{as } j \to \infty, \tag{2.2}$$

$$\lim_{|u| \to \infty} \max_{|v|, |w| < \infty} |\kappa_{abcd}(0, u, v, w)| = 0, \qquad 1 \leq a, b, c, d \leq K, \tag{2.3}$$

where $\kappa_{abcd}(0, u, v, w)$ is the fourth cumulant of x_{a0}, x_{bu}, x_{cv}, x_{dw}, and x_{it} is x_t's ith element.

THEOREM 1. *Under* (1.1) *and Conditions* 1–7, *it follows that*

$$n^{1/2}\left(\hat{\beta}_\phi - \beta \right), n^{1/2}\left(\tilde{\beta}_\phi - \beta \right) \to_d N\left(0, \Sigma_\phi^{-1} \Sigma_\psi \Sigma_\phi^{-1}\right), \tag{2.4}$$

and thus, when $f(\lambda) > 0$ *and* $\phi(\lambda) = f(\lambda)^{-1}$ *for all* λ,

$$n^{1/2}\left(\hat{\beta}_{f^{-1}} - \beta \right), n^{1/2}\left(\tilde{\beta}_{f^{-1}} - \beta \right) \to_d N\left(0, \Sigma_{f^{-1}}^{-1}\right). \tag{2.5}$$

The proof is reserved for the following section, and we first discuss the conditions.

Condition 1 is a relaxation of Eicker's (1967) condition on his ε_t in (1.15); see also Hannan (1979). It is restrictive in the linearity it imposes, but not otherwise.

Condition 2 implies that $\phi_j = O(\tilde{\phi}_j) = o(j^{-1})$ as $j \to \infty$. Then Condition 3 is automatically satisfied if also $\gamma_j = o(j^{-1})$ as $j \to \infty$. On weakening the latter requirement to $\gamma_j = O(\log^p(2 + |j|)/(1 + |j|))$, for any $p > 0$, we can satisfy Condition 3 by $\phi_j = O(1/(1 + |j|)\log^{p+2}(2 + |j|))$. On the other hand, if only $\gamma_j = O(1/\log^p(2 + |j|))$, for any $p > 0$, then $\phi_j = O(\log^{p-2}(2 + |j|)/(1 + |j|))$ suffices. The first statement of Condition 3 implies, by the Riemann–Lebesgue lemma [Zygmund (1977), page 45], the mild type of ergodicity condition $\lim_{j \to \infty} \gamma_j = 0$, and (2.1) entails no extra condition on γ_j when $\sum_0^\infty \tilde{\phi}_j^{1/2} < \infty$ [which implies $\phi_j = o(j^{-2})$].

Perhaps the conditions of most interest under which Condition 3 holds lie between these extremes: when, for some $d \in (0, \frac{1}{2})$,

$$\gamma_j = O\big(L(|j|)^{-1}(1+|j|)^{2d-1}\big), \tag{2.6}$$

$$\phi_j = O\big(L(|j|)(1+|j|)^{-2d-1}\big), \tag{2.7}$$

where L is slowly varying at ∞ [see, e.g., Bingham, Goldie and Teugels (1987)]. The special case of (2.6)

$$\gamma_j \sim Dj^{2d-1}, \qquad D > 0 \quad \text{as } j \to \infty \tag{2.8}$$

(where "$\sim$" means that the ratio of left- and right-hand sides tends to 1) holds in case of fractional autoregressive integrated moving average (FARIMA) models with differencing parameter d, where f in (1.7) is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 - e^{i\lambda}|^{-2d}\left|\frac{a(e^{i\lambda})}{b(e^{i\lambda})}\right|^2, \qquad -\pi < \lambda \le \pi, \tag{2.9}$$

where a and b are polynomials of finite degree, all of whose zeros are outside the unit circle, and $\sigma^2 > 0$. The model (2.9) may have originated in Adenstedt (1974), and it also satisfies (1.16). However, we stress that "O" in (2.6) has the usual "upper bound" meaning, not "exact rate," so even when $L(j) \equiv 1$ it can hold also when the γ_j not only decay slowly, but indefinitely oscillate. For example, consider the spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi}\prod_{j=1}^{h}|1 - 2\cos\omega_j e^{i\lambda} + e^{2i\lambda}|^{-2d_j}\left|\frac{a(e^{i\lambda})}{b(e^{i\lambda})}\right|^2, \qquad -\pi < \lambda \le \pi, \tag{2.10}$$

where a, b and σ^2 are as before, and the ω_j are distinct numbers in $[0, \pi]$. In case $h = 1$ and $\omega_1 = 0$, (2.10) reduces to (2.9) with $d_1 = d/2$, so $0 < d_1 < \frac{1}{4}$ is required for (2.6) to hold. When $h = 1$ but $0 < \omega_1 \le \pi$, we have a "cyclic" FARIMA, f having singularity at the nonzero frequency ω_1. In case $h = 1$ and $\omega_1 = \pi$, Gray, Zhang and Woodward (1989) showed that $\gamma_j \sim D(-1)^j j^{4d_1 - 1}$, so again we need $0 < d_1 < \frac{1}{4}$ for (2.6) to hold, but now (2.8) does not hold. In case $h = 1$ and $0 < \omega_1 < \pi$, Gray, Zhang and Woodward (1989, 1994) showed that $\gamma_j \sim Dj^{2d_1 - 1}\cos(j\omega_1)$, so (2.6) holds when $0 < d_1 < \frac{1}{2}$, but not (2.8). For $h > 1$ there is more than one spectral singularity in $[0, \pi]$, as may be a reasonable model for seasonal processes [e.g., for monthly data take $h = 7$ and $\omega_j = (j-1)\pi/6$, $j = 1, \dots, 7$]. We conjecture that (2.6) will hold, with $d = \max(a_1, \dots, a_h)$ and $0 < a_j/2 = d_j < \frac{1}{4}$ if $\omega_j = 0$ or π, $0 < a_j = d_j < \frac{1}{2}$ otherwise. The problem of testing hypotheses on the d_j in (2.9) and (2.10) has been studied by Robinson (1994b). Conditions that link (1.16) and (2.8) outside the classes (2.9) and (2.10) can be inferred from Yong (1974), and sufficient conditions for (2.6) with $L(j) \equiv 1$ in terms of upper bounds on f and its derivative are in Lemma 4 of Fox and Taqqu (1986), which can also be generalized to the case of singularities at nonzero frequencies.

As d decreases from $\frac{1}{2}$ to 0, (2.6) becomes stronger but (2.7) becomes weaker, and while "O" in the latter again means "upper bound," we have

$$\phi_j \sim D' j^{-2d-1}, \qquad D' > 0 \quad \text{as } j \to \infty, \tag{2.11}$$

along with (2.8) in case $\phi = f^{-1}$ and u_t is a noncyclic FARIMA as in (2.9). For this ϕ, the ϕ_j are proportional to the inverse autocorrelations of u_t. Condition 4 is implied by Condition 2, if $\phi = f^{-1}$, and indicates that in order to choose ϕ merely to guarantee an $n^{1/2}$-consistent $\hat{\beta}_\phi$ we have to know the location of the singularity or singularities of f, and design ϕ to have a zero or zeros of sufficient order to cancel them. Suppose that, for some $\omega \in [0, \pi)$, $f(\lambda) \sim C|\lambda - \omega|^{-2d}/L(1/|\lambda - \omega|)$ as $\lambda \to \omega$ for $d \in (0, \frac{1}{2})$, f is continuous for $\lambda \neq \pm\omega$ and (2.6) holds [as is true in case of (2.9)]. Then, for any such d,

$$\phi(\lambda) = |\lambda - \omega|, \qquad 0 \le \lambda < \pi, \tag{2.12}$$

satisfies Conditions 2, 3 [because $\psi_j = O(j^{-2})$] and 4. When f has several singularities we can extend (2.12) to have zeros of degree 1 at each.

Nonsingularity in Condition 5 is likely to hold in case of no multicollinearity in x_t. Finiteness of Σ_ϕ is a consequence of Parseval's equality and Condition 2.

Condition 6 is very restrictive. We believe it could be relaxed to independence up to some moment order, but at the cost of greater structure on x_t than required in Condition 7. Such structure (e.g., a linear process similar to that assumed for u_t) could also lead to a reduction in the fourth moment condition on x_t, but we find the extremely mild condition (2.2) aesthetically so appealing as to offset both considerations. The fourth cumulant condition (2.3) is vacuous in case of Gaussianity, and milder than the summability conditions on cumulants frequently employed. Finally, the covariance stationarity condition on x_t is very strong. There may be scope for replacing it by a stochastic version of "Grenander's conditions." Neither Condition 7 nor "Grenander's conditions" are satisfied by unit root behavior, a popular recent assumption. While there are certainly many series lending empirical support for this sort of assumption, not only is it strictly not "weaker" than Condition 7, but it actually describes a very specialized form of nonstationarity, and covers a far narrower spectrum of the "$I(d)$" processes than does Condition 7. Note that the model (1.1) permits the vector x_t to consist of lags or leads of a basic time series [see, e.g., Hannan (1967)]; parsimonious parameterizations have been proposed in this case, entailing linear restrictions on β, and it is easy to extend our results to cover these.

