The Annals of Statistics
1999, Vol. 27, No. 1, 178–203

WHITTLE ESTIMATOR FOR FINITE-VARIANCE NON-GAUSSIAN TIME SERIES WITH LONG MEMORY

By Liudas Giraitis[1] and Murad S. Taqqu[2]

Boston University

We consider time series $Y_t = G(X_t)$ where X_t is Gaussian with long memory and G is a polynomial. The series Y_t may or may not have long memory. The spectral density $g_\theta(x)$ of Y_t is parameterized by a vector θ and we want to estimate its true value θ_0. We use a least-squares Whittle-type estimator $\widehat{\theta}_N$ for θ_0, based on observations $Y_1, \dots, Y_N$. If Y_t is Gaussian, then $\sqrt{N}(\widehat{\theta}_N - \theta_0)$ converges to a Gaussian distribution. We show that for non-Gaussian time series Y_t, this $\sqrt{N}$ consistency of the Whittle estimator does not always hold and that the limit is not necessarily Gaussian. This can happen even if Y_t has short memory.

1. Introduction. A time series X_t, $t \in \mathbb{Z}$ is said to be *strongly dependent* (possesses *long memory* or *long-range dependence*) if it has a spectral density $f(x)$, satisfying

$$f(x) = |x|^{-\alpha} L(1/|x|), \qquad x \in [-\pi, \pi], \ (0 < \alpha < 1), \tag{1.1}$$

where L is a slowly varying function at infinity. Since such time series are used as models in many applications, it is important to be able to estimate the long-memory parameter α. It is well known that the strong dependence renders many results in statistical inference invalid, for example, for confidence intervals [see Beran (1992)], or U-statistics [see Dehling and Taqqu (1989)] or for testing the change-points of the distribution function [see Giraitis, Leipus and Surgailis (1996)]. Fox and Taqqu (1986) have discovered the surprising fact that when X_t is Gaussian, the Whittle estimator of the long-memory parameter continues to satisfy the central limit theorem (CLT) and is $\sqrt{N}$-weakly consistent; that is, it has the same type of asymptotic properties as under short memory ($\alpha = 0$). This is because the Whittle estimator compensates for the underlying strong dependence. Giraitis and Surgailis (1990) showed that a similar result holds for linear sequences. Dahlhaus (1989), extending the result in the Gaussian case to the maximum likelihood, proved that the maximum likelihood estimator is efficient and asymptotically normal.

In the semiparametric setup, when the knowledge about the behavior of the spectral density is localized at frequency $x = 0$, Robinson (1994, 1995) developed methods for estimating the memory parameter, based on local Whittle

Received October 1996; revised October 1998.

[1]Supported by NSF Grant NCR-94-04931 and ESRC Grant R000235892 at the London School of Economics.

[2]Supported in part by NSF Grants DMS-94-04093 and NCR-94-04931 at Boston University.

AMS 1991 *subject classifications*. Primary 62E20, 62F10; secondary 60G18.

Key words and phrases. Hermite polynomials, non-central limit theorem, long-range dependence, quadratic forms, time series.

and modified Geweke–Porter–Hudak estimators [see also Giraitis and Koul (1997)]. These estimators are consistent and satisfy the CLT, but the rates of convergence are slower than in the parametric cases. A comparative study of the effectiveness of the various methods for estimating the long-memory parameter was considered by Taqqu, Teverovsky and Willinger (1995) and Taqqu and Teverovsky (1998). The least-squares based Whittle method, it turns out, is very effective for Gaussian or linear time series when the parametric model is accurately specified.

We show in this paper that the compensation effect in the Whittle estimator which appears when the observations X_t are pure Gaussian or linear is rather the exception than the rule. The results below imply that, in general, for non-Gaussian time series Y_t, the $\sqrt{N}$ consistency of the Whittle estimator does not always hold and the limit, moreover, is not necessarily Gaussian. Moreover, as shown in Section 5, it is possible that Y_t has short memory and that $\widehat{\theta}_N$ converges, nevertheless, to a non-Gaussian distribution.

We suppose that (X_t) is a mean-zero Gaussian stationary time series with long memory; that is, with spectral density (1.1). Its covariance is $r(t) = EX_0X_t = \int_{-\pi}^{\pi} e^{itx} f(x)\,dx$. We shall refer to the exponent α in (1.1) as the long-memory parameter of (X_t). Suppose that the time series

$$Y_t = G(X_t), \qquad t = 1, \dots, N, \tag{1.2}$$

is observed, where G is a polynomial and (Y_t) has zero mean. We suppose that $Y_t,\ t \in \mathbb{Z}$ may display long memory, that is, it has a spectral density $s_\theta(x) = \sigma^2 g_\theta(x)$, $|x| \le \pi$, $\theta \in \Theta$, $\sigma > 0$, where $\Theta \subset \mathbb{R}^p$ is a compact set and g_θ satisfies

$$g_\theta(x) = |x|^{-\alpha_G(\theta)}\, L_{G,\theta}(1/|x|), \qquad |x| \le \pi, \tag{1.3}$$

where $0 \le \alpha_G(\theta) < 1$ and $L_{G,\theta}$ is a slowly varying function. The sequence (Y_t) is said to have short memory if $\alpha_G(\theta) = 0$, in which case, g_θ is bounded if $L_{G,\theta}$ is bounded. (Y_t) is said to have long memory if $\alpha_G(\theta) > 0$.

The Whittle estimator $\widehat{\theta}_N$ of θ is a function of the observations $Y_t,\ t = 1, \dots, N$. Our goal is to characterize its asymptotic properties when the number N of observations goes to infinity. We want to understand why key features of the Whittle estimator for Gaussian or linear observations such as compensation of the long memory, $\sqrt{N}$ consistency and asymptotic normality may cease to hold when $Y_t = G(X_t)$ is nonlinear. We restrict ourselves to polynomial G because such a choice already illustrates the problems associated with nonlinear transformations of Gaussian data. The case of a general G, which will be considered in a different paper, involves, in addition, delicate questions of convergence.

Let θ_0 denote the true (unknown) value of θ. To obtain the asymptotic behavior of $\widehat{\theta}_N$, we use Lemma 6.2 below to approximate $\widehat{\theta}_N - \theta_0$ by

$$T_N = \frac{1}{N} \sum_{t,s=1}^{N} \nabla a_{\theta_0}(t-s) G(X_t) G(X_s), \tag{1.4}$$

where

$$a_\theta(t) = \int_{-\pi}^{\pi} e^{itx} g_\theta^{-1}(x)\, dx \tag{1.5}$$

and where ∇a_{θ_0} denotes the derivative of a_θ with respect to θ evaluated at θ_0 [see (2.9)]. The kernel $a_\theta(t)$ involves the spectral density g_θ of the observations Y_t and is defined in the same way as in the case of the Gaussian–Whittle estimator. Because the product $G(X_t)G(X_s)$ in (1.4) involves the joint vector (X_t, X_s), we will expand it in bivariate Hermite polynomials $H_{m,n}(X_t, X_s)$ [see (2.14) for a definition]. One gets

$$T_N = \frac{1}{N} \sum_{m,n\geq 0} S_N^{(m,n)},$$

where

$$S_N^{(m,n)} = \sum_{t,s=1}^{N} v_{m,n}(t-s) H_{m,n}(X_t, X_s), \tag{1.6}$$

and where

$$v_{m,n}(t-s) = \frac{1}{m!n!}[EG^{(m)}(X_t)G^{(n)}(X_s)]\nabla a_{\theta_0}(t-s), \tag{1.7}$$

$G^{(0)} \equiv G$ and $G^{(m)}(x) = (d^m/dx^m)G(x)$. To determine the exponent γ for which $N^\gamma(\widehat{\theta}_N - \theta_0)$ converges to a limit, we will show that for each (m, n), there is an exponent $\kappa(m, n)$ such that $N^{-\kappa(m,n)} S_N^{(m,n)}$ converges in distribution. The value of the exponent $\kappa(m, n)$ decreases to 1/2 as $m+n$ increases. The asymptotic behavior of $S_N^{(m,n)}$ is controlled both by the dependence structure of the bivariate Hermite polynomials $H_{m,n}(X_t, X_s)$ and the weights $v_{m,n}(t)$ in (1.6). This explains, in particular, the compensation that occurs in the linear case $G(x) = x$. For such a G, $T_N = N^{-1} S_N^{(1,1)} = N^{-1} \sum_{t,s=1}^{N} v_{1,1}(t - s)H_{1.1}(X_t, X_s)$. Because $H_{1,1}(X_t, X_s) = X_t X_s$ and because of the special form of the weights $v_{1,1}(t)$ (see Example 4.1), there is compensation of the long memory: the sum $S_N^{(1,1)}$ converges to a Gaussian distribution with normalizing factor $N^{-1/2}$, and hence, the limiting distribution of $\sqrt{N}(\widehat{\theta}_N - \theta_0)$ is Gaussian. In the case of non-Gaussian observations $G(X_t)$, such a compensation is the exception rather than the rule and the class of limit distribution is then much richer. The limit for $\widehat{\theta}_N - \theta_0$ may be Gaussian but with a normalization different from $\sqrt{N}$ (Section 2), but it can also be non-Gaussian; for example, it may have the (non-Gaussian) Rosenblatt distribution (Section 3).

The paper is structured as follows. The main results are stated in Section 2. We provide a more detailed analysis of the asymptotic behavior of the Whittle estimator in Sections 3 and 4. Section 5 treats the special case of Hermite filters. The results stated in Sections 2 and 3 are then proved in Sections 6 and 7, respectively.

2. Main results. We want to estimate the parameters (θ, σ) that characterize the spectral density $s_\theta(x) = \sigma^2 g_\theta(x)$ of the process (Y_t), using observations $Y_1, \dots, Y_N$. We shall estimate the true value θ_0 of θ, assuming, as usual, that θ_0 lies in the interior of the compact set Θ. We also assume that if $\theta \neq \theta_0$, then the set $\{x: g_\theta(x) \neq g_{\theta_0}(x)\}$ has positive Lebesgue measure, so that θ corresponds to a dependence structure different from the one associated with θ_0.

We use the standard (least-square) Whittle estimator $\widehat{\theta}_N$ of θ_0 [see Fox and Taqqu (1986)], defined as follows: $\widehat{\theta}_N$ is the value of θ that minimizes

$$\sigma_N^2(\theta) = N^{-1} Y' \, A_{N,\theta} Y, \tag{2.1}$$

where Y denotes the column vector $(Y_1, \dots, Y_N)$, Y' its transpose and where the entries of the Toeplitz matrix $A_{N,\theta} = \{a_\theta(t-s)\}_{t,s=1,\dots,N}$, are given in (1.5). We then estimate the true value σ_0^2 of σ^2 by

$$\widehat{\sigma}_N^2 = (2\pi)^{-2} \sigma_N^2(\widehat{\theta}_N). \tag{2.2}$$

[Fox and Taqqu (1986) included the factor $(2\pi)^{-2}$ in (1.5) instead of (2.2).]