It is our insistence on allowing for any degree of fractional differencing, at any frequencies, in x_t which prevents the least squares choice of $\phi \equiv 1$ [see (1.6)] from being covered by the clt, and we have preferred not to investigate in detail the trade-offs between F and ϕ which the discussion of Section 1 indicates is possible. However, the following result will be of some use in Section 4. Its proof is a simple consequence of some of the properties derived in the following section, and the Toeplitz lemma [see, e.g., Stout (1974), pages 120 and 121], and is omitted.

THEOREM 2. *Under* (1.1) *and Conditions* 6 *and* 7 *and the nonsingularity of* Γ_0,

$$\hat{\beta}_1 = \beta + o_p\left[n^{-1}\sum_0^n|\gamma_j|\right]^{1/2} \quad as\ n \to \infty. \tag{2.13}$$

As an application of Theorem 2, consider a problem of relatively minor interest, estimating α in (1.1). An obvious estimate is

$$\hat{\alpha}_\phi = \bar{y} - \hat{\beta}'_\phi \bar{x} = \alpha + \bar{u} - \left(\hat{\beta}_\phi - \beta\right)' \bar{x}, \tag{2.14}$$

where $\bar{u} = n^{-1}\Sigma_t u_t$. If $\hat{\beta}_\phi = \beta + o_p(\{n^{-1}\Sigma_0^n|\gamma_j|\}^{1/2})$, as in Theorem 2, and if (2.6) holds, then $n^{1/2-d}(\hat{\alpha}_\phi - \alpha)$ has the same limit distribution as $n^{1/2-d}\bar{u}$, and if Condition 1 holds this is $N(0, D/d(2d+1))$ [cf. Taqqu (1975)]. On the other hand, if $f(\lambda)$ is continuous and positive at $\lambda = 0$ and $\hat{\beta}_\phi$ is $n^{1/2}$-consistent (as in Theorem 1), then the last term on the rightmost expression in (2.14) also contributes, unless $E(x_1) = 0$, while it dominates if $\hat{\beta}_\phi$ is less-than-$n^{1/2}$-consistent.

3. Proof of Theorem 1. Dropping the ϕ subscripts from (1.2), (1.3), (1.17) and (1.18), write $\hat{\beta} - \beta = \hat{A}^{-1}\hat{a}$, $\tilde{\beta} - \beta = \tilde{A}^{-1}\tilde{a}$, where

$$\hat{a} = n^{-1}\sum_s\sum_t (x_t - \bar{x})(u_s - \bar{u})\phi_{t-s}, \quad \tilde{a} = n^{-1}\sum_{j=1}^{n-1} I_{xu}(\lambda_j)\phi(\lambda_j),$$

for

$$I_{xu}(\lambda) = (2\pi n)^{-1}\left(\sum_t x_t e^{it\lambda}\right)\left(\sum_t u_t e^{-it\lambda}\right).$$

With $Ex_1 = 0$, without loss of generality, write $A = n^{-1}\Sigma_s\Sigma_t x_t x'_s \phi_{t-s}$, $a = n^{-1}\Sigma_s\Sigma_t x_t u'_s \phi_{t-s}$. Define $r_t = \Sigma_s x_s \phi_{t-s}$, $s_u = \Sigma_t r_t \tau_{t-u}$, with $\tau_j = 0$, $j < 0$, and $a_1 = n^{-1}\Sigma_{-N}^n s_u \varepsilon_u$, $a_2 = a - a_1$, with $N = N_n$ yet to be chosen. Write

$$D_0 = E(naa' \,|\{x_t\}), \quad D_1 = E(na_1a'_1 \,|\{x_t\}), \quad D_2 = D_0 - D_1,$$
$$E = (2\pi)^{-1}\int_{-\pi}^{\pi} I_x(\lambda)\psi(\lambda)\,d\lambda.$$

Write

$$n^{1/2}\left(\hat{\beta} - \beta\right) = A^{-1}(A\hat{A}^{-1})\Sigma_\psi^{1/2}\left(\Sigma_\psi^{-1/2}E^{1/2}\right)\left(E^{-1/2}D_0^{1/2}\right)$$
$$\times\left(D_0^{-1/2}D_1^{1/2}\right)D_1^{-1/2}n^{1/2}(a_1 + a_2),$$
$$n^{1/2}\left(\tilde{\beta} - \hat{\beta}\right) = (\hat{A}^{-1} - \tilde{A}^{-1})n^{1/2}\hat{a} + \tilde{A}^{-1}n^{1/2}(\hat{a} - \tilde{a}),$$

where $X^{1/2}$ satisfies $(X^{1/2})^2 = X$, and noting that $E\|D_2\| \le E\|n^{1/2}a_2\|^2$, where $\|X\| = \mathrm{tr}^{1/2}(X'X)$, the proof for $\hat{\beta}$ follows immediately from Propositions 1–6 below, and that for $\tilde{\beta}$ from Propositions 1–5 and 7.

PROPOSITION 1. *For N increasing suitably with n,*

$$\lim_{n\to\infty} E\|n^{1/2}a_2\|^2 = 0.$$

PROOF. With C denoting a generic constant,

$$\begin{aligned} E\big(n\|a_2\|^2\big) &= \frac{\sigma^2}{n}\sum_{u=-\infty}^{-N-1}\sum_t\sum_s\sum_r\sum_q E(x_r x_q')\phi_{t-r}\phi_{s-q}\tau_{t-u}\tau_{s-u} \\ &\le \frac{C}{n}\Big(\sum_t\sum_s|\phi_{t-s}|\Big)^2\sum_N^\infty \tau_u^2 \le Cn\sum_N^\infty \tau_u^2, \end{aligned}$$

which tends to 0 as $n\to\infty$ for a suitable sequence N_n, by Condition 1. □

PROPOSITION 2. *As $n\to\infty$,*

$$D_1^{-1/2}n^{1/2}a_1 \to_d N(0,I), \tag{3.1}$$

where I is the identity matrix.

PROOF. Equation (3.1) is implied if the convergence holds conditional on $\{x_t\}$ and we establish the latter. For any vector z such that $\|z\|=1$ and any N, n, define $d_u = n^{-1/2}z'D_1^{-1/2}s_u$. Then from Scott (1973), it suffices to show that, as $n, N\to\infty$,

$$\sum_{-N}^{n} d_u^2 E\big(\varepsilon_u^2|F_{u-1}\big) \to_p 1, \tag{3.2}$$

$$E\Bigg(\sum_{-N}^{n} d_u^2 E\big[\varepsilon_u^2 I(|d_u\varepsilon_u|>\eta)|\{x_t\}\big]\Bigg) \to 0 \quad \text{for all } \eta>0. \tag{3.3}$$

Trivially, (3.2) follows from Condition 1 and $\sum_{-N}^n d_u^2 = 1/\sigma^2$. For $\delta>0$ the left-hand side of (3.3) is bounded by

$$E\Bigg[\sum_{-N}^{n} d_u^2 E\big[\varepsilon_u^2 I(|\varepsilon_u|>\eta/\delta)\big]\Bigg] + P\Big(\max_u|d_u|>\delta\Big). \tag{3.4}$$

The first term can be made arbitrarily small by choosing δ small enough, from uniform integrability. We now modify an argument of Eicker (1967). Propositions 1, 4 and 5 below and Condition 5 imply that $D_1^{-1}=O_p(1)$, so we can consider the set $\|D_1^{-1/2}\|\le C$, on which, for any $\varepsilon>0$,

$$\begin{aligned} \max_u|d_u| &\le \max_u Cn^{-1/2}\sum_t\|r_t\||\tau_{t-u}| \\ &\le C\Big\{\frac{\varepsilon}{n}\sum_t\|r_t\|^2\Big\}^{1/2} + CLn^{-1/2}\max_{1\le t\le n}\|r_t\|\max_u|\tau_u|, \end{aligned}$$

where $L = L(\varepsilon)$ is chosen such that $\sum_{u \geq L} \tau_u^2 < \varepsilon$. Then the proof of (3.3) is completed by

$$E\|r_t\|^2 \leq \sum_s \sum_u E\|x_s\|^2 |\phi_{t-s}\phi_{t-u}| \leq C < \infty,$$

$$\max_{1 \leq t \leq n} \|r_t\| \leq \max_{1 \leq t \leq n} \|x_t\| \sum |\phi_u| \leq C \left\{ \sum_t \|x_t\|^4 \right\}^{1/4} = O_p(n^{1/4}),$$

using Markov's inequality and Conditions 2 and 7. □

PROPOSITION 3. *$A \to_p \Sigma_\phi$ as $n \to \infty$.*

PROOF. Write $C_j = n^{-1} \sum_{1 \leq t, t+j \leq n} x_t x'_{t+j}$. By easy use of formula (30) on page 464 of Anderson (1971), it is seen that Condition 7 implies $E\|C_j - \Gamma_j\|^2 \to 0$ for all fixed j. With a simple truncation argument and Condition 2, it follows that $\sum_{1-n}^{n-1} \phi_j (C_j - \Gamma_j) \to_p 0$. But Condition 2 also implies that $\sum_{|j| \geq n} \phi_j \Gamma_j \to 0$. □

PROPOSITION 4. *$E \to_p \Sigma_\psi$ as $n \to \infty$.*

PROOF. Let $\psi_J(\lambda) = 2\pi \sum_{j=-J}^{J} (1 - |j|/J) \psi_j e^{-ij\lambda}$ be the Cesaro sum to J terms of the Fourier series of $\psi(\lambda)$, where $\psi_j = (2\pi)^{-2} \int_{-\pi}^{\pi} \psi(\lambda) \cos j\lambda \, d\lambda$, and write

$$\begin{aligned} E - \Sigma_\psi = {} & \frac{1}{2\pi} \int_{-\pi}^{\pi} I_x(\lambda) \{\psi(\lambda) - \psi_J(\lambda)\} \, d\lambda \\ & + \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_J(\lambda) \{I_x(\lambda) \, d\lambda - dF(\lambda)\} \\ & + \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\psi_J(\lambda) - \psi(\lambda)\} \, dF(\lambda). \end{aligned}$$

The second term on the right-hand side is

$$\sum_{j=-J}^{J} \psi_j \left(1 - \frac{|j|}{J}\right) (C_j - \Gamma_j) \to_p 0$$

for any fixed J, as in the proof of Proposition 3, while Condition 4 implies that for any $\varepsilon > 0$ we can choose J so large that $\sup_\lambda |\psi(\lambda) - \psi_J(\lambda)| < \varepsilon$ from Fejèr's theorem [Zygmund (1977), page 89] so that the norms of the first and last terms on the right-hand side are less than $(\varepsilon/2\pi)\mathrm{tr}(C_0)$ and $(\varepsilon/2\pi)\mathrm{tr}(\Gamma_0)$. □

PROPOSITION 5. *$E - D \to_p 0$ as $n \to \infty$.*

PROOF. We have

$$\psi_m = \sum \sum \phi_j \phi_k \gamma_{m-j-k},$$

so that

$$\begin{aligned} E - D &= \frac{1}{n} \sum_t \sum_u x_t x_u' \sum_{1-n}^{n-1} \gamma_j {\sum_q}' \phi_{t-j-q} \phi_{u-q} \\ &\quad + \frac{1}{n} \sum_t \sum_u x_t x_u' \sum_{|j| \geq n} \gamma_j \Delta_{t-j-u} \\ &= R_1 + R_2, \end{aligned}$$

where Σ_q' is a sum over $\{q > \min(n, n-j)\} \cup \{q < \max(1, 1-j)\}$ and $\Delta_j = \Sigma \phi_t \phi_{t-j}$. To prove $R_1 \to_p 0$ and $R_2 \to_p 0$, it is convenient to introduce first a series of lemmas, some of which will also be useful in proving subsequent properties. Let $a_{jtu} = \Sigma_{q > n-j} \phi_{t-j-q} \phi_{u-q}$.

LEMMA 1. *For $|r| < n$,*

$$\sum_t \Phi_{|r-t|/2} \leq 4 \sum_{t=0}^{n-1} \Phi_t.$$

PROOF. $\Phi_{|r-t|/2} = \Phi_{[|r-t|/2]}$, and for $1 \leq t \leq n$, $0 \leq r < n$ we have $0 \leq [\frac{1}{2}|r-t|] \leq n-1$, with duplications due to division by 2 and taking integer parts, and to positive and negative values of $r - t$. □

LEMMA 2.

$$\sum_{t=0}^{n-1} \Phi_t = \sum_t t|\phi_t| + n\Phi_n.$$

PROOF. Elementary.

LEMMA 3.

$$\sum_t t\tilde{\phi}_t \leq \left(\sum_t \tilde{\phi}_t^{1/2} \right)^2.$$

PROOF. $\Sigma_t t\tilde{\phi}_t \leq \Sigma_t \tilde{\phi}_t^{1/2} \Sigma_{u=1}^t \tilde{\phi}_u^{1/2} \leq (\Sigma_t \tilde{\phi}_t^{1/2})^2$. □

LEMMA 4. *Under Condition 2, for all integers j and all $t, u \in [1, n]$,*

$$|a_{jtu}| \leq C\tilde{\phi}_{n-t}^{1/2} \tilde{\phi}_{|t-j-u|/2}^{1/2} + \tilde{\phi}_{n-t} \Phi_{|t-j-u|/2}.$$

PROOF. The left-hand side is bounded by

$$\sum_{i \in A} |\phi_{t-j-u-i} \phi_{-i}| + \sum_{i \in B} |\phi_{t-j-u-i} \phi_{-i}|,$$

where $A = \{i: i > n - j - u, |i| \le \frac{1}{2}|t - j - u|\}$, $B = \{i: i > n - j - u, |i| > \frac{1}{2}|t - j - u|\}$. Because $i \in A$ or $i \in B$ imply $t - j - u - i < t - n \le 0$ while $i \in A$ implies $t - j - u - i \le \frac{1}{2}(t - j - u) \le 0$, we can bound the two sums respectively by $\tilde{\phi}_{n-t}^{1/2} \tilde{\phi}_{|t-j-u|/2}^{1/2} \Sigma |\phi_i|$ and $\tilde{\phi}_{n-t} \Phi_{|t-j-u|/2}$, whence the result follows from Condition 2. □

LEMMA 5. *For* $0 \le j < n$,

$$\sum_t \sum_u |a_{jtu}| \le C \left\{ \left(\sum_0^n \tilde{\phi}_t^{1/2} \right)^2 + n \Phi_n \right\}. \tag{3.5}$$

PROOF. Applying Lemma 4,

$$\begin{aligned} \sum_t \sum_u |a_{jtu}| &\le C \sum_t \tilde{\phi}_{n-t}^{1/2} \sum_u \tilde{\phi}_{|t-j-u|/2}^{1/2} + \sum_t \tilde{\phi}_{n-t} \sum_u \Phi_{|t-j-u|/2} \\ &\le C \left\{ \left(\sum_0^{n-1} \tilde{\phi}_t^{1/2} \right)^2 + \sum_t t |\phi_t| + n \Phi_n \right\} \end{aligned}$$

by Lemmas 1 and 2, where (3.5) follows from Lemma 3. □

LEMMA 6. $\sum_{j > 2n} |\Delta_j| \le C \Phi_n$.

PROOF. We have

$$\begin{aligned} \sum_{j > 2n} |\Delta_j| &\le \sum_{|q| \le n} |\phi_q| \sum_{j > 2n} |\phi_{q-j}| + \sum_{|q| > n} |\phi_q| \sum_{j > 2n} |\phi_{q-j}| \\ &\le \sum_{|q| \le n} |\phi_q| \sum_{j > n} |\phi_j| + 2 \sum_{q > n} |\phi_q| \sum |\phi_j| \\ &\le C \Phi_n. \end{aligned}$$

□

LEMMA 7. *For* $j > 0$,

$$|\Delta_j| \le C \tilde{\phi}_{j/2}.$$

PROOF.

$$\begin{aligned} |\Delta_j| &\le \sum_{|q| \le j/2} |\phi_q \phi_{q-j}| + \sum_{|q| > j/2} |\phi_q \phi_{q-j}| \\ &\le C \tilde{\phi}_{j/2}. \end{aligned}$$

□

Consider now

$$\begin{aligned} R_1 &= \sum_t \sum_u x_t x_u' b_{tu} + \sum_t \sum_u x_{n-t+1} x_{n-u+1}' b_{tu} \\ &\quad + \sum_t \sum_u x_u x_t' \tilde{b}_{tu} + \sum_t \sum_u x_{n-t+1} x_{n-u+1}' \tilde{b}_{tu} \\ &= R_{11} + R_{12} + R_{13} + R_{14}, \end{aligned}$$

where $b_{tu} = n^{-1}\sum_{j=0}^{n-1}\gamma_j a_{jtu}$, $\tilde{b}_{tu} = n^{-1}\sum_{j=1}^{n-1}\gamma_j a_{jtu}$. Then

$$|ER_1| \le \left|\sum_t \sum_u (\Gamma_{t-u} + \Gamma_{u-t}) b_{tu}\right| + \left|\sum_t \sum_u (\Gamma_{t-u} + \Gamma_{u-t}) \tilde{b}_{tu}\right|$$
$$\le \frac{4}{n} \sum_t \sum_u c_{t-u} \bar{b}_{tu},$$

where $c_j = \|\Gamma_j\|$, $\bar{b}_{tu} = \sum_{j=0}^{n-1}|\gamma_j a_{jtu}|$. By Lemma 5 and Condition 3,

$$\sum_t \sum_u \bar{b}_{tu} = O(n), \tag{3.6}$$

so that $ER_1 \to 0$ by Condition 7 and the Toeplitz lemma. Next

$$\begin{aligned} E\|R_1 - ER_1\|^2 &\le 4 \sum_{j=1}^{4} E\|R_{1j} - ER_{1j}\|^2 \\ &\le \sum_t \sum_u \sum_v \sum_w h_{tuvw} \bar{b}_{tu} \bar{b}_{vw}, \end{aligned} \tag{3.7}$$

where

$$h_{tuvw} = c_{t-v}c_{u-w} + c_{t-w}c_{u-v} + \sum\sum\sum\sum_{a,b,c,d=1}^{K} |\kappa_{abcd}(t,u,v,w)|$$
$$= h_{tuvw}^{(1)} + h_{tuvw}^{(2)} + h_{tuvw}^{(3)}.$$