To allow prediction, we suppose $\int_{-\pi}^{\pi} \log(\sigma^2 g_\theta(x))\, dx > -\infty$ and, as in Hannan (1973) and Fox and Taqqu (1986), we suppose without loss of generality that g_θ is suitably normalized so that

$$\int_{-\pi}^{\pi} \log g_\theta(x)\, dx = 0, \qquad \theta \in \Theta. \tag{2.3}$$

(For more details about this normalization, see the beginning of Section 4.)

The following theorem shows that, as in the Gaussian case considered in Fox and Taqqu (1986), $\widehat{\theta}_N$ and $\widehat{\sigma}_N$ are strongly consistent.

THEOREM 2.1. *Assume that* (2.3) *holds and that* $g_\theta^{-1}(x)$ *is a continuous function. Then*

$$\lim_{N\to\infty} \widehat{\theta}_N = \theta_0 \quad \textit{and} \quad \lim_{N\to\infty} \widehat{\sigma}_N^2 = \sigma_0^2 \quad \textit{a.s.} \tag{2.4}$$

We now focus on the study of the asymptotic behavior of $\widehat{\theta}_N - \theta_0$. In the next theorems, we show that contrary to the linear case $G(X_j) = X_j$, the convergence $N^\gamma(\widehat{\theta}_N - \theta_0)$ as $N \to \infty$ to a nondegenerate limit requires, typically, $\gamma < 1/2$. The limit, moreover, may be either Gaussian or non-Gaussian.

2.1. *Gaussian limit.* We shall need a number of technical conditions similar to those in Fox and Taqqu (1986). They are widely used in the statistical literature to control the behavior of the spectral density $s_\theta(x)$ around the pole $x = 0$.

STANDARD ASSUMPTIONS. In addition to (1.1), (1.3) and (2.3), we assume that $(\partial^2/\partial\theta_i\partial\theta_j)g_\theta^{-1}(x)$ is a continuous function in (x, θ),

$$\left|\frac{\partial}{\partial\theta_j}g_\theta^{-1}(x)\right| \le C|x|^{\alpha_G(\theta)-\varepsilon}, \qquad |x| \le \pi \quad \text{for } \theta = \theta_0 \tag{2.5}$$

and

$$\left|\frac{\partial^2}{\partial x\partial\theta_j}g_\theta^{-1}(x)\right| \le C|x|^{\alpha_G(\theta)-1-\varepsilon}, \qquad |x| \le \pi \quad \text{for } \theta = \theta_0, \tag{2.6}$$

where $\varepsilon > 0$ is any small fixed number. We also assume that the spectral density f of the Gaussian sequence (X_t) satisfies

$$\left|\frac{d}{dx}f(x)\right| \le C|x|^{-\alpha-1-\varepsilon}, \qquad |x| \le \pi, \tag{2.7}$$

where $\varepsilon = \varepsilon(\theta) > 0$ is any small fixed number.

We start with some notation. Define the column vector

$$\rho_1 = 2\sum_{t\in\mathbb{Z}}[E\dot{G}(X_t)G(X_0)]\nabla a_{\theta_0}(t), \tag{2.8}$$

where $\dot{G}$ denotes the derivative of G,

$$\nabla a_\theta(t) = \left(\frac{\partial}{\partial\theta_1}a_\theta(t), \dots, \frac{\partial}{\partial\theta_p}a_\theta(t)\right)', \qquad t \in \mathbb{Z}, \tag{2.9}$$

and $\nabla a_{\theta_0}(t) = \nabla a_\theta(t)|_{\theta=\theta_0}$.

Since by (2.6), $(\partial/\partial x)(\partial/\partial\theta_i)g_\theta^{-1} \in L^\gamma[-\pi, \pi]$, $i = 1, \dots, p$ for some $\gamma > 1$, then by a well-known property of Fourier coefficients [see Zygmund (1979), Theorem VI.3.8, Vol. I],

$$\sum_{t\in\mathbb{Z}}|\nabla a_{\theta_0}(t)| < \infty. \tag{2.10}$$

Since $|E\dot{G}(X_t)G(X_0)| \le (E(\dot{G}(X_0))^2(EG(X_0))^2)^{1/2}$, this implies that $|\rho_1| < \infty$.

Now introduce the $k\times k$ variance-covariance matrix $W_\theta = (w_\theta(i, j))_{i, j=1,\dots,k}$ with entries

$$w_\theta(i, j) = \int_{-\pi}^{\pi} g_\theta(x)\frac{\partial^2}{\partial\theta_i\partial\theta_j}g_\theta^{-1}(x)\,dx. \tag{2.11}$$

THEOREM 2.2. *Suppose that the standard assumptions hold, that $W_{\theta_0}^{-1}$ exists and $\rho_1 \ne 0$. Then*

$$\widehat{\theta}_N - \theta_0 = -(2\pi\sigma_0^2)^{-1}W_{\theta_0}^{-1}\rho_1\left(N^{-1}\sum_{j=1}^{N}X_j\right)(1 + o_P(1)). \tag{2.12}$$

This theorem is remarkable, in that it indicates that, under strong dependence, $\widehat{\theta}_N - \theta_0$ behaves asymptotically like the sample mean of the underlying Gaussian vector X_t when $\rho_1 \neq 0$. Since $\{X_t\}$ has strong dependence, the normalization will not $\sqrt{N}$ as the following corollary indicates.

COROLLARY 2.1. *Theorem* 2.2 *implies that*

$$[N^{1-\alpha}L^{-1}(N)]^{1/2}(\widehat{\theta}_N - \theta_0) \Rightarrow (2\pi\sigma_0^2)^{-1}W_{\theta_0}^{-1}\rho_1\xi, \tag{2.13}$$

where α is the long-memory parameter of the Gaussian sequence X_t appearing in (1.1), *and where ξ is a Gaussian random variable with zero mean and variance $E\xi^2 = 2/(\alpha(\alpha+1))$.*

Corollary 2.1 follows immediately from Theorem 2.2 because

$$\mathrm{Var}\left(\sum_{j=0}^{N} X_j\right) = \int_{-\pi}^{\pi}\left|\frac{e^{i(N+1)x}-1}{e^{ix}-1}\right|^2 f(x)\,dx \sim N^{1+\alpha}L(N)\int_{-\infty}^{\infty}\left|\frac{e^{ix}-1}{ix}\right|^2|x|^{-\alpha}\,dx$$

as $N \to \infty$ and

$$\int_{-\infty}^{\infty}\left|\frac{e^{ix}-1}{ix}\right|^2|x|^{-\alpha}\,dx = \int_0^1\int_0^1 |x-y|^{-1+\alpha}\,dx\,dy = 2/(\alpha(\alpha+1)).$$

REMARK. Fox and Taqqu (1986) have shown that when $G(X_t) = X_t$, one has compensation and the rate of convergence is $\sqrt{N}$. In this case, $\dot{G}(X_t) = 1$, $E\dot{G}(X_t)G(X_t) = EX_t = 0$ and $\rho_1 = 0$ by (2.8). Hence Corollary 2.1 does not apply.

2.2. *Asymptotic expansion for $\widehat{\theta}_N - \theta_0$.* Theorem 2.2 is a consequence of the general asymptotic expansion for $\widehat{\theta}_N$ which is given in Theorem 2.3 below. To characterize this expansion, we now define the bivariate Hermite polynomials $H_{n_1,n_2}(x_1, x_2)$, $n_1, n_2 = 0, 1, \ldots$, which are particular cases of multivariate Appell polynomials [see Avram and Taqqu (1987), Giraitis and Surgailis (1986), Giraitis and Taqqu (1998)].

They are defined by the recurrence relations

$$\begin{aligned} &\frac{\partial}{\partial x_1}H_{n_1,n_2}(x_1,x_2) = n_1 H_{n_1-1,n_2}(x_1,x_2);\\ &\frac{\partial}{\partial x_2}H_{n_1,n_2}(x_1,x_2) = n_2 H_{n_1,n_2-1}(x_1,x_2);\\ &EH_{n_1,n_2}(X_{t_1}, X_{t_2}) = 0. \end{aligned} \tag{2.14}$$

The first two relations indicate that these polynomials behave like power functions. The last relation provides the constants of integration and relates the polynomial to the joint distribution of the X_t's. Finally, the bivariate Hermite polynomials are orthogonal; that is, for any t, s, u, v,

$$EH_{m,n}(X_t, X_s)H_{m',n'}(X_u, X_v) = 0 \quad \text{if } m+n \neq m'+n'. \tag{2.15}$$

We now focus on the random variables $S_N^{(m,n)}$ which were defined in (1.6). Observe first that since G is a polynomial, $S_N^{(m,n)}$ is zero for large enough m and n. Moreover, (2.15) implies

$$ES_N^{(m,n)}S_N^{(m',n')} = 0 \quad \text{if } m+n \neq m'+n'.$$

We want to normalize each of the $S_N^{(m,n)}$ suitably. We shall assume at first that $1/(1-\alpha)$ is not an integer, that is, $\alpha \neq 1/2,\ 2/3,\ 3/4\ \ldots$ and let

$$k^* = [1/(1-\alpha)]$$

denote the smallest integer less than $1/(1-\alpha)$.

When $m = n = 0$, define the nonrandom term

$$\mu_N := N^{-1/2}S_N^{(0,0)} = N^{-1/2}\sum_{t,s=1}^{N} \nabla a_{\theta_0}(t-s)\ EG(X_t)G(X_s). \tag{2.16}$$

When $1 \le m+n \le k^*$, set

$$T_{k,N} = \sum_{m,n\ge 0:\, m+n=k} [N^{2-k(1-\alpha)}]L^k(N)]^{-1/2}S_N^{(m,n)}, \qquad 1 \le k \le k^*; \tag{2.17}$$

and, gathering the finitely many remaining $S_N^{(m,n)}$, set

$$V_N = N^{-1/2}\sum_{m,n\ge 0:\, m+n>k^*} S_N^{(m,n)}. \tag{2.18}$$

Finally, using the definition of $v_{m,n}(t)$ in (1.7), set

$$\begin{aligned}\rho_k &= \sum_{m,n\ge 0:\, m+n=k}\ \sum_{t\in\mathbb{Z}} v_{m,n}(t)\\ &= \sum_{m,n\ge 0:\, m+n=k} \frac{1}{m!n!}\sum_{t\in\mathbb{Z}}[EG^{(m)}(X_t)G^{(n)}(X_0)]\nabla a_{\theta_0}(t),\end{aligned} \tag{2.19}$$

and note that

$$\rho_k = \frac{1}{k!}\sum_t E\left[\frac{d^k}{d\mu^k}G(\mu+X_t)G(\mu+X_0)\right]\bigg|_{\mu=0}\nabla a_{\theta_0}(t).$$

The sum (2.19) converges absolutely, since $\sum_t |\nabla a_{\theta_0}(t)| < \infty$ [see (2.10)].