By Condition 7, for any $\varepsilon > 0$ there exists $L = L(\varepsilon)$ such that $h_{tuvw}^{(1)} < \varepsilon$ if $\max(|t-v|, |u-w|) \ge L$, $h_{tuvw}^{(2)} < \varepsilon$ if $\max(|t-w|, |u-v|) \ge L$ and $h_{tuvw}^{(3)} < \varepsilon$ if $\max(|t-u|, |t-v|, |t-w|, |u-v|, |u-w|, |v-w|) \ge L$. Thus, because $\sum_t\sum_u\sum_v\sum_w \bar{b}_{tu}\bar{b}_{vw} = O(n^2)$ from (3.6), it follows that $E\|R_1 - ER_1\|^2 \to 0$, and then $R_1 \to_p 0$. Now consider R_2. We have

$$|ER_2| \le \frac{1}{n} \sum_t \sum_u c_{t-u} d_{tu},$$

where $d_{tu} = \sum_{|j|>n}|\gamma_j \Delta_{t-j-u}|$. Applying Lemmas 3, 6 and 7 and Condition 3,

$$\begin{aligned} \frac{1}{n}\sum_t \sum_u d_{tu} &\le \tilde{\gamma}_n \sum_{1-n}^{n-1} \sum_{|j|>n} |\Delta_{r-j}| \\ &\le \tilde{\gamma}_n \left(\sum_{1}^{2n} |j\Delta_j| + n \sum_{j>2n} |\Delta_j| \right) \\ &\le C\tilde{\gamma}_n \left(\sum_{1}^{2n} j\tilde{\phi}_j + n\Phi_n \right) = O(1). \end{aligned} \tag{3.8}$$

Thus $ER_2 \to 0$ by the Toeplitz lemma. Next

$$E\|R_2 - ER_2\|^2 \le \frac{1}{n^2} \sum_t \sum_u \sum_v \sum_w h_{tuvw} |d_{tu} d_{vw}| \to 0$$

from (3.8) and the properties of h_{tuvw} previously indicated and the Toeplitz lemma. □

PROPOSITION 6. *As* $n \to \infty$,

$$n^{1/2}(\hat{a} - a) \to_p 0, \qquad \hat{A} - A \to_p 0.$$

PROOF. Put $n^{1/2}(\hat{a} - a) = b_1 + b_2 + b_3$, where

$$b_1 = -n^{-1/2} \bar{x} \sum_t \sum_s u_s \phi_{t-s}, \quad b_2 = -n^{-1/2} \bar{u} \sum_t \sum_s x_t \phi_{t-s}$$

and

$$b_3 = \bar{u}\bar{x} n^{-1/2} \sum_t \sum_s \phi_{t-s}.$$

Note first that

$$\bar{u} = O_p\left(\left\{ n^{-1} \sum_{1-n}^{n-1} (1 - |j|/n) \gamma_j \right\}^{1/2} \right), \qquad \bar{x} \to_p 0, \tag{3.9}$$

where the second result uses Condition 7 and the Toeplitz lemma. First suppose $\phi(0) = 0$. Because $\phi(\lambda) = 2\pi \sum \phi_j \cos j\lambda$,

$$\begin{aligned}
\sum_t \sum_s \phi_{t-s} &= \sum_{1-n}^{n-1} (n - |j|) \phi_j \\
&= -2nq_n - 2 \sum_t t\phi_t \\
&= O\left(\left\{ \sum_0^n \tilde{\phi}_t^{1/2} \right\}^2 + n\Phi_n \right)
\end{aligned}$$

by Lemma 3, where $q_n = \sum_{t > n} \phi_t$. Thus $b_3 \to_p 0$ by (3.9) and Condition 3. Next, $\sum_t \sum_s x_t \phi_{t-s} = -\sum_t x_t (q_{t+1} + q_{n-t})$, so that

$$E\left\| \sum_t \sum_s x_t \phi_{t-s} \right\|^2 = \sum_{1-n}^{n-1} \operatorname{tr}(\Gamma_j) P_j,$$

where

$$P_j = \sum_{\max(1, 1-j)}^{\min(n, n-j)} (q_{t+1} + q_{n-t})(q_{t+j+1} + q_{n-t-j}).$$

Because $P_j = O(\sum_0^{n+1} \Phi_t)$ uniformly in j, it follows from Lemmas 2 and 3, (3.9), Condition 3 and the Toeplitz lemma that $b_2 \to_p 0$. Similarly,

$$E\left(\sum_t \sum_s u_s \phi_{t-s}\right)^2 = \sum_{1-n}^{n-1} \gamma_j P_j = O(n),$$

and then $b_1 \to_p 0$ by (3.9). Now suppose $\phi(0) \neq 0$, when Condition 4 implies $f(\lambda)$ is continuous at $\lambda = 0$. Thus, from (3.9) and Fejér's theorem, $\bar{u} = O_p(n^{-1/2})$, and because Condition 2 implies $\Sigma_t \Sigma_s \phi_{t-s} = O(n)$, it follows that $b_3 \to_p 0$. Because

$$E\left\| \sum_t \sum_s x_t \phi_{t-s} \right\|^2 \leq \left(\sum |\phi_j|\right)^2 \sum_t \sum_s c_{t-s} = o(n^2),$$

we then have $b_2 \to_p 0$. To deal with b_1,

$$E\left(\sum_t \sum_s u_t \phi_{t-s}\right)^2 = \sum_t \sum_s \phi_{t-s} \sum_v \sum_{j=t-n}^{t-1} \gamma_j \phi_{t-v-j} = \sum_{i=1}^{5} d_i,$$

where

$$d_1 = \sum_t \sum_s \psi_{t-s},$$

$$d_2 = -\sum_{t>n} \sum_s \phi_{t-s} \sum_v \zeta_{t-v},$$

$$d_3 = -\sum_{t\leq 0} \sum_s \phi_{t-s} \sum_t \zeta_{s-v},$$

$$d_4 = -\sum_t \sum_s \phi_{t-s} \sum_{j\geq t} \gamma_j \sum_v \phi_{t-v-j},$$

$$d_5 = -\sum_t \sum_s \phi_{t-s} \sum_{j<t-n} \gamma_j \sum_v \phi_{t-v-j},$$

where $\zeta_k = \Sigma \gamma_j \phi_{k-j}$. Now $d_1 = (n/2\pi)\psi_n(0) = O(n)$ by Condition 4, as in the proof of Proposition 4. Uniformly in t.

$$\sum_v |\zeta_{t-v}| \leq \sum_v \left\{ \sum_{-n}^{n} |\gamma_j \phi_{t-v-j}| + \sum_{j>n} |\gamma_j \phi_{t-v-j}| + \sum_{j<-n} |\gamma_j \phi_{t-v-j}| \right\}$$

$$= O\left(\sum_0^n |\gamma_j| + n\tilde{\gamma}_n \right),$$

so that d_2 and d_3 are $O(n)$ by Lemma 3 and Condition 3. It is easily seen that

$$|d_4| \leq C \sum_t \sum_{j\geq t} |\gamma_j| \sum_v |\phi_{t-v-j}|$$

$$\leq \left(\sum_j |\gamma_j| + n\tilde{\gamma}_n \right) \left\{ \sum_j j|\phi_j| + n\Phi_n \right\} = O(n)$$

by Condition 3. In the same way $d_5 = O(n)$. Thus $b_1 \to_p 0$ by (3.9). The proof that $\hat{A} - A \to_p 0$ is omitted because it follows from similar but more straightforward calculations to those above after writing

$$\hat{A} = A = \overline{x}\overline{x}'\frac{1}{n}\sum_t\sum_s \phi_{t-s} - \overline{x}\frac{1}{n}\sum_t\sum_s x_s'\phi_{t-s} - \frac{1}{n}\sum_t\sum_s x_t\phi_{t-s}\overline{x}'. \qquad \square$$

PROPOSITION 7. *As* $n \to \infty$,

$$n^{1/2}(\tilde{a} - a) \to_p 0, \qquad \tilde{A} - A \to_p 0.$$

PROOF. Because

$$\sum_{j=0}^{n-1} \exp(i\ell\lambda_j) = \begin{cases} n, & \ell = 0, \bmod n, \\ 0, & \text{otherwise}, \end{cases}$$

we have

$$n^{1/2}(\tilde{a} - a) = n^{-1/2}\sum_t\sum_s x_t u_s \omega_{t-s} - n^{1/2}\overline{x}\overline{u}\sum \phi_j,$$

where $\omega_j = \sum_{|i|\geq 1}\phi_{j+in}$. The second term on the right-hand side is identically 0 if $\phi(0$ is $0)$; otherwise, because Condition 4 implies $f(0) < \infty$, and thus $\overline{u} = O_p(n^{-1/2})$, it is $o_p(1)$. The first term has mean 0 and variance bounded by

$$\frac{1}{n}\sum_s\sum_t\sum_u\sum_v c_{t-u}|\gamma_{s-v}\,\omega_{t-s}\,\omega_{u-v}|.$$

By the Toeplitz lemma it suffices to show that

$$\sum_s\sum_t\sum_u\sum_v|\gamma_{s-v}\,\omega_{t-s}\,\omega_{u-v}| = O(n).$$

The left-hand side is $O(\sum_{j=0}^{n}|\gamma_j|\sum_{1-n}^{n-1}|\omega_j|\sum_t\sum_s|\omega_{t-s}|)$. In view of Condition 3 and Lemma 3, the proof that $n^{1/2}(\tilde{a} - a) \to_p 0$ is completed by the following lemmas.