We provide in the following theorem an asymptotic expansion for the Whittle estimator. Similar types of expansions for M-estimators of the location parameter are discussed in the recent paper by Koul and Surgailis (1997) and, for empirical distribution functions, in Ho and Hsing (1996) [see also Ho and Hsing (1997)].

THEOREM 2.3. *Suppose that the standard assumptions hold, that* $W_{\theta_0}^{-1}$ *exists and that* $1/(1-\alpha)$ *is not an integer. Then*

$$\hat{\theta}_N = \theta_0 - (1 + o_P(1))(2\pi\sigma_0^2)^{-1} W_{\theta_0}^{-1} \times \left[\sum_{1 \le k \le k^*} N^{-k(1-\alpha)/2} L^{k/2}(N) T_{k,N} + N^{-1/2} V_N + N^{-1/2} \mu_N \right] + o_P(N^{-1}), \tag{2.20}$$

where μ_N, $T_{1,N}, \dots, T_{k^*,N}, V_N$ *are defined in* (2.16), (2.17), (2.18), *respectively. The vectors* $T_{1,N}, \dots, T_{k^*,N}$ *and* V_N *are uncorrelated. Moreover, as* $N \to \infty$,

$$(T_{1,N}, \dots, T_{k^*,N}) \Rightarrow (\rho_1 I_1, \dots, \rho_{k^*} I_{k^*}), \tag{2.21}$$

where I_k *is the* k*-tuple Itô–Wiener integral* (2.23), *while*

$$V_N \Rightarrow \mathcal{N}(0, D_{k^*}) \tag{2.22}$$

is asymptotically normally distributed with zero mean and covariance matrix D_{k^*}, *where the entries of the* $k \times k$ *matrix* D_{k^*} *are given below in* (6.22) *or* (6.23). *In addition,*

$$\mu_N \to 0.$$

The k-tuple Itô–Wiener integral I_k is defined by

$$I_k = \int_{R^k}'' \frac{\exp(i(x_1 + \cdots + x_k)) - 1}{i(x_1 + \cdots + x_k)} |x_1|^{-\alpha/2} \cdots |x_k|^{-\alpha/2} Z(dx_1) \cdots Z(dx_k), \quad k = 1, \dots, k^*, \tag{2.23}$$

where $Z(dx) = \overline{Z(-dx)}$ is a standard Gaussian complex measure with zero mean and variance $E|Z(dx)|^2 = dx$. The symbol $\int''$ indicates that one does not integrate on the hyperdiagonals $x_i = \pm x_j$, $i, j = 1, \dots, k$. The integral is well defined if

$$\int_{\mathbb{R}^k} \left| \frac{\exp(i(x_1 + \cdots + x_k)) - 1}{i(x_1 + \cdots + x_k)} \right|^2 |x_1|^{-\alpha} \cdots |x_k|^{-\alpha} dx_1 \cdots dx_k < \infty,$$

and this relation holds for $k = 1, \dots, k^*$, as long as $k^* < 1/(1-\alpha)$.

Applications of Theorem 2.3 will be found in the next sections. Theorem 2.3 provides in particular the limits of $T_{1,N}, \dots, T_{k*,N}$ and V_N, which are properly normalized sums of bivariate Hermite polynomials. The theorem also indicates how $\hat{\theta}_N - \theta_0$ relates to these quantities.

The univariate Hermite polynomials $H_n(x)$ are defined by the relations

$$H_0(x) = 1, \qquad \frac{d}{dx} H_n(x) = n H_{n-1}(x), \qquad E H_n(X_t) = 0, \ n \ge 1.$$

If instead of expanding $G(X_t)G(X_s)$ in the bivariate Hermite polynomials $H_{m,n}(X_t, X_s)$, we had expanded separately each $G(X_t)$ in univariate Hermite polynomials, we would not have easily obtained the correct normalization factors. In fact, the first nonzero term in the bivariate expansion is not determined by the first nonzero term in the univariate expansion.

The following theorem concerns the boundary case $\alpha = 1 - 1/k^*$.

THEOREM 2.4. *If* $1/(1-\alpha)$ *is an integer in Theorem* 2.3, *that is*, $\alpha = 1-1/k^*$, *then as* $N \to \infty$,

$$(T_{1,N}, \dots, T_{k^*-1,N}) \Rightarrow (\rho_1 I_1, \dots, \rho_{k^*-1} I_{k^*-1}) \tag{2.24}$$

and

$$\operatorname{Var} T_{k^*,N} = O(N^{1+\varepsilon}) \tag{2.25}$$

for any $\varepsilon > 0$.

The results of this section are proved in Section 6.

3. Convergence to the Rosenblatt distribution. We now analyze the asymptotic expansion (2.20) in more detail. We show that, if the observed process is $Y_t = G(X_t)$, then the deviation of the Whittle estimate $\widehat{\theta}_N$ from the true value of parameter θ_0, after suitable rescaling can have, asymptotically, either a Gaussian or non-Gaussian distribution, in particular, the Rosenblatt distribution. This is the distribution of

$$I_2 = \int_{\mathbb{R}^2}'' \frac{\exp(it(x_1+x_2)) - 1}{i(x_1+x_2)} |x_1|^{-\alpha}|x_2|^{-\alpha} Z(dx_1)Z(dx_2), \qquad \alpha > 1/2, \tag{3.1}$$

and arises when the dominant term in (2.20) is $k = 2$. The expansion (2.20) indicates, that $\widehat{\theta}_N$ could have, in principle, limit distributions represented by multiple Wiener–Itô integrals of third or higher order.

Theorem 2.2 implies that $\rho_1 \neq 0$ is a sufficient condition for the limit to be Gaussian. The next result indicates when the limit I_2 can appear.

THEOREM 3.1. *Let* $\rho_1 = 0$, $\rho_2 \neq 0$.
If $1/2 < \alpha < 1$, *then*

$$N^{(1-\alpha)} L^{-1}(N)(\widehat{\theta}_N - \theta_0) \Rightarrow (2\pi\sigma_0^2)^{-1} W_{\theta_0}^{-1} \rho_2\, I_2, \qquad N \to \infty, \tag{3.2}$$

where I_2 *has the Rosenblatt distribution.*
If $0 < \alpha < 1/2$, *then*

$$\sqrt{N}(\widehat{\theta}_N - \theta_0) \Rightarrow \mathcal{N}(0, (2\pi\sigma_0^2)^{-2} W_{\theta_0}^{-1} D W_{\theta_0}^{-1}), \tag{3.3}$$

where D is $p \times p$ matrix with entries

$$d(i, j) = \sum_{t \in \mathbb{Z}} \left[\sum_{s_1, s_2 \in \mathbb{Z}} \dot{a}_{\theta_0}^{(i)}(s_1) \dot{a}_{\theta_0}^{(j)}(s_2) \operatorname{Cov}(G(X_t)G(X_{t+s_1}), G(X_0)G(X_{s_2})) \right]. \tag{3.4}$$

Theorem 3.1 is proved in Section 7.

REMARK 1. The convergence in (3.4) has to be understood as

$$\lim_{T \to \infty} \sum_{|t| \le T} \sum_{s_1, s_2} \cdot$$

The order of summation is important because $\sum_{t, s_1, s_2} |\cdot| = \infty$.

REMARK 2. As indicated by the theorem, in the case (3.2), the limit is not Gaussian. As noted in Section 5 below, this can happen even when the observations $Y_t = G(X_t)$ are weakly dependent, for example, if $Y_t = H_k(X_t)$ with $k > 1/(1 - \alpha)$.

REMARK 3. Relation $1/2 < \alpha < 1$ implies $k^* \ge 2$ and $0 < \alpha < 1/2$ implies $k^* = 1$. In the latter case, the term that determines the limit is V_N [see (2.18)].

REMARK 4. To understand the essence of the difficulty in the proof of Theorem 3.1, note that the convergence (2.21) in the case $1/2 < \alpha < 1$ implies $T_{1, N} \Rightarrow 0$ and $T_{2, N} \Rightarrow \rho_2 I_2$ since $\rho_1 = 0$, $\rho_2 \ne 0$. If the expansion (2.20) were to be applied in a simple-minded fashion, one would get

$$\begin{aligned} &(2\pi\sigma_0^2 W_{\theta_0}) N^{(1-\alpha)} L^{-1}(N)(\widehat{\theta}_N - \theta_0) \\ &\quad \sim N^{(1-\alpha)/2} L^{-1/2}(N) T_{1, N} + T_{2, N} \sim \infty \cdot 0 + \rho_2 I_2. \end{aligned}$$

In the proof, we show that this "$\infty \cdot 0$" is in fact 0.

4. An alternative expression for ρ_k. Theorem 3.1 and Corollary 2.1 show the important role that the ρ_k's play in determining the limit distribution of $\widehat{\theta}_N - \theta_0$. The expression for ρ_k given in (2.19) is, however, difficult to work with. To derive an alternative one, we first need to make explicit the functional relationship that results from the normalization (2.3).

The spectral density $s_\theta(x)$ of the observations (Y_t) has the two following equivalent expressions:

$$s_\theta(x) = \sigma^2 g_\theta(x) = v^2 h_\theta(x), \tag{4.1}$$

where $v^2 = \operatorname{Var} Y_t$. Because of the normalization (2.3), the factors σ^2 and $g_\theta(x)$ are related to v^2 and $h_\theta(x)$ as follows:

$$\begin{aligned} &\sigma^2 = \sigma^2(v^2, \theta) = \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[v^2 h_\theta(x)]\, dx \right\} \quad \text{and} \\ &g_\theta(x) = v^2 h_\theta(x) / \sigma^2(v^2, \theta), \end{aligned} \tag{4.2}$$

and, in particular,

$$\sigma^2(v^2, \theta) = v^2\sigma^2(1, \theta) \quad \text{and} \quad g_\theta(x) = h_\theta(x)/\sigma^2(1, \theta). \tag{4.3}$$

The parameter $\sigma^2(v^2, \theta)$ equals the one-step prediction variance. If the variance v^2 is unknown, σ^2 will be unknown even if θ is known. This is why both θ and σ have to be estimated and can be regarded as independent parameters.

Introduce also the coefficients

$$J(\ell) = EG(X_0)H_\ell(X_0), \qquad \ell \geq 0 \tag{4.4}$$

$(J(0) = EG(X_t) = 0)$ in the expansion

$$G(X_t) = \sum_{\ell\geq 1} \frac{J(\ell)}{\ell!} H_\ell(X_t) \tag{4.5}$$

of G in univariate Hermite polynomials [see Taqqu (1975)].

We can now derive the following alternative expression for ρ_k.

LEMMA 4.1. *For all* $k \geq 1$,

$$\rho_k = \sigma_0^2\left[\int_{-\pi}^{\pi} \frac{\lambda_k(x)}{s_{\theta_0}(x)}\,dx \int_{-\pi}^{\pi} \frac{\nabla s_{\theta_0}(x)}{s_{\theta_0}(x)}\,dx - 2\pi\int_{-\pi}^{\pi} \frac{\lambda_k(x)}{s_{\theta_0}(x)}\frac{\nabla s_{\theta_0}(x)}{s_{\theta_0}(x)}\,dx\right], \tag{4.6}$$

where s_{θ_0} *is given in* (7.3),

$$\lambda_k(x) = \sum_{\substack{m,\,n\geq 0\\ m+n=k}} \frac{1}{m!n!} h_{m,n}(x), \tag{4.7}$$

$$h_{m,n}(x) = \sum_{\ell\geq 1} \frac{1}{\ell!} J(\ell+m)J(\ell+n)f^{(*\ell)}(x), \tag{4.8}$$

and the $J(\ell)$, $\ell \geq 1$, *are the coefficients* (4.4) *in the expansion of* G *in Hermite polynomials*.

PROOF. Since

$$G^{(m)}(x) = \sum_{\ell\geq m} \frac{J(\ell)}{(\ell-m)!}\ H_{\ell-m}(x),$$

we have

$$\begin{aligned} EG^{(m)}(X_t)G^{(n)}(X_0) &= \sum_{\ell,\,\ell'\geq 0} \frac{J(\ell+m)}{\ell!}\frac{J(\ell+n)}{\ell'!} EH_\ell(X_t)H_{\ell'}(X_0)\\ &= J(m)J(n) + \sum_{\ell\geq 1} \frac{J(\ell+m)J(\ell+n)}{\ell!} r^\ell(t), \end{aligned}$$

where $r(t) = EX_tX_0 = \int_{-\pi}^{\pi} e^{itx} f(x)\,dx$, so that by (4.8), we have

$$EG^{(m)}(X_t)G^{(n)}(X_0) = J(m)J(n) + \int_{-\pi}^{\pi} e^{itx} h_{m,n}(x)\,dx.$$

We now incorporate this relation in the expression (2.19) for ρ_k. Since

$$\sum_t \nabla a_{\theta_0}(t) = 2\pi \nabla g_{\theta_0}^{-1}(0) = 0$$

by (2.5), the constant term $J(m)J(n)$ contributes nothing to ρ_k. Applying the Parseval identity and the relations (4.7) and (1.5), we obtain

$$\rho_k = 2\pi \int_{-\pi}^{\pi} \lambda_k(x) \nabla g_{\theta_0}^{-1}(x)\, dx. \tag{4.9}$$

By (4.3), $g_\theta(x) = h_\theta(x)/\sigma^2(1, \theta)$, and hence

$$\nabla g_{\theta_0}^{-1} = -g_{\theta_0}^{-2} \nabla g_{\theta_0} = -\sigma^4(1, \theta_0) h_{\theta_0}^{-2} [(\nabla \sigma^{-2}(1, \theta)) h_\theta + \sigma^{-2}(1, \theta) \nabla h_\theta]\big|_{\theta = \theta_0}.$$

Since (4.2) implies

$$\begin{aligned} \nabla \sigma^{-2}(1, \theta_0) = \nabla(\sigma^2(1, \theta_0))^{-1} &= -\frac{1}{\sigma^4(1, \theta_0)} \nabla \sigma^2(1, \theta_0) \\ &= -\sigma^{-2}(1, \theta_0) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\nabla h_{\theta_0}(u)}{h_{\theta_0}(u)}\, du, \end{aligned} \tag{4.10}$$

we get

$$\nabla g_{\theta_0}^{-1}(x) = \sigma^2(1, \theta_0) \left[\frac{1}{2\pi} \frac{1}{h_{\theta_0}(x)} \int_{-\pi}^{\pi} \frac{\nabla h_{\theta_0}(u)}{h_{\theta_0}(u)}\, du - \frac{1}{h_{\theta_0}(x)} \frac{\nabla h_{\theta_0}(x)}{h_{\theta_0}(x)} \right].$$

Using (4.1) and (4.3), we can now replace $h_{\theta_0}(x)$ by $s_{\theta_0}(x)/v^2$, $\sigma^2(1, \theta_0)$ by σ_0^2/v^2 and then use (4.9) to get (4.6). □

REMARK 1. The parameter σ_0^2 in ρ_k is estimated by $\widehat{\sigma}_N^2$ [see (2.2)].

REMARK 2. The integrands in (4.6), in particular $\nabla s_{\theta_0} = \nabla s_\theta|_{\theta=\theta_0}$, are non-necessarily strictly positive or negative functions.

Particular cases.

$$\lambda_1(x) = h_{1,0}(x) + h_{0,1}(x) = 2 \sum_{\ell \ge 1} J(\ell) J(\ell+1) f^{(*\ell)}(x),$$

$$\begin{aligned} \lambda_2(x) &= h_{2,0}(x)/2 + h_{0,2}(x)/2 + h_{11}(x) \\ &= \sum_{\ell \ge 1} \frac{1}{\ell!} [J(\ell+2) J(\ell) + J^2(\ell+1)] f^{(*\ell)}(x). \end{aligned}$$

EXAMPLE 4.1. If $G(X_t) = X_t$, then $J(1) = 1$ and $J(\ell) = 0$ for $\ell \neq 1$. Hence $\rho_k = 0$ for all $k \ge 1$. In this case, as was already proved in Fox and Taqqu (1986), we have "compensation," and $\widehat{\theta}_N - \theta_0$ converges to a Gaussian limit after a $\sqrt{N}$ normalization.

EXAMPLE 4.2. If $G(X_t) = H_\ell(X_t)$, $\ell \ge 1$, then $\rho_1 = \rho_3 = \rho_5 = \cdots = 0$.

EXAMPLE 4.3. One gets $\rho_1 = 0$ if G is such that $J(\ell)J(\ell+1) = 0$ for all $\ell \geq 1$, for example if $G(X_t) = H_1(X_t) + H_3(X_t) = X_t^3 - 2X_t$.

5. Hermite filters. Suppose $Y_t = H_\ell(X_t)$ with $\ell \geq 2$ and that $\{X_t,\ t \geq 0\}$ has long memory ($1/2 < \alpha < 1$). If $\ell < 1/(1-\alpha)$, then the spectral density of Y_t diverges at the origin. In this case, Y_t has also long memory and satisfies a noncentral limit theorem [Taqqu (1979) and Dobrushin and Major (1979)]. On the other hand, if $\ell > 1/(1-\alpha)$, then the spectral density of Y_t is continuous. In this second case, Y_t has short memory and satisfies a central limit theorem [Breuer and Major (1983), Giraitis and Surgailis (1985)]. What happens to the Whittle estimator of θ in either of these two cases? The following corollary provides the answer.

COROLLARY 5.1. *If* $Y_t = H_\ell(X_t),\ \ell \geq 2,\ 1/2 < \alpha < 1$, *then the convergence* (3.2) *holds with*

$$\rho_2 = \sigma_0^2 \ell \biggl[\int_{-\pi}^{\pi} \frac{f^{(*(\ell-1))}(x)}{f^{(*\ell)}(x)}\, dx \int_{-\pi}^{\pi} \frac{\nabla f^{(*\ell)}(x)}{f^{(*\ell)}(x)}\, dx - 2\pi \int_{-\pi}^{\pi} \frac{f^{(*(\ell-1))}(x)}{f^{(*\ell)}(x)} \frac{\nabla f^{(*\ell)}(x)}{f^{(*\ell)}(x)}\, dx \biggr], \tag{5.1}$$

where σ_0^2 *is the true value of* σ^2.

The proof follows directly from Theorem 3.1 and Lemma 4.1 by using $J(\ell) = \ell!$ and $J(j) = 0$ for $j \neq \ell$.

If $\alpha > 1/2$ and $\rho_2 \neq 0$, then Theorem 3.1 implies that $\widehat{\theta}_N$ converges to a non-Gaussian distribution. We see, therefore, that if ℓ is large enough, namely $\ell > 1/(1-\alpha)$, then, on one hand, Y_t has short memory and its normalized sums converge to a Gaussian distribution, but, on the other hand, the corresponding Whittle estimator converges to a non-Gaussian distribution. Because this is a situation where we have weakly dependent observations, we could have expected the Whittle estimator to behave as in the weakly dependent case. In reality, as Corollary 5.1 indicates, the asymptotic behavior of the estimator is still strongly influenced by the underlying long memory. This apparent paradox can be explained by the fact that the quadratic forms characterizing the estimator depend on the two-dimensional process $(G(X_t), G(X_s)),\ t, s \geq 0$ and not merely on the weakly dependent one-dimensional process $G(X_t),\ t \geq 0$.

6. Proof of the results of Section 2.

6.1. *Proof of Theorem* 2.1. The assumptions $\sigma > 0$ and (2.3) imply that Y_t can be represented as one-sided linear process $Y_t = \sum_{k=0}^{\infty} a(k, \theta)\varepsilon(t-k)$ with uncorrelated innovation sequence ε_k with variance σ^2 [see for example, Theorems 5.7.1 and 5.7.2 in Brockwell and Davis (1991)]. The Gaussian sequence (X_j) is, moreover, an ergodic sequence, because it possesses a spectral density and hence a spectral measure which does not have an atom at frequency 0.

Thus, $G(X_j)$ is also an ergodic sequence. We can now apply Theorem 1 of Hannan (1973) to obtain (2.4). □

6.2. *Proof of Theorem* 2.3. We will need a number of preliminary lemmas. The first involves the expansion of $G(X_t)G(X_s)$ in bivariate Hermite polynomials. The expansion holds pointwise because G is a polynomial.