LEMMA 8.

$$\sum_{j=1-n}^{n-1}|\omega_j| < \infty.$$

PROOF. The left-hand side is bounded by $2\sum|\phi_j| < \infty$. $\square$

LEMMA 9.

$$\sum_t\sum_s|\omega_{t-s}| \leq C\left(\sum_j j|\phi_j| + n\Phi_n\right).$$

PROOF. The left-hand side is bounded by

$$2\sum_{j=1}^{n-1} j|\phi_j| + 4\sum_{\ell\geq 1}\sum_{\ell n}^{(\ell+1)n-1}|\phi_j|(n - |j - \ell n|).$$

The second term is bounded by $4n\Phi_n$. $\square$

Finally, the proof that $\tilde{A} - A \to_p 0$ follows from

$$\tilde{A} - A = \frac{1}{n} \sum_t \sum_s x_t x_s' \omega_{t-s} - \bar{x}\bar{x}' \sum \phi_j.$$

The second term is clearly $o_p(1)$. The first has mean $O(n^{-1}\sum_t \sum_s c_{t-s}|\omega_{t-s}|) \to 0$ from the Toeplitz lemma and Lemma 8. The latter property is also used to show that the variance of the first term tends to 0, on dealing with this term as we did with R_1 and R_2 in the proof of Proposition 5. □

4. Feasible generalized least squares. We now wish to estimate β efficiently in the absence of knowledge of f up to scale. We write $f(\lambda) = f(\lambda; \theta_0)$, where $f(\lambda; \theta)$ is a known function of λ and the p-dimensional vector θ, and θ_0 is unknown. For an estimate $\hat{\theta}$ of θ_0, define $\hat{\beta}_{\hat{f}^{-1}}, \tilde{\beta}_{\hat{f}^{-1}}$ by putting $\phi(\lambda) = f^{-1}(\lambda; \hat{\theta})$ in $\hat{\beta}_\phi$ and $\tilde{\beta}_\phi$. If θ includes the scale factor of f, and this is functionally unrelated to the remaining elements of θ, then these estimates of β are invariant to it, but the scale factor has to be estimated in order to estimate the limiting covariance matrix $\Sigma_{f^{-1}}^{-1}$.

Now introduce the following condition.

CONDITION 8.

$$\hat{\theta} - \theta_0 = O_p\left(\left\{ \sum_0^n |\gamma_j| \right\}^{-1/2} \right) \tag{4.1}$$

and, defining $N_\delta = \{\theta: \|\theta - \theta_0\| < \delta\}$, there exists $\delta > 0$ such that

$$\phi_j(\theta) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} f(\lambda; \theta)^{-1} \cos j\lambda \, d\lambda$$

is differentiable in N_δ, for each j, and

$$\sum \sup_{\theta \in N_\delta} \left\| \frac{\partial \phi_j(\theta)}{\partial \theta} \right\| < \infty. \tag{4.2}$$

THEOREM 3. *Under* (1.1) *and Conditions* 1–8, *it follows that*

$$n^{1/2}\left(\hat{\beta}_{\hat{f}^{-1}} - \beta\right), \qquad n^{1/2}\left(\tilde{\beta}_{\hat{f}^{-1}} - \beta\right) \to_d N\left(0, \Sigma_{f^{-1}}^{-1}\right), \tag{4.3}$$

and $\Sigma_{f^{-1}}$ *is consistently estimated by both*

$$\frac{1}{n} \sum_t \sum_s (x_s - \bar{x})(x_t - \bar{x})' \phi_{t-s}(\hat{\theta}) \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n-1} I_x(\lambda_j) \phi\left(\lambda_j; \hat{\theta}\right). \tag{4.4}$$

The proof is omitted because it makes straightforward use of Theorem 1 and its proof, the mean value theorem and the fact that

$$E\|n^{-1} \sum_{1 \le t,\, t+j \le n} (x_t - Ex_1) u_t\|^2 = o\left(n^{-1} \sum_{j=0}^{n-1} |\gamma_j| \right)$$

as $n \to \infty$ uniformly in j, to show that $\hat{\beta}_{\hat{f}^{-1}} - \hat{\beta}_{f^{-1}}$ and $\tilde{\beta}_{\hat{f}^{-1}} - \tilde{\beta}_{f^{-1}}$ are $o_p(n^{1/2})$. We discuss Condition 8, however.

In the simple model

$$\phi_j(\theta) = \theta_1 j^{-2\theta_2 - 1}, \qquad j = 1, 2, \dots, \ \theta_1 > 0, 0 < \theta_2 < \tfrac{1}{2},$$

(4.2) clearly holds. More generally, Lemma 5 of Fox and Taqqu (1986) establishes (4.2) under certain regularity conditions. These conditions entail a singularity in $f(\lambda; \theta)$ at $\lambda = 0$ only, and are satisfied in the FARIMA case (2.9). It seems likely that Fox and Taqqu's (1986) result can be extended to enable (4.2) to be checked in case of singularities at nonzero frequencies.

Condition (4.1) is clearly milder than the requirement

$$\hat{\theta} - \theta_0 = O_p(n^{-1/2}). \tag{4.5}$$

It is necessary to say something about how $\hat{\theta}$ is obtained. First, pretend that the u_t are observable, and denote by $\tilde{\theta}$ an estimate of θ_0 based on them. In any given model any number of $\tilde{\theta}$ is, in principle, available, some of which have been explicitly discussed in the literature, and for some of these, rates of convergence have been obtained. For example, defining $I(\lambda) = (2\pi n)^{-1} |\sum_{t=1}^{n} (u_t - \bar{u}) e^{it\lambda}|^2$, consider $\tilde{\theta}$ minimizing

$$\int_{-\pi}^{\pi} \left\{ \log f(\lambda; \theta) + \frac{I(\lambda)}{f(\lambda; \theta)} \right\} d\lambda. \tag{4.6}$$

Conditions for $\tilde{\theta} - \theta_0 = O_p(n^{-1/2})$ have here been given by Fox and Taqqu (1986), Dahlhaus (1989) and Giraitis and Surgailis (1991) in case f has a singularity at $\lambda = 0$ only, while Dahlhaus (1989) also considered approximations to (4.6) which lead to the same rate of convergence. Hosoya (1993) has similar results in case of singularities at known, nonzero frequencies; see (2.10). It appears that $n^{1/2}$-consistency can be achieved by many other types of estimates of parametric models for long-range dependence. While no rigorous asymptotics exist for it, this seems true of the explicitly defined estimate (or a modification thereof) of Kashyap and Eom (1988) for the FARIMA $(0, d, 0)$ model given by (2.9) with $a = b \equiv 1$, and an extension of this type of estimate to the model

$$f(\lambda; \theta) = \exp\left[\sum_{j=1}^{p-1} \theta_j \cos\{(j-1)\lambda\} \right] |1 - e^{i\lambda}|^{-2\theta_p}$$

proposed by Robinson (1994a).

Now suppose that for some such $\tilde{\theta}$ we define $\hat{\theta}$ correspondingly after replacing u_t in $\tilde{\theta}$ by

$$\hat{u}_t = y_t - \bar{y} - \hat{\beta}_{\phi}'(x_t - \bar{x}) \tag{4.7}$$

for some ϕ. (Of course, $\tilde{\beta}_{\phi}$ can be used in place of $\hat{\beta}_{\phi}$.) Then, given that

$\tilde{\theta} - \theta_0 = O_p(\{\Sigma_0^n|\gamma_j|\}^{-1/2})$, it suffices to show that $\hat{\theta} - \tilde{\theta} = O_p(\{\Sigma_0^n|\gamma_j|\}^{-1/2})$. Unfortunately, this requires a somewhat case-by-case treatment, and in a given case a reasonably detailed proof may be lengthy, especially for estimates that are only implicitly defined, where a preliminary consistency proof for $\hat{\theta}$ may be needed; this is true of estimates minimizing (4.6) with $I(\lambda)$ replaced by $\hat{I}(\lambda) = (2\pi n)^{-1}|\Sigma_t \hat{u}_t e^{it\lambda}|^2$, which are among those of most interest because of their efficiency under Gaussianity. There seems little interest in presenting the details in even a single case, especially as the statistical literature contains many demonstrations that errors can be replaced by suitable residuals without affecting first-order asymptotics, and which suggest that the same is likely to be true in our problem, at least if $\hat{\beta}_\phi$ is $n^{1/2}$-consistent, because $\hat{u}_t = u_t - \bar{u} - (\hat{\beta}_\phi - \beta)'(x_t - \bar{x})$.