LEMMA 6.1.

$$G(X_t)G(X_s) = \sum_{m,\, n\geq 0} \frac{1}{m!n!}[EG^{(m)}(X_t)G^{(n)}(X_s)]H_{m,\, n}(X_t, X_s). \tag{6.1}$$

PROOF. The expansion

$$G(X_t)G(X_s) = \sum_{m,\, n\geq 0} \frac{J_{t,\, s}(m,\, n)}{m!n!} H_{m,\, n}(X_t, X_s)$$

has only a finite number of terms since G is a polynomial. To identify the coefficients, use the differentiation rules (2.14), to get

$$G^{(m_0)}(x)G^{(n_0)}(y) = \sum_{m\geq m_0,\, n\geq n_0} \frac{1}{(m-m_0)!}\frac{1}{(n-n_0)!} J_{t,\, s}(m, n)H_{m-m_0,\, n-n_0}(x, y).$$

Since $EH_{m-m_0,\, n-n_0}(X_t, X_s)$ equals 1 if $m = m_0$, $n = n_0$, and 0 otherwise, one obtains $EG^{(m_0)}(X_t)G^{(n_0)}(X_s) = J_{t,\, s}(m_0, n_0)$, which identifies the coefficients. This concludes the proof. □

REMARK. Observe, that contrary to the univariate case, the coefficients in the expansion (6.1) are not constants; they depend on t and s. Since the sequence X_t is stationary, they only depend, in fact, on the difference $t - s$.

In the following lemma, we show that $\widehat{\theta}_N - \theta_0$ is a sum of a negligible $o_P(N^{-1})$ term and a quadratic form.

LEMMA 6.2. *Under the conditions of Theorem* 2.3,

$$\widehat{\theta}_N = \theta_0 - (1 + o_P(1))(2\pi\sigma_0^2)^{-1}W_{\theta_0}^{-1}N^{-1}(Y'\nabla A_{N,\, \theta_0}Y) + o_P(N^{-1}).$$

PROOF. By the mean value theorem,

$$Y'\nabla A_{N,\, \widehat{\theta}_N}Y - Y'\nabla A_{N,\, \theta_0}Y = (Y'\nabla^2 A_{N,\, \theta_N^*}Y)(\widehat{\theta}_N - \theta_0), \tag{6.2}$$

where $|\theta_N^* - \theta_0| \leq |\widehat{\theta}_N - \theta_0|$ and $\nabla A_{N,\, \theta} = ((\partial/\partial\theta_1)A_{N,\, \theta}, \dots, (\partial/\partial\theta_p)A_{N,\, \theta})$. Since $Y'\widehat{\theta}_N Y$ minimizes $Y'A_{N,\, \theta}Y$ for $\theta \in \Theta$, we have $\nabla A_{N,\, \widehat{\theta}_N}Y = 0$ if $\widehat{\theta}_N$ belongs to the interior Θ^0 of Θ.

If $|Y'\nabla A_{N,\, \widehat{\theta}_N}Y| > 0$ then $\widehat{\theta}_N$ must lie on the boundary $\partial\Theta$ of Θ and since θ_0 is in the interior, the distance between $\widehat{\theta}_N$ and θ_0 will be at least as big as the distance $\delta = \min_{\theta\in\partial\Theta} |\theta - \theta_0| > 0$ between θ_0 and the boundary $\partial\Theta$. Therefore, using the same argument as in Dahlhaus (1989),

$$P(|Y'\nabla A_{N,\widehat{\theta}_N}Y|>0)\le P(\widehat{\theta}_N\in\partial\Theta)\le P(|\widehat{\theta}_N-\theta_0|\ge\delta)\to 0\ (N\to\infty) \tag{6.3}$$

by Theorem 2.1. Thus

$$-N^{-1}Y'\nabla A_{N,\theta_0}Y=(N^{-1}Y'\nabla^2A_{N,\theta_N^*}Y)(\widehat{\theta}_N-\theta_0)+o_P(N^{-1}). \tag{6.4}$$

By the assumption (2.5), $(\partial^2/\partial\theta_i\partial\theta_j)g_\theta^{-1}(x)$ is a continuous function in (x,θ). Therefore, as in Lemma 1 and 2 of Fox and Taqqu (1986), we get

$$(2\pi)^{-2}N^{-1}Y'\nabla^2A_{N,\theta_N^*}Y\to\frac{\sigma_0^2}{2\pi}W_{\theta_0},\qquad N\to\infty$$

with probability 1. □

Since $Y_t=G(X_t)$, we can write, using Lemma 6.1,

$$\begin{aligned}N^{-1}Y'\nabla A_{N,\theta_0}Y&=N^{-1}\sum_{t,s=1}^{N}\nabla a_{\theta_0}(t-s)G(X_t)G(X_s)\\&=N^{-1}\sum_{m,n\ge 0}S_N^{(m,n)}\\&=\sum_{k=1}^{k^*}N^{-k(1-\alpha)/2}L^{k/2}(N)T_{k,N}+N^{-1/2}V_N+N^{-1/2}\mu_N,\end{aligned} \tag{6.5}$$

where $S_N^{(m,n)}$, μ_N, $T_{k,N}$ and V_N are defined in (1.6), (2.16), (2.17) and (2.18), respectively. The asymptotic expansion (2.20) now follows from relation (6.5) and Lemma 6.2.

Because of (2.3) and (2.6),

$$\mu_N\to 0\quad\text{as } N\to\infty,$$

by Lemma 4 in Giraitis and Surgailis (1990). It is therefore sufficient to focus on $T_{k,N}$, $1\le k\le k^*$ and V_N. These terms involve the quadratic forms $S_N^{(m,n)}$ defined in (1.6). The weights $v_{m,n}(t)$ in the expression of $S_N^{(m,n)}$ satisfy

$$\sum_t|v_{m,n}(t)|<\infty, \tag{6.6}$$

because

$$\sum_t|\nabla a_{\theta_0}(t)E[G^{(m)}(X_t)G^{(n)}(X_0)]|\le C\sum_t|\nabla a_{\theta_0}(t)|<\infty, \tag{6.7}$$

since $|EG^{(m)}(X_t)G^{(n)}(X_0)|\le(E|G^{(m)}(X_0)|^2E|G^{(n)}(X_0)|^2)^{1/2}<\infty$ and (2.10). Relation (6.6) implies that the Fourier transform $\widehat{v}_{m,n}(x):=(2\pi)^{-1}\sum_{t\in\mathbb{Z}}\cdot e^{-itx}v_{m,n}(t)$ of the weights

$$v_{m,n}(t)=\int_{-\pi}^{\pi}e^{itx}\widehat{v}_{m,n}(x)\,dx$$

is a bounded and continuous function, and $\widehat{v}_{m,n}(0)=(2\pi)^{-1}\sum_t v_{m,n}(t)$.

The vectors $T_{1,N}, \dots, T_{k^*,N}$ and V_N in (6.5) are uncorrelated because the bivariate Hermite polynomials $H_{m,n}$ are orthogonal for different values of $m+n$. To derive the asymptotic behavior of the various terms in the expansion (6.5), we shall use central and noncentral limit theorems for quadratic forms obtained in Giraitis, Taqqu and Terrin (1998) and Giraitis and Taqqu (1998). These theorems involve the quantities $d_k^+(\alpha)$, $k \geq 0$, defined as follows. For any $0 < \alpha < 1$,

$$d_0^+(\alpha) = 1 \quad \text{and} \quad d_k^+(\alpha) := \begin{cases} \alpha, & \text{if } k = 1, \\ \max(d_k(\alpha), 0), & \text{if } k \neq 1, \end{cases} \tag{6.8}$$

where

$$d_k(\alpha) = 1 - k(1 - \alpha).$$

Let us consider first the case where $S_N^{(m,n)}$, properly normalized, satisfies a noncentral limit theorem, that is, when it requires a normalization different from $\sqrt{N}$. The limit may or may not be Gaussian. We shall use the following general proposition which may be useful in other contexts as well.

PROPOSITION 6.1. *Suppose that*

$$Q_N^{(m,n)} = \sum_{t,s=1}^{N} \lambda_{m,n}(t-s) H_{m,n}(X_t, X_s),$$

where $m, n \geq 0$, $1 \leq m + n$, (X_t) *is a Gaussian sequence with spectral density* (1.1), *and*

$$d_m^+(\alpha) + d_n^+(\alpha) > 1, \tag{6.9}$$

where d_n^+, $n \geq 0$ *is defined in* (6.8). *Then, for any* $\{\lambda_{m,n}(t)\}$ *satisfying*

$$\sum_{t \in \mathbb{Z}} |\lambda_{m,n}(t)| < \infty, \tag{6.10}$$

the normalized quadratic form

$$[N^{d_m^+(\alpha)+d_n^+(\alpha)} L^{m+n}(N)]^{-1/2} Q_N^{(m,n)}$$

converges in distribution as $N \to \infty$ *to the multiple Itô–Wiener integral*

$$\begin{aligned} I_{m,n} := &\left((2\pi)^{-1} \sum_{t \in \mathbb{Z}} \lambda_{m,n}(t) \right) \\ &\times \int''_{\mathbb{R}^{m+n}} \left[\int_{-\infty}^{\infty} \frac{\exp(i(x_1 + \cdots + x_m + u)) - 1}{i(x_1 + \cdots + x_m + u)} \right. \\ &\qquad \left. \times \frac{\exp(i(x_{m+1} + \cdots + x_{m+n} - u)) - 1}{i(x_{m+1} + \cdots + x_{m+n} - u)} \, du \right] \\ &\times |x_1|^{-\alpha/2} \cdots |x_{m+n}|^{-\alpha/2} Z(dx_1) \cdots Z(dx_{m+n}). \end{aligned} \tag{6.11}$$

The convergence also holds in the sense of finite-dimensional distributions for any finite collection of (m, n).

That proposition follows from the following theorems in Giraitis, Taqqu and Terrin (1998): Theorem 2.1 (when $m \geq 1$, $n \geq 1$), Theorem 2.2 (when $m \geq 1$, $n = 0$ or $m = 0$, $n \geq 1$) and Theorem 3.3 (in the multivariate case). [In the notation of Giraitis, Taqqu and Terrin (1998), we are considering cases (A.1) and (A.5), with $\beta = 0$ and $L_1(N) \sim \widehat{\lambda}_{m,n}(0) = (2\pi)^{-1}\sum_{t\in\mathbb{Z}} \lambda_{m,n}(t)$ $(N \to \infty)$.]

We shall now verify that the conditions of Proposition 6.1 are satisfied with $\lambda_{m,n} = v_{m,n}$ and $Q_N^{(m,n)} = S_N^{(m,n)}$. Relation (6.10) holds because of (6.6). To verify Relation (6.9), recall that $1/(1-\alpha)$ is assumed non-integer, and consequently $d^+_{m+n}(\alpha) > 0$ for $m + n \leq k^*$. Since $d^+_m(\alpha) \geq d^+_{m+n}(\alpha) > 0$ and $d^+_n(\alpha) \geq d^+_{m+n}(\alpha) > 0$, we have

$$d^+_m(\alpha) + d^+_n(\alpha) = 1 + d^+_{m+n}(\alpha) > 1.$$

Proposition 6.1 and Lemma 6.3 below then imply

$$\begin{aligned}(6.12)\qquad &([N^{1+d^+_{m+n}(\alpha)}L^{m+n}(N)]^{-1/2} S_N^{(m,n)})_{m,n\geq 0:\,1\leq m+n\leq k^*}\\ &\Rightarrow (I_{m,n})_{m,n\geq 0:\,1\leq m+n\leq k^*},\end{aligned}$$

where $\Rightarrow$ denotes convergence of finite-dimensional distributions, and where

$$\begin{aligned}(6.13)\qquad I_{m,n} = \sum_t v_{m,n}(t) \int''_{\mathbb{R}^{m+n}} &\frac{\exp(i(x_1+\cdots+x_{m+n}))-1}{i(x_1+\cdots+x_{m+n})}|x_1|^{-\alpha/2}\\ &\cdots |x_{m+n}|^{-\alpha/2} Z(dx_1)\cdots Z(dx_m).\end{aligned}$$

LEMMA 6.3. *For all* $x \in \mathbb{R}$,

$$(6.14)\qquad \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\exp(i(x+u))-1}{i(x+u)}\;\frac{\exp(-iu)-1}{-iu}\,du = \frac{\exp(ix)-1}{ix}.$$

PROOF. Since

$$\sum_{j=0}^{N} \exp(ijx) = \frac{\exp(i(N+1)x)-1}{\exp(ix)-1}$$

and

$$\lim_{N\to\infty} N^{-1}\frac{\exp(ix)-1}{\exp(ix/(N+1))-1} = \frac{\exp(ix)-1}{ix},$$

we have

$$\begin{aligned}&\frac{\exp(ix)-1}{ix}\\ &\quad= \lim_{N\to\infty} N^{-1}\frac{\exp(ix)-1}{\exp(ix/(N+1))-1} = \lim_{N\to\infty} N^{-1}\sum_{j=0}^{N}\exp(ij(x/(N+1)))\\ &\quad= \lim_{N\to\infty}\frac{1}{2\pi}N^{-1}\int_{-\pi}^{\pi}\sum_{t,s=0}^{N}\exp(it(x/(N+1)+u))\exp(-isu)\,du\end{aligned}$$

$$= \frac{1}{2\pi} \lim_{N\to\infty} N^{-1} \int_{-\pi}^{\pi} \left[\frac{\exp(i(x+(N+1)u))-1}{\exp(ix/(N+1)+u)-1} \frac{\exp(-i(N+1)u)-1}{\exp(-iu)-1} \right] du$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{N\to\infty} \mathbb{1}(|u| \le \pi(N+1)) N^{-1}$$

$$\times \left[(N+1)^{-1} \frac{\exp(i(x+u))-1}{\exp(i(x+u)/(N+1))-1} \frac{\exp(-iu)-1}{\exp(-iu/(N+1))-1} \right] du$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(i(x+u))-1}{i(x+u)} \frac{\exp(-iu)-1}{-iu} \, du. \qquad \square$$

Hence relations (2.17), (6.12) and (6.13) imply the convergence (2.21) of $(T_{1,N}, \ldots, T_{k^*,N})$. To prove the convergence (2.22) involving V_N, we shall use Corollary 5.1 of Giraitis and Taqqu (1998), in the form of the following proposition.

PROPOSITION 6.2. *Suppose that each quadratic form*

$$Q_N^{(i)} = \sum_{t,s=1}^{N} b_i(t-s) H_{m_i,n_i}(X_t, X_s), \qquad i = 1, \ldots, k$$

satisfies the assumptions

$$(6.15) \qquad \sum_{t=-\infty}^{\infty} |r(t)|^p < \infty, \qquad \sum_{t=-\infty}^{\infty} |b_i(t)|^{q_i} < \infty, \ \ i = 1, \ldots, k \ (p, q_1, \ldots, q_k \ge 1)$$

and

$$(6.16) \qquad \min(m_i p^{-1}, 1) + \min(n_i p^{-1}, 1) + 2q_i^{-1} \ge 3.$$

Then, as $N \to \infty$, the CLT holds:

$$(6.17) \qquad N^{-1/2}(Q_N^{(1)}, \ldots, Q_N^{(k)}) \Rightarrow (Z^{(1)}, \ldots, Z^{(k)}),$$

where $(Z^{(1)}, \ldots, Z^{(k)})$ is the Gaussian vector with zero mean and cross-covariances

$$(6.18) \qquad \begin{aligned} \sigma_{i,j} &\equiv EZ^{(i)}Z^{(j)} \\ &:= \sum_{l_1,l_2,t\in\mathbb{Z}} b_i(l_1) b_j(l_2) \operatorname{Cov}(H_{m_i,n_i}(X_t, X_{t+l_1}), H_{m_j,n_j}(X_0, X_{l_2})). \end{aligned}$$

We apply this Proposition 6.2 to

$$(6.19) \qquad V_N = N^{-1/2} \sum_{0\le m,n:\, m+n>k^*} \sum_{t,s=1}^{N} v_{m,n}(t-s) H_{m,n}(X_t, X_s),$$

letting b_i's be the $v_{m,n}$'s. There are only a finite number of summands in (6.19) because G is polynomial and hence $v_{m,n} = 0$ for large m and n.

Since $v_{m,n}(t)$ is absolutely summable by (6.6), we can set $q_i = 1$ in (6.16). Thus, to check (6.16), it is sufficient to show that

$$(6.20) \qquad \min(mp^{-1}, 1) + \min(np^{-1}, 1) \ge 1.$$

In view of Lemma 4 in Fox and Taqqu (1986), relations (1.1) and (2.7) imply

$$r(t) = O(|t|^{\alpha-1+\varepsilon}), \qquad t \to \infty \tag{6.21}$$

for every $\varepsilon > 0$, in particular $2\varepsilon < 1-\alpha$, and thus $\sum_{t=-\infty}^{\infty}|r(t)|^p < \infty$ for $p = (1-\alpha-2\varepsilon)^{-1}$. Clearly, (6.20) must be proved only in the case $mp^{-1} < 1$, $np^{-1} < 1$. But then

$$\begin{aligned}\min(mp^{-1}, 1) + \min(np^{-1}, 1) &= mp^{-1} + np^{-1} = (m+n)p^{-1}\\ &= (m+n)(1-\alpha) - (m+n)2\varepsilon > 1,\end{aligned}$$

for small enough $\varepsilon > 0$. Indeed, by definition of k^*, $k^* < 1/(1-\alpha) < k^*+1$, and hence we have $(m+n)(1-\alpha) \ge (k^*+1)(1-\alpha) > 1$. Thus (6.20) holds and therefore, Proposition 6.2 applies to $(S_N^{(m,n)})_{0\le m,n:\, m+n>k^*}$. Since V_N involves a sum over different pairs (m, n) and since the bivariate Hermite polynomials are orthogonal, we obtain

$$V_N \Rightarrow N(0, D_{k^*}),$$

where the entries of the matrix D_{k^*} equal

$$\begin{aligned} d(i, j) = &\sum_{0\le m_1, n_1, m_2, n_2:\, m_1+n_1=m_2+n_2>k^*} (m_1!n_1!m_2!n_2!)^{-1}\\ &\times \sum_{t, s_1, s_2, \in\mathbb{Z}} \dot a_{\theta_0}^{(i)}(s_1)\dot a_{\theta_0}^{(j)} b(s_2)\\ &\times E[G^{(m_1)}(X_0)G^{(n_1)}(X_{s_1})]E[G^{(m_2)}(X_0)G^{(n_2)}(X_{s_2})]\\ &\times \operatorname{Cov}(H_{m_1, n_1}(X_t, X_{t+s_1}), H_{m_2, n_2}(X_0, X_{s_2})).\end{aligned} \tag{6.22}$$

Here $\dot a_{\theta_0}^{(j)}(s) := (\partial/\partial\theta_j)a_\theta(s)|_{\theta=\theta_0}$, $j = 1, \dots, p$. Observe that by (6.1), $d(i, j)$ can also be expressed as

$$\begin{aligned} d(i, j) = \sum_{t\in\mathbb{Z}} \sum_{s_1, s_2, \in\mathbb{Z}} &\dot a_{\theta}^{(i)}(s_1)\dot a_{\theta}^{(j)}(s_2)\\ &\times \Big[\operatorname{Cov}(G(X_t)G(X_{t+s_1}), G(X_0)G(X_{s_2}))\\ &\quad - \sum_{0\le m_1, n_1, m_2, n_2:\, 1\le m_1+n_1=m_2+n_2\le k^*} (m_1!n_1!m_2!n_2!)^{-1}\\ &\quad \times E[G^{(m_1)}(X_0)G^{(n_1)}(X_{s_1})]E[G^{(m_2)}(X_0)G^{(n_2)}(X_{s_2})]\\ &\quad \times \operatorname{Cov}(H_{m_1, n_1}(X_t, X_{t+s_1}), H_{m_2, n_2}(X_0, X_{s_2}))\Big].\end{aligned} \tag{6.23}$$