It appears that (4.1) is capable of significant relaxation, subject to additional regularity conditions [such as on higher-order derivatives of the $\phi_j(\theta)$] and that a slower rate of convergence of $\hat{\beta}_\phi$ would suffice, such that possibly least squares $\hat{\beta}_1$ (see Theorem 2) could be used in (4.7). However, verification of these conjectures would not only be lengthy but of limited practical value because it seems desirable in finite-sample practice to use estimates of θ_0 and β in $\hat{\beta}_{\hat{f}^{-1}}$ or $\tilde{\beta}_{\hat{f}^{-1}}$ with the maximum, $n^{1/2}$, rate of convergence, and Theorem 1 and our recent discussion indicate that this is likely to be possible. In practice, we may also wish to iterate the above practice, employing $\hat{\beta}_{\hat{f}^{-1}}$ or $\tilde{\beta}_{\hat{f}^{-1}}$ to recalculate residuals (4.7) and thence new estimates of β which would have no greater asymptotic efficiency but might have better finite-sample properties.

5. Nonlinear regression models. An important extension of (1.1) consists of regression models with nonlinearity in parameters and possibly in regressors also. Consider

$$y_t = \alpha + z_t(\beta_0) + u_t, \qquad t = 1, 2, \dots, \tag{5.1}$$

where $z_t(\beta)$ is a given function of the vector β and of a stochastic regression vector observed at time t, whose presence is indicated only by the t-subscript. Note that β is zero-subscripted, as was θ in the previous section, because of the need to explicitly define an objective function for estimation in nonlinear problems. As an example of $z_t(\beta)$, consider the multiplicative form

$$z_t(\beta) = \beta_1 \prod_{i=2}^{K} x_{it}^{\beta_i}. \tag{5.2}$$

Following the work of Jennrich (1969) and Malinvaud (1970), on nonlinear least squares in case of independent u_t, authors such as Gallant and Goebel (1975), Hannan (1971) and Robinson (1972) studied asymptotic theory for estimates of models such as (5.1) in case of weakly dependent u_t. We allow u_t to have long-range dependence. Nonlinear extremum estimates in regression models with long-range-dependent errors have been discussed by Koul (1992), Koul and Muhkerjee (1993) and Robinson (1994a). However, the regression models considered in the former two references are linear ones, the nonlinearity coming from the nonquadratic objective functions used, and the conditions and theoretical treatment are very different from ours, while the latter reference provides only a heuristic account of broad issues.

Define

$$\hat{\beta}_\phi = \arg\min_B \hat{Q}_\phi(\beta),$$

$$\hat{Q}_\phi(\beta) = \int_{-\pi}^{\pi} \left\{ |w_y(\lambda) - w_z(\lambda;\beta)|^2 - I_y(\lambda) \right\} \phi(\lambda)\, d\lambda$$

$$= \int_{-\pi}^{\pi} \left[I_z(\lambda;\beta) - 2R\left\{ I_{zy}(\lambda;\beta) \right\} \right] \phi(\lambda)\, d\lambda,$$

$$I_z(\lambda;\beta) = |w_z(\lambda;\beta)|^2, \qquad I_{zy}(\lambda;\beta) = w_z(\lambda;\beta) w_y(-\lambda),$$

$$w(\lambda;\beta) = \frac{1}{(2\pi n)^{1/2}} \sum_t \{ z_t(\beta) - \bar{z}(\beta) \} e^{it\lambda}, \qquad \bar{z}(\beta) = \frac{1}{n} \sum_t z_t(\beta),$$

where B is a compact subset of $\Re^K$. As in Section 1 it can be argued that the clt with $n^{1/2}$ convergence rate can fail for the nonlinear least squares choice $\phi(\lambda) \equiv 1$, whereas $\phi = f^{-1}$ is an efficient choice among those ϕ which do lead to $n^{1/2}$-consistency.

This setup does not quite cover models with infinite-dimensional leads or lags, such as where

$$z_t(\beta) = \sum \chi_j(\beta) x_{t-j}, \tag{5.3}$$

where the regression vector x_t is observed only at $t = 1, \dots, n$ and the row vector $\chi_j(\beta)$ can be nonlinear in β; for example, a popular choice has been the geometric weights

$$\chi_j(\beta) = \begin{cases} \beta_1 \beta_2^j, & j \geq 0, \\ 0, & j < 0, \end{cases} \qquad |\beta_2| < 1,\ K = 1. \tag{5.4}$$

However, we can replace $z_t(\beta)$ in (5.3) by

$$\tilde{z}_t(\beta) = \sum_{t-n}^{t-1} \chi_j(\beta) x_{t-j}, \tag{5.5}$$

and it is easily shown that the error in $\hat{\beta}_\phi$ is no more than

$$O_p\left(n^{-1} \sum_{j=-n}^{n} |j| \, \| \chi_j(\beta_0) \| + \sum_{|j|>n} \| \chi_j(\beta_0) \| \right),$$

and by the Kronecker lemma we can apply (5.10) of Condition 14 below to show that this is $o_p(n^{-1/2})$, as desired.

We define

$$\tilde{\beta}_\phi = \arg\min_B \tilde{Q}(\beta),$$

$$\tilde{Q}_\phi(\beta) = \frac{1}{n} \sum_{j=1}^{n-1} \left[I_z(\lambda_j;\beta) - 2R\left\{ I_{zy}(\lambda_j;\beta) \right\} \right] \phi(\lambda_j).$$

In case (5.3) we may replace $\tilde{Q}_{\phi}(\beta)$ by

$$(5.6)\qquad \frac{1}{n}\sum_{j=1}^{n-1}\left[\chi(\lambda_j;\beta)I_x(\lambda_j)\chi'(-\lambda_j;\beta) - 2R\{\chi(\lambda_j;\beta)I_{xy}(\lambda_j)\}\right]\phi(\lambda_j),$$

where

$$\chi(\lambda;\beta) = \sum \chi_j(\beta)e^{ij\lambda},$$

the frequency response function $\chi(\lambda;\beta)$ typically being of simpler form than the $\chi_j(\beta)$ and written down by inspection, for example $\chi(\lambda;\beta) = \beta_1(1 - \beta_2 e^{i\lambda})^{-1}$ in case (5.4).

We now introduce some additional conditions. Define

$$N_\delta(\beta) = \{b: b \in B, \|b - \beta\| < \delta\}.$$

CONDITION 9. β_0 is an interior point of the compact set B.

CONDITION 10. For all $\beta \in B$, $\{z_t(\beta)\}$ is strictly stationary, $z_1(\beta)$ is continuous, there exists $\delta > 0$ such that

$$(5.7)\qquad E\Big(\sup_{b \in N_\delta(\beta)} z_1^2(b)\Big) < \infty,$$

and, for all $b, \beta \in B$,

$$(5.8)\qquad \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\gamma_j^2(b,\beta) = 0,$$

where $\gamma_j(b,\beta) = \operatorname{cov}\{z_1(b), z_{1+j}(\beta)\}$, and, for all k,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\left|\operatorname{cum}\{z_1(b), z_{1+j}(b), z_{1+k}(\beta), z_{1+j+k}(\beta)\right| = 0.$$

CONDITION 11. $\{z_t(\beta)\}$ and $\{\varepsilon_t\}$ are independent for all $\beta \in B$.

CONDITION 12. For all $\beta \in B \setminus \{\beta_0\}$,

$$(5.9)\qquad \begin{aligned}\int_{-\pi}^{\pi}\phi(\lambda)\,d\{G(\lambda;\beta,\beta) &- G(\lambda;\beta,\beta_0)\\ &-G(\lambda;\beta_0,\beta) + G(\lambda;\beta_0,\beta_0)\} > 0,\end{aligned}$$

where, for all $b, \beta \in B$, G is given by

$$\gamma_j(b,\beta) = \int_{-\pi}^{\pi} e^{ij\lambda}\,dG(\lambda;b,\beta),$$

and the matrix

$$\begin{bmatrix} G(\lambda; b, b) & G(\lambda; \beta, b) \\ G(\lambda; \beta, b) & G(\lambda; \beta, \beta) \end{bmatrix}$$

has Hermitian nonnegative definite increments and is continuous from the right.