This concludes the proof of Theorem 2.3. □

6.3. *Proof of Theorem* 2.4. The proof of (2.24) is the same as that of (2.21). To verify (2.25), recall that $T_{k^*,N}$ involves sums $S_N^{(m,n)} = \sum_{t,s=1}^N v_{m,n}(t-s)H_{m,n}(X_t, X_s)$ such that $m+n=k^*$. The Fourier transform $\widehat{v}_{m,n}(x)$ of the weights $v_{m,n}(t)$ is bounded. Thus $|\widehat{v}_{m,n}(x)| \le C|x|^{-\beta}$, $|x| \le \pi$ with $\beta = 0$. Since $k^* = 1/(1-\alpha)$, the parameter $\gamma = 2\beta + d_m^+(\alpha) + d_n^+(\alpha) = d_m^+(\alpha) + d_n^+(\alpha)$ in (2.12) of Giraitis, Taqqu and Terrin (1998), becomes $\gamma = 1$, corresponding to the boundary case between CLT ($\gamma < 1$) and non-CLT ($\gamma > 1$) for the $S_N^{(m,n)}$'s. Using the techniques of that paper, one can verify that (2.25) holds. [Here are some details. Referring to the labeling and notation of Giraitis, Taqqu and Terrin (1998), we have to show that (3.15) can be replaced by $N^{-\gamma^*}\operatorname{Var} Q_N(r_{N,4}^{(j)}) <$ const., where $\gamma^* = 1+\varepsilon$. The proof of Proposition 4.1 applies almost verbatim with $h_\Delta \equiv 0$. The only difference is in $r_{N,4}$, which should be renormalized by $N^{-\gamma^*}$ instead of $[N^\gamma L^*(N)]^{-1}$. Fix K and bound $f(x)$ by $C|x|^{-\alpha-\varepsilon'}$, $\varepsilon' > 0$. (Our ε is then a function of ε' and is small if ε' is small.) Since $\gamma^* > 1$, the argument of the paper applies and yields $N^{-\gamma^*} r_{N,4} \le$ const.] □

7. Proof of Theorem 3.1. We shall apply Theorem 2.3 and show that the contribution of the first term in the expansion (2.20) is negligible. This first term involves

$$N^{-(1-\alpha)/2}L^{1/2}(N)T_{1,N} = 2N^{-1}S_N^{(0,1)}$$

by (2.17) and $S_N^{(0,1)} = S_N^{(1,0)}$. If we can prove that

$$N^{-1}S_N^{(0,1)} = O_P(N^{-1+\delta/2}) = o_P(N^{-1/2}), \tag{7.1}$$

then this term remains negligible even when multiplied by the normalization factor $N^{(1-\alpha)}L^{-1}(N)$ when $1/2 < \alpha < 1$ and $N^{1/2}$ when $0 < \alpha < 1/2$.

To see that Theorem 2.3 and Relation (7.1) then yield the result, observe that if $1/2 < \alpha < 1$, we have $k^* \ge 2$ and hence the determining term is $T_{2,N}$ which converges to $\rho_2 I_2$. If $0 < \alpha < 1/2$, then $1 \le k^* < 2$, that is, $k^* = 1$. In this case, the determining term is V_N, which converges to $N(0, D_1)$. The entries $d(i,j)$ of the variance-covariance matrix D_1 are given by (6.22) or (6.23), but in view of Lemma 7.1 below, they have the simpler expression (3.4) which is used in Theorem 3.1.

Thus, to prove the theorem, it is sufficient to establish (7.1). We must then estimate

$$S_N^{(0,1)} = \sum_{t,s=1}^N v_{0,1}(t-s)X_t = \sum_{t,s=1}^N [E\dot{G}(X_t)G(X_0)]\nabla a_{\theta_0}(t)X_t.$$

To evaluate $E\dot{G}(X_t)G(X_0)$, it is easier to use univariate expansion in Hermite polynomials. Let $k_0 \ge 1$ denote the *Hermite rank* of G, that is, the index at which the expansion (4.5) of G in univariate Hermite polynomials effectively starts. Thus,

$$G(X_t) = \sum_{k \ge k_0} \frac{J(k)}{k!} H_k(X_t), \tag{7.2}$$

where $J(k_0) \neq 0$.

First we obtain an expression of the spectral density of the sequence $(G(X_t))_{t \in \mathbb{Z}}$. Since the Hermite polynomials are orthogonal, it follows from (7.2) that

$$\begin{aligned} E\ G(X_t)G(X_s) &= \sum_{k \ge k_0} \frac{J(k)^2}{k!^2}\ EH_k(X_t)H_k(X_s) \\ &= \sum_{k \ge k_0} \frac{J(k)^2}{k!} r^k(t-s) \\ &= \int_{-\pi}^{\pi} e^{ix(t-s)} s_{\theta_0}(x)\, dx, \end{aligned}$$

where $s_{\theta_0}(x) = \sigma_0^2 g_{\theta_0}(x)$ denotes the spectral density of $G(X_t)$. Expressed in terms of the spectral density $f(x)$ of X_t, it equals

$$s_{\theta_0}(x) = \sigma_0^2 g_{\theta_0}(x) = \sum_{k \ge k_0} \frac{J(k)^2}{k!} f^{(*k)}(x), \qquad |x| \le \pi, \tag{7.3}$$

where

$$f^{(*k)}(y) = \int_{[-\pi,\, \pi]^{k-1}} f(y - x_1 - \cdots - x_{k-1}) f(x_1) \cdots f(x_{k-1}) dx_1 \cdots dx_{k-1} \tag{7.4}$$

is the kth ($k \ge 1$) convolution. (We assume that f and all spectral densities are periodically extended to $\mathbb{R}$ with period 2π.)

A similar argument together with the differentiation rule $\dot{H}_m(x) = mH_{m-1}(x)$, $m \ge 1$, implies

$$\begin{aligned} EG(X_t)\dot{G}(X_s) = E\dot{G}(X_t)G(X_s) &= \sum_{k,\, k' \ge k_0} \frac{J(k)}{k!} \frac{J(k')}{k'!} k' EH_k(X_t)H_{k'-1}(X_s) \\ &= \sum_{k \ge k_0} \frac{J(k)J(k+1)}{k!} r^k(t-s) = \int_{-\pi}^{\pi} e^{i(t-s)} h_{0,1}(x)\, dx, \end{aligned}$$

where

$$h_{0,\,1}(x) = \sum_{k \ge k_0} \frac{J(k)J(k+1)}{k!} f^{(*k)}(x). \tag{7.5}$$

If

$$|f(x)| \le C|x|^{-\mu}, \qquad |x| \le \pi,\ 0 < \mu < 1,$$

then, for $k \ge 1$,

$$f^{(*k)}(x) \le C|x|^{-d_k^+(\mu)}, \qquad |x| \le \pi \tag{7.6}$$

as long as $d_k(\mu) \neq 0$ [obvious for $k = 1$; for $k \geq 2$ see, e.g., Lemma 5.2 in Giraitis and Taqqu (1997)]. Therefore the assumption $f(x) = |x|^{-\alpha} L(1/|x|) \leq C'|x|^{-\alpha-\varepsilon'}$, $\varepsilon' > 0$, and the relations (7.3) and (7.5) easily imply that for any fixed $\varepsilon > 0$,

$$g_{\theta_0}(x) \leq C|x|^{-d^+_{k_0}(\alpha)-\varepsilon}, \quad h_{0,1}(x) \leq C|x|^{-d^+_{k_0}(\alpha)-\varepsilon}. \tag{7.7}$$

In fact, in view of (1.3), one has

$$d^+_{k_0}(\alpha) = \alpha_G(\theta_0) \tag{7.8}$$

[see also the proof of Lemma 5.2 in Giraitis and Taqqu (1997)]. Then, according to the definitions (1.7) and (1.5),

$$\begin{aligned} v_{0,1}(t) = v_{1,0}(t) &= [E\dot{G}(X_t)G(X_0)]\nabla a_{\theta_0}(t) \\ &= \int_{-\pi}^{\pi} e^{it(x+y)} h_{0,1}(x)\nabla g^{-1}_{\theta_0}(y)\,dx\,dy \\ &= \int_{-\pi}^{\pi} e^{itu}\widehat{v}_{0,1}(u)\,du, \end{aligned}$$

where

$$\widehat{v}_{0,1}(u) = \int_{-\pi}^{\pi} \nabla g^{-1}_{\theta_0}(u-x) h_{0,1}(x)\,dx.$$

Now (2.8) and the assumption of the theorem imply

$$\rho_1 = 2\sum_t [E\dot{G}(X_t)G(X_0)]\nabla a_{\theta_0}(t) = 0. \tag{7.9}$$

But the Fourier transform of $E\dot{G}(X_t)G(X_0)$ is $h_{0,1} \in L^p$ for some $p > 1$ by (7.7), and that of $\nabla a_\theta(t)$ is $\nabla g^{-1}_{\theta_0} \in L^\infty$. Thus, by Theorem VII.6.11 of Zygmund [(1979), Vol. I], Parseval's equality holds and

$$2\pi \int_{-\pi}^{\pi} h_{0,1}(x)\nabla g^{-1}_{\theta_0}(x)\,dx = \sum_t [E\dot{G}(X_t)G(X_0)]\nabla a_{\theta_0}(t) = 0, \tag{7.10}$$

so that $\widehat{v}_{0,1}(0) = 0$. Since $g_{\theta_0}(x) = g_{\theta_0}(|x|)$, we have

$$\widehat{v}_{0,1}(u) = \widehat{v}_{0,1}(u) - \widehat{v}_{0,1}(0) = \int_{-\pi}^{\pi} h_{0,1}(x)(\nabla g^{-1}_{\theta_0}(u-x) - \nabla g^{-1}_{\theta_0}(x))\,dx.$$

Since powers are monotone functions, we obtain, by the mean value theorem,

$$\begin{aligned} |\widehat{v}_{0,1}(u)| &\leq C\int_{-\pi}^{\pi} |h_{0,1}(x)||\nabla g^{-1}_{\theta_0}(u-x) - \nabla g^{-1}_{\theta_0}(x)|^{1-\delta}\,dx \\ &\leq C|u|^{1-\delta}\int_{-\pi}^{\pi} |h_{0,1}(x)| \sup_{|u-x|\leq|y|\leq|x|} \left|\frac{d}{dx}\nabla g^{-1}_{\theta_0}(y)\right|^{1-\delta} dx. \end{aligned} \tag{7.11}$$

Using (2.6), (7.7) and (7.8), we get, that for $\delta > 0$,

$$\begin{aligned} |\widehat{v}_{0,1}(u)| &\le C|u|^{1-\delta}\int_{-\pi}^{\pi}(|u-x|^{(\alpha_G(\theta_0)-1-\varepsilon)(1-\delta)} \\ &\qquad + |x|^{(\alpha_G(\theta_0)-1-\varepsilon)(1-\delta)})|x|^{-\alpha_G(\theta)-\varepsilon}\,dx \\ &\le C|u|^{1-\delta}\int_{-\pi}^{\pi}(|u-x|^{-1+\gamma}+|x|^{-1+\gamma})\,dx \le C|u|^{1-\delta} \end{aligned} \tag{7.12}$$

uniformly in $|u| \le \pi$, when $\delta > 0$, $\varepsilon > 0$ are chosen such that $\gamma = \delta(1+\varepsilon-\alpha_G(\theta_0)) - 2\varepsilon > 0$. Observe that the argument applies also for $\alpha_G = 0$. (The constants C change from line to line.) Hence

$$|\widehat{v}_{0,1}(u)| \le C|u|^{1-\delta} \le C|u|^{-\beta}, \tag{7.13}$$

where $\beta = -(1+\alpha-\delta)/2$, when $\delta > 0$ is chosen small enough, and α is the long-memory parameter in (1.1). From (7.13) and (1.1) [see the proof of Theorem 2.2 in Giraitis, Taqqu and Terrin (1998)] it follows that

$$E(S_N^{(0,1)})^2 = E\left[\sum_{t,s=1}^{N} v_{0,1}(t-s)X_t\right]^2 \le CN^{\gamma},$$

where $\gamma \le 2\beta+\alpha+1 = \delta$. Hence $N^{-1}S_N^{(0,1)} = O_P(N^{\gamma-1}) = o_P(N^{-1/2})$, which proves (7.1). This concludes the proof of the theorem. □

The preceding proof used the following lemma.