CONDITION 13. $z_t(\beta)$ is twice differentiable in β, $(\partial/\partial\beta)z_1(\beta)$ and $(\partial^2/\partial\beta\,\partial\beta')z_1(\beta)$ are continuous at β_0 and there exists $\delta > 0$ such that

$$E\left\{\sup_{\beta\in N_\delta(\beta_0)}\left\|\frac{\partial}{\partial\beta}z_1(\beta)\right\|^2 + \sup_{\beta\in N_\delta(\beta_0)}\left\|\frac{\partial^2 z_1(\beta)}{\partial\beta\,\partial\beta'}\right\|^2\right\} < \infty.$$

CONDITION 14. In case $z_t(\beta)$ is given by (5.3), and $z_t(\beta)$ is replaced by (5.5) in $\hat{\beta}_\phi$, or $\tilde{Q}_\phi(\beta)$ is replaced by (5.6), we have $\max_t E\|x_t\|^2 < \infty$ and

$$\sum |j|^{1/2} \| \chi_j(\beta_0)\| < \infty. \tag{5.10}$$

We will apply Conditions 5 and 7 and the notation in (1.8) with x_t there replaced by

$$\frac{\partial z_t(\beta_0)}{\partial\beta}. \tag{5.11}$$

Conditions 9–11 are introduced partly for the consistency proof which is needed as a preliminary to the clt's in Theorems 4 and 5 below. Condition 9 is fairly standard in asymptotic theory for extremum estimates, and can be removed with respect to a multiplicative scale parameter in $z_t(\beta)$. Conditions 10 and 13 can be checked for (5.2) if the x_{it} are positive, bounded away from 0, depend only on a stationary and ergodic Gaussian vector and satisfy a suitable moment condition. The components (5.8) of Condition 10, and (2.2) of Condition 7 with (5.11) substituted, can more readily be explained in terms of conditions on the observables on which $z_t(\beta)$ (in general nonlinearly) depends than can "Grenander's conditions," linear processes or unit roots; (5.8) is equivalent to the absence of jumps in $G(\lambda; b, \beta)$, and is implied if $\gamma_j(b, \beta) \to 0$ as $j \to \infty$. Conditions 10 and 13 are of a different character from the corresponding ones of Hannan (1971) and Robinson (1972), which are expressed in terms of limits. Condition 11 holds in (5.2) under Condition 6, and, in general, implies that the derivatives in Condition 13 will also be independent of the ε_t. Condition 12 is an inevitable identifiability condition which, like Condition 7 with (5.11), is violated in (5.2) if, for example, $x_{i1} = x_{j1}$ a.s., for some $i \neq j$.

THEOREM 4. *Under* (5.1) *and Conditions* 1–5, 7 *and* 9–14, *it follows that* (2.4) *holds, and thus* (2.5) *holds in case* $f(\lambda) > 0$ *and* $\phi(\lambda) = f(\lambda)^{-1}$.

THEOREM 5. *Under* (5.1) *and Conditions* 1–5 *and* 7–14 *and* $f(\lambda) > 0$ *for all* λ, *it follows that* (4.3) *holds and* $\Sigma_{f^{-1}}$ *is consistently estimated by* (4.4) *with* x_t *replaced by* $(\partial/\partial\beta)z_t(\hat{\beta}_{f^{-1}})$.

It suffices to indicate how the proof of Theorem 4 differs from the many other proofs for extremum estimates, and how it uses Theorem 1. We discuss only $\hat{\beta}_\phi$. The first step is to show that $\hat{\beta}_\phi \to_p \beta_0$, where the broad argument is similar to that of, for example, Malinvaud (1970); see also the proofs of a.s. convergence of Jennrich (1969) and others. Writing $Q_\phi(\beta)$ in time domain form using (1.5), because of Conditions 3 and 13 it suffices to show that, for all fixed j,

$$(5.12) \qquad \frac{1}{n}\sum_t{}' z_t(\beta) u_{t+j} \to_p 0, \qquad \frac{1}{n}\sum_t{}' z_t(b) z_{t+j}(\beta) \to_p \gamma_j(b, \beta)$$

uniformly in $b, \beta \in B$, where Σ_t' is a sum over $1 \le t,\ t+j \le n$. Pointwise mean square convergence is an easy consequence of Conditions 2 and 10, and uniformity follows from this, Condition 9 and equicontinuity resulting from the Cauchy inequality and the fact the $\sup_{b \in N_\delta(\beta)} |z_1(b) - z_1(\beta)|^2$ for all $\beta \in B$ as $\delta \to 0$, which is due to continuity, the domination condition (5.5) and a routine extension of DeGroot (1970), page 206. Next we have

$$0 = \frac{\partial Q_\phi(\hat{\beta}_\phi)}{\partial \beta} = \frac{\partial Q_\phi(\beta_0)}{\partial \beta} + \tilde{H}(\hat{\beta}_\phi - \beta_0),$$

with probability approaching 1 as $n \to \infty$, where $\tilde{H}$ is a matrix with ith row $(\partial^2/\partial\beta_i\, \partial\beta')Q_\phi(\beta^{(i)})$ such that $\|\beta^{(i)} - \beta_0\| \le \|\hat{\beta}_\phi - \beta_0\|$, $i = 1, \dots, K$. In view of (5.11) we can rewrite $(\partial/\partial\beta)Q_\phi(\beta)$ as $-2\hat{a}$ (where $\hat{a}$ is defined at the start of Section 3) so after applying Theorem 1 it suffices to show that $2\hat{A} - \tilde{H} \to_p 0$. This follows straightforwardly from consistency of $\hat{\beta}_\phi$ and Condition 13 using techniques in the proof of (5.12).

6. Monte Carlo simulations. Finite-sample performance of estimates was investigated in a small Monte Carlo study. In the linear model (1.1) we took $K = 1$ and $\alpha = 0$, $\beta = 1$; our results are invariant to the choice of α and β. The scalar processes x_t and u_t were both Gaussian FARIMA's with spectra [see (2.9)] $dF(\lambda)/d\lambda = (2\pi)^{-1}|1 - e^{i\lambda}|^{-2c}$ and $f(\lambda) = (2\pi)^{-1}|1 - e^{i\lambda}|^{-2d}$, for the grid of values $c, d = 0.05(0.1)0.45$. In $\hat{\beta}_\phi$ and $\tilde{\beta}_\phi$ we took $\phi(\lambda) = |2\sin\frac{1}{2}\lambda|$, which satisfies our conditions for all the above c, d, noting that $\phi_j = 2/\{\pi^2(1 - 4j^2)\}$, $j = 0, \pm 1, \dots$. The asymptotic relative efficiency is

$$\frac{\{\Gamma(2-2c)\Gamma(1-c+d)\Gamma(2-c-d)\}^2}{\Gamma(3/2-c)^4\Gamma(1-2c+2d)\Gamma(3-2c-2d)},$$

which is given in Table 1. The efficiency increases in d and decreases in c, such that it is effectively 100% for all c when $d = 0.45$, and is generally very satisfactory for the other larger values of d, and falls below 50% only when $c = 0.45$, $d = 0.05$. Notice that least squares is asymptotically normal for all cases above the south–west/north–east diagonal.

TABLE 1
Asymptotic relative efficiency of $\hat{\beta}_\phi$ or $\tilde{\beta}_\phi$ to $\hat{\beta}_{f^{-1}}$ or $\tilde{\beta}_{f^{-1}}$ with $\phi(\lambda) = |2 \sin \lambda/2|$

		d				
		0.05	**0.15**	**0.25**	**0.35**	**0.45**
	0.05	0.8295	0.8992	0.9494	0.9820	0.9980
	0.15	0.7792	0.8701	0.9351	0.9769	0.9975
c	0.25	0.7039	0.8270	0.9139	0.9695	0.9966
	0.35	0.5850	0.7597	0.8811	0.9580	0.9954
	0.45	0.3823	0.6470	0.8268	0.9392	0.9933

We generated 1000 replications of series x_t and u_t of lengths $n = 64$, 128 and 256, via the algorithm of Davies and Harte (1987), noting that x_t and u_t have autocovariances

$$\Gamma_j = \frac{(-1)^j \Gamma(1-2c)}{\Gamma(1+j-c)\Gamma(1-j-c)},$$

$$\gamma_j = \frac{(-1)^j \Gamma(1-2d)}{\Gamma(1+j-d)\Gamma(1-j-d)}, \qquad j = 0, \pm 1, \dots$$

[see Adenstedt (1974)]. Deriving the y_t from (1.1), we then computed $\hat{\beta}_\phi$, $\hat{\beta}_{\hat{f}^{-1}}$, $\tilde{\beta}_\phi$ and $\tilde{\beta}_{\hat{f}^{-1}}$ in each case, where

$$\hat{d} = \arg \min_{\varepsilon \le \delta \le 1/2 - \varepsilon} \sum_{j=1}^{n-1} \frac{\left|\sum_{t=1}^{n}\left(y_t - \hat{\beta}_\phi x_t\right) e^{it\lambda_j}\right|^2}{f(\lambda_j; \delta)}$$

for $f(\lambda; \delta) = (2\pi)^{-1}|1 - e^{i\lambda}|^{-2\delta}$ and $\varepsilon = 0.001$.

Table 2 displays the ratio of the asymptotic variance of $\hat{\beta}_{f^{-1}}$, namely,

$$n^{-1}\Sigma_{f^{-1}}^{-1} = \Gamma(1-c+d)^2 / n\Gamma(1-2c+2d),$$

to the Monte Carlo mean-squared errors of $\hat{\beta}_\phi$ and $\tilde{\beta}_\phi$. The differences between the $\hat{\beta}_\phi$ and $\tilde{\beta}_\phi$ values decrease as n increases, but even for $n = 64$ they seem rather slight so we detect no finite-sample grounds for preferring one of these estimates over the other. Throughout Table 2 there are noticeable discrepancies relative to Table 1 (mostly the Table 2 values are smaller), with some departures from monotonicity in c and d, and little evidence of convergence over the range of n considered; indeed, in some cases the ratios drift in the wrong direction. However, nowhere is the asymptotic theory grossly misleading, and generally it seems to have performed fairly well.

Table 3 contains ratios of $\Sigma_{f^{-1}}^{-1}$ to Monte Carlo mean-squared errors of $\hat{\beta}_{\hat{f}^{-1}}$ and $\tilde{\beta}_{\hat{f}^{-1}}$. They are predominantly less than 1, and for the smaller values of c and d substantially so. In some cases there are significant, even large, differences between the $\hat{\beta}_{\hat{f}^{-1}}$ and $\tilde{\beta}_{\hat{f}^{-1}}$ values, though often they are very

TABLE 2

Ratios of $n\Sigma_{\hat{f}^{-1}}^{-1}$ to Monte Carlo MSE's of $\hat{\beta}_\phi$ and $\tilde{\beta}_\phi$ with $\phi(\lambda) = |2\sin\lambda/2|$

	$\hat{\beta}_\phi$					$\tilde{\beta}_\phi$				
					$n = 64$					
c/d	0.05	0.15	0.25	0.35	0.45	0.05	0.15	0.25	0.35	0.45
0.05	0.829	0.889	0.918	0.964	0.963	0.821	0.881	0.918	0.959	0.962
0.15	0.768	0.875	0.821	0.945	1.042	0.760	0.869	0.809	0.938	1.008
0.25	0.709	0.756	0.922	0.905	0.920	0.700	0.747	0.916	0.886	0.901
0.35	0.584	0.744	0.890	0.898	1.018	0.567	0.730	0.870	0.870	0.979
0.45	0.395	0.640	0.777	0.930	0.930	0.385	0.629	0.765	0.916	0.897
					$n = 128$					
c/d	0.05	0.15	0.25	0.35	0.45	0.05	0.15	0.25	0.35	0.45
0.05	0.813	0.894	0.880	0.988	0.973	0.809	0.891	0.882	0.981	0.973
0.15	0.813	0.818	0.970	0.990	0.989	0.809	0.808	0.964	0.990	0.976
0.25	0.668	0.796	1.006	0.955	0.998	0.661	0.790	1.007	0.939	0.980
0.35	0.607	0.766	0.849	1.049	0.971	0.601	0.750	0.838	1.029	0.939
0.45	0.406	0.662	0.769	0.971	0.979	0.403	0.647	0.762	0.956	0.955
					$n = 256$					
c/d	0.05	0.15	0.25	0.35	0.45	0.05	0.15	0.25	0.35	0.45
0.05	0.837	0.894	0.915	0.937	0.995	0.833	0.895	0.910	0.934	0.994
0.15	0.758	0.909	0.965	0.950	0.984	0.755	0.906	0.959	0.945	0.969
0.25	0.699	0.833	0.898	0.941	1.003	0.697	0.834	0.892	0.936	0.997
0.35	0.611	0.758	0.927	1.018	0.975	0.607	0.749	0.922	0.998	0.955
0.45	0.409	0.620	0.821	0.980	1.027	0.406	0.615	0.806	0.969	1.001

slight. The $\tilde{\beta}_{\hat{f}^{-1}}$ demonstrate the greater degree of convergence to 1 with n. One expects that a major effect on the results is the estimation of d and that they could deteriorate if a richer model of $f(\lambda)$ were estimated. Overall, while there is clearly a degree of sensitivity to c and d, the estimates considered appear to cope adequately with long-range dependence in both x_t and u_t.

Acknowledgments. We are very grateful for the comments of two referees and an Associate Editor on the first version of this paper. One referee, in particular, showed how to considerably abbreviate our original proofs. Though the final version of the paper substantially relaxes the conditions in the original version, we have made use of this referee's techniques in doing so. We also thank Yoshihiko Nishiyama for carrying out the Monte Carlo simulations.

TABLE 3

Ratios of $n\Sigma_{f^{-1}}^{-1}$ *to Monte Carlo MSE's of* $\hat{\beta}_{\hat{f}^{-1}}$ *and* $\tilde{\beta}_{\hat{f}^{-1}}$

	$\hat{\beta}_{\hat{f}^{-1}}$					$\tilde{\beta}_{\hat{f}^{-1}}$				
					n = 64					
c / d	**0.05**	**0.15**	**0.25**	**0.35**	**0.45**	**0.05**	**0.15**	**0.25**	**0.35**	**0.45**
0.05	0.964	0.935	0.933	0.955	0.955	0.964	0.934	0.936	0.951	0.954
0.15	0.916	0.931	0.848	0.947	0.955	0.916	0.931	0.843	0.942	0.992
0.25	0.906	0.874	0.901	0.912	1.020	0.905	0.869	0.903	0.898	0.893
0.35	0.737	0.882	0.911	0.907	0.907	0.734	0.878	0.903	0.885	0.941
0.45	0.583	0.792	0.809	0.933	0.906	0.584	0.788	0.802	0.922	0.878
					n = 128					
c / d	**0.05**	**0.15**	**0.25**	**0.35**	**0.45**	**0.05**	**0.15**	**0.25**	**0.35**	**0.45**
0.05	0.981	0.980	0.926	1.003	0.962	0.981	0.981	0.928	0.998	0.962
0.15	0.997	0.908	1.004	0.984	0.978	0.997	0.905	1.004	0.986	0.966
0.25	0.891	0.899	1.028	0.956	1.001	0.891	0.898	1.033	0.942	0.984
0.35	0.846	0.951	0.936	1.048	0.960	0.847	0.945	0.925	1.036	0.931
0.45	0.614	0.905	0.854	0.962	0.960	0.616	0.907	0.852	0.954	0.939
					n = 256					
c / d	**0.05**	**0.15**	**0.25**	**0.35**	**0.45**	**0.05**	**0.15**	**0.25**	**0.35**	**0.45**
0.05	1.023	0.966	1.001	0.916	0.899	1.021	0.968	0.998	0.941	0.989
0.15	0.899	1.021	0.999	0.911	0.923	0.898	1.023	1.013	0.957	0.971
0.25	1.034	0.990	0.927	0.846	0.808	1.032	0.990	0.951	0.933	0.988
0.35	0.940	0.953	0.963	0.909	0.727	0.937	0.950	1.002	1.009	0.946
0.45	0.654	0.794	0.892	0.858	0.613	0.653	0.797	0.932	1.014	0.986

REFERENCES

ADENSTEDT, R. K. (1974). On large-sample estimation of the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

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

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

DAHLHAUS, R. (1995). Efficient location and regression estimation for long range dependent regression models. *Ann. Statist.* **23** 1029–1047.

DAVIES, R. B. and HARTE, D. S. (1987). Tests for Hurst effect. *Biometrika* **74** 95–101.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

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.

GALLANT, A. R. and GOEBEL, J. J. (1976). Nonlinear regression with autocorrelated errors. *J. Amer. Statist. Assoc.* **71** 961–967.

GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (1994). Asymptotic normality of regression estimators with long memory errors. *Statist. Probab. Lett.* To appear.

GIRAITIS, L. and SURGAILIS, D. (1991). 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.

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

GRAY, H. L., ZHANG, N.-F. and WOODWARD, W. A. (1994). On generalized fractional processes—a correction. *J. Time Ser. Anal.* **15** 561–562.

HAMON, B. V. and HANNAN, E. J. (1963). Estimating relations between time series. *J. Geophys. Res.* **68** 1033–1041.

HANNAN, E. J. (1967). The estimation of a lagged regression relation. *Biometrika* **54** 409–418.

HANNAN, E. J. (1971). Non-linear time series regression. *J. Appl. Probab.* **8** 767–780.

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

HOSOYA, Y. (1993). A limit theory in stationary processes with long-range dependence and related statistical models. Unpublished manuscript.

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

JENNRICH, R. I. (1969). Non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

KASHYAP, R. L. and EOM, K.-B. (1988). Estimation in long-memory time series model. *J. Time Ser. Anal.* **9** 35–41.

KOUL, H. L. (1992). *M*-estimators in linear models with long range dependent errors. *Statist. Probab. Lett.* **14** 153–164.

KOUL, H. L. and MUKHERJEE, K. (1993). Asymptotics of R-, MD- and LAD estimators in linear regression models with long range dependent errors. *Probab. Theory Related Fields* **95** 538–553.

KÜNSCH, H., BERAN, J. and HAMPEL, R. (1993). Contrasts under long-range correlations. *Ann. Statist.* **21** 943–964.

MALINVAUD, E. (1970). The consistency of nonlinear regression. *Ann. Math. Statist.* **41** 959–969.

ROBINSON, P. M. (1972). Non-linear regression for multiple time series. *J. Appl. Probab.* **9** 758–768. [To be reprinted in *Nonlinear Models* (H. Bierens and A. R. Gallant, eds.). International Library of Critical Writings in Econometrics, Edward Elgar.]

ROBINSON, P. M. (1994a). 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. (1994b). Efficient tests of nonstationary hypotheses. *J. Amer. Statist. Assoc.* **89** 1420–1437.

SCOTT, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Adv. in Appl. Probab.* **5** 119–137.

STOUT, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

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

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

YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19** 158–177.

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: robinso1@vax.lse.ac.uk