LEMMA 7.1. *If* $\rho_1 = 0$ *and* $\alpha < 1/2$, *then* $k^* = 1$ *and the* $d(i,j)$ *in* (6.22) *can be expressed as* (3.4).

PROOF. In view of (6.1), by adding and subtracting

$$\begin{aligned} d_{\le}(i,j) = \sum_{t\in\mathbb{Z}}\ \sum_{s_1,s_2\in\mathbb{Z}}\ \sum_{\substack{0\le m_1,n_1,m_2,n_2\\ 1\le m_1+n_1=m_2+n_2\le k^*}} &\dot{a}_{\theta_0}^{(i)}(s_1)\dot{a}_{\theta_0}^{(j)}(s_2)(m_1!n_1!m_2!n_2!)^{-1} \\ &\times E[G^{(m_1)}(X_0)G^{(n_1)}(X_{s_1})]E[G^{(m_2)}(X_0)G^{(n_2)}(X_{s_2})] \\ &\times \operatorname{Cov}(H_{m_1,n_1}(X_t,X_{t+s_1}), H_{m_2,n_2}(X_0,X_{s_2})) \end{aligned}$$

to (6.22), we will get (3.4), provided that we can prove that $d_{\le}(i,j) = 0$. Here $d_{\le}$ is defined as $\lim_{T\to\infty}\sum_{|t|\le T}\sum_{s_1,s_2}$.

Because $\alpha < 1/2$, we have $k^* = 1$, and therefore

$$d_{\le}(i,j) = \sum_{t}\sum_{s_1,s_2} v^{(i)}(s_1)v^{(j)}(s_2)[r(t)+r(t-s_2)+r(t-s_1)+r(t+s_1-s_2)],$$

where

$$v^{(j)}(s) = \dot{a}_{\theta_0}^{(j)}(s)[E\dot{G}(X_0)G(X_s)] = \dot{a}_{\theta_0}^{(j)}(s)[E\dot{G}(X_s)G(X_0)], \qquad j = 1,\dots,p.$$

As in (7.10), $\rho_1 = 0$ implies

$$\sum_s v^{(j)}(s) = 0.$$

Therefore $d_{\leq}(i, j)$ reduces to

$$d_{\leq}(i, j) = \sum_t \left[\sum_{s_1, s_2} v^{(i)}(s_1) v^{(j)}(s_2) r(t + s_1 - s_2) \right].$$

Since $\sum_t |r(t)| = \infty$, one has to be careful about the order of summation.

Heuristically, the term in brackets is a convolution with Fourier transform

$$\widehat{w}(x) = \widehat{v}^{(i)}(x) \widehat{v}^{(j)}(x) f(x), \tag{7.14}$$

and therefore $d_{\leq}(i, j)$ should equal to $2\pi \widehat{w}(0) = 0$. This short argument glosses over many difficulties. To be precise, we consider first

$$R(t) = \sum_{s_2} v^{(j)}(s_2) r(t - s_2).$$

Since, as in (7.13), we have for some small $\varepsilon > 0$,

$$|\widehat{v}^{(j)}(x)| \leq C|x|^{1-\varepsilon}, \qquad |x| \leq \pi,$$

bounded, and $g \in L^p$, for some $p > 1$, we can apply the Parseval equality to $R(t)$ and get

$$R(t) = 2\pi \int_{-\pi}^{\pi} e^{itx} \widehat{v}^{(j)}(x) f(x)\, dx.$$

Using also (1.1),

$$|\widehat{v}^{(j)}(x) f(x)| \leq C|x|^{1-\varepsilon} |x|^{-\alpha-\varepsilon} = C|x|^{1-\alpha-2\varepsilon}, \qquad |x| \leq \pi,$$

is also bounded and therefore we can apply the Parseval inequality again and get

$$d_{\leq}(i, j) = \sum_t \left[\sum_{s_1} v^{(i)}(s_1) R(t + s_1) \right] = \sum_t (2\pi)^2 \int_{-\pi}^{\pi} e^{itx} \widehat{v}^{(i)}(x) \widehat{v}^{(j)}(x) f(x)\, dx.$$

Let $w(t) = \int_{-\pi}^{\pi} e^{itx} \widehat{w}(x)\, dx$, where $\widehat{w}(x)$ was introduced in (7.14). Then

$$d_{\leq}(i, j) = (2\pi)^2 \sum_t w(t).$$

We can apply the previous inequalities to verify that $\widehat{w}(0) = 0$. The delicate part is to check that

$$\sum_t w(t) = 2\pi \widehat{w}(0),$$

that is, that the Fourier series $\sum_t e^{-itx} w(t)$ converges to $2\pi \widehat{w}(x)$ at $x = 0$. From Theorem II.10.7 of Zygmund [(1979), Vol. I], it is sufficient to show that

$$\widehat{w}(0 + y) - \widehat{w}(0) = O(|\log |y||^{-1}) \quad \text{as } y \to 0 \tag{7.15}$$

and

$$w(t) = O(t^{-\delta}) \quad \text{for some } \delta > 0. \tag{7.16}$$

The first relation follows immediately from the previous inequalities. To verify the second, apply a similar argument as in (7.11) and (7.12) to get $|\widehat{v}^{(j)}(u_2) - \widehat{v}^{(j)}(u_1)| \le C|u_2 - u_1|^{1-\varepsilon}$ and then use Theorem II.4.7 of Zygmund (1979), Vol. I, to obtain $|v^{(j)}(s)| = O(|s|^{-1+\varepsilon})$ as $s \to \infty$. Since $r(t) = O(|t|^{\alpha-1+\varepsilon})$ as $t \to \infty$ by (6.21), we have

$$|R(t)| \le \sum_{s_1} |v^{(j)}(s_2) r(t - s_2)| \le C \sum_{s_2} |s_2|^{-1+\varepsilon} |t - s_2|^{\alpha-1+\varepsilon} \le C|t|^{\alpha-1+2\varepsilon},$$

$$|w(t)| \le \sum_{s_1} |v^{(i)}(s_1) R(t + s_1)| \le C \sum_{s_1} |s_1|^{-1+\varepsilon} |t + s_1|^{\alpha-1+2\varepsilon} \le C|t|^{\alpha-1+3\varepsilon},$$

establishing (7.16) and hence the lemma.

This completes the proof of Theorem 3.1. □

Acknowledgments. We thank Hans-Rudi Künsch for raising the question about the Whittle estimator for $G(X_t)$. We also thank an Associate Editor and two referees for their numerous comments and suggestions, which greatly improved the presentation of the paper.

The second author thanks Boston University for its hospitality.

REFERENCES

AVRAM, F. and TAQQU, M. S. (1987). Noncentral limit theorems and Appell polynomials. *Ann. Probab.* **15** 767–775.

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

BREUER, P. and MAJOR, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13** 425–441.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

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

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783.

DOBRUSHIN, R. L. and MAJOR, P. (1979). Non-central limit theorems for non-linear functions of Gaussian fields. *Z. Wahrsch. Verw. Gebiete* **50** 27–52.

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.

GIRAITIS, L. and KOUL, H. (1997). Estimation of the dependence parameter in linear regression with long-range dependent errors. *Stochastic Process. Appl.* **71** 207–224.

GIRAITIS, L., LEIPUS, R. and SURGAILIS, D. (1996). The change-point problem for dependent observations. *J. Statist. Plann. Inference* **53** 297–310.

GIRAITIS, L. and SURGAILIS, D. (1985). CLT and other limit theorems for functionals of Gaussian processes. *Probab. Theory Related Fields* **70** 191–212.

GIRAITIS, L. and SURGAILIS, D. (1986). Multivariate Appell polynomials and the central limit theorem. In *Dependence in Probability and Statistics* (E. Eberlein and M. S. Taqqu, eds.) 21–71. Birkhäuser, Boston.

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

GIRAITIS, L. and TAQQU, M. S. (1997). Limit theorems for bivariate Appell polynomials I. Central limit theorems. *Probab. Theory Related Fields* **107** 359–381.

GIRAITIS, L. and TAQQU, M. S. (1998). Central limit theorems for quadratic forms with time-domain conditions. *Ann. Probab.* **26** 377–398.

GIRAITIS, L., TAQQU, M. S. and TERRIN, N. (1998). Limit theorems for bivariate Appell polynomials II. Non-central limit theorems. *Probab. Theory Related Fields* **110** 333–367.

HANNAN, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* **10** 130–145.

HO, H. C. and HSING, T. (1996). On the asymptotic expansion of the empirical process of long memory moving averages. *Ann. Statist.* **24** 992–1024.

HO, H. C. and HSING, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636–1669.

KOUL, H. L. and SURGAILIS, D. (1997). Asymptotic expansion of M-estimators with long memory errors. *Ann. Statist.* **25** 818–850.

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

ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range 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.

TAQQU, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Verw. Gebiete* **50** 53–83.

TAQQU, M. S. and TEVEROVSKY, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance series. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (R. Adler, R. Feldman and M. S. Taqqu, eds.) 177–217. Birkhäuser, Boston.

TAQQU, M. S., TEVEROVSKY, V. and WILLINGER, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals* **3** 785–798. [Reprinted in *Fractal Geometry and Analysis* (C. J. G. Evertsz, H.-O. Peitgen and R. F. Voss, eds.) World Scientific, Singapore.]

ZYGMUND, A. (1979). *Trigonometric Series I, II*. Cambridge Univ. Press.

DEPARTMENT OF MATHEMATICS
BOSTON UNIVERSITY
111 CUMMINGTON STREET
BOSTON, MASSACHUSETTS 02215
E-MAIL: liudas@math.bu.edu
murad@math.bu.edu