A central limit theorem for the empirical process of a long memory linear sequence

Liudas Giraitis[1,2] and Donatas Surgailis[3]

[1] Heidelberg University, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany
[2] Research supported by the Alexander von Humboldt-Foundation
[3] Institute of Mathematics and Informatics, 2600 Vilnius, Lithuania

Abstract. A central limit theorem for the normalized empirical process, based on a (non-Gaussian) moving average sequence $X_t, t \in \mathbf{Z}$ with long memory, is established, generalizing the results of Dehling and Taqqu (1989). The proof is based on the (Appell) expansion

$$\mathbf{1}(X_t \le x) = F(x) + f(x)X_t + ...$$

of the indicator function, where $F(x) = P[X_t \le x]$ is the marginal distribution function, $f(x) = F'(x)$, and the covariance of the remainder term decays faster than the covariance of X_t . As a consequence, the limit distribution of M-functionals and U-statistics based on such long memory observations is obtained.

1. Introduction .

Statistical inference for long memory time series has gained considerable attention in recent years; see e.g. Beran (1991), Dehling and Taqqu (1989), Koul and Mukherdjee (1993) and the comprehensive survey by Beran (1992) for additional references. Most of the studies deal with Gaussian observations $X_t, t \in \mathbf{Z}$, or (eventually) with instanteneous functionals $X_t = G(Y_t)$ of a Gaussian long memory series $Y_t, t \in \mathbf{Z}$, using well-developped techniques of Hermite expansions (Major (1981)).

A natural generalization of Gaussian long memory series constitute linear (or moving average) processes

$$X_t = \sum_{s \le t} b_{t-s}\xi_s, \qquad t \in \mathbf{Z}, \tag{1.1}$$

with hyperbolically decaying coefficients

$$b_t = L(t)\, t^{-(D+1)/2}, \tag{1.2}$$

where the main parameter $D \in (0,1)$ characterizes the decay rate of the "memory", $L(\cdot)$ is a slowly varying at infinity function, and $\xi_t, t \in \mathbf{Z}$ is an iid sequence with zero mean and variance 1. In particular, (1.1) includes the case of (non-Gaussian) fractional ARMA processes (see Granger and Joyeux (1980)). More generally, condition (1.2) can be replaced by the hyperbolic decay condition of the covariance

$$r(t) = EX_sX_{t+s} = \sum b_k b_{t+k} = L_2(t)t^{-D} \tag{1.3}$$

where $L_2(t) \sim dL^2(t)$ is a slowly varying at infinity function, $d = \int_0^\infty (u(1+u))^{-(1+D)/2}du$.

The main object of this paper is the study of the asymptotics of the *empirical process*

$$F_N(x,t) = \frac{1}{N}\sum_{s=1}^{[Nt]}\big[\mathbf{1}(X_s \le x) - F(x)\big], \tag{1.4}$$

where $X_t, t \in \mathbf{Z}$ is the linear process of (1.1)-(1.2), and

$$F(x) = P[X_t \le x] \tag{1.5}$$

is the (marginal) probability distribution function, which does not depend on $t \in \mathbf{Z}$, thanks to the strict stationarity of the observation process X_t (1.1). Our main result (Theorem 1 below) is a generalization of Dehling and Taqqu (1989), Theorem 1.1, who considered the case of instanteneous functionals of a Gaussian sequence.

Introduce the (two-parameter Skorokhod) space $\mathcal{D} = \mathcal{D}([-\infty, +\infty] \times [0,1])$ of real functions $g = g(x,t)$, $(x,t) \in [-\infty, +\infty] \times [0,1]$, with the sup-topology, and write $\stackrel{\mathcal{D}}{\Longrightarrow}$ for weak convergence of random elements in $\mathcal{D}$ (see Dehling and Taqqu (1989) for details).

The *fractional Brownian motion* $Z(t), t \in [0,1]$ is a (continuous) Gaussian process with zero mean and the covariance

$$EZ(t)Z(s) \;=\; \frac{1}{2}\big(|t|^{2-D} + |s|^{2-D} - |t-s|^{2-D}\big).$$

Put

$$c_D \;=\; \int_0^1 \int_0^1 |x-y|^{-D} dx dy \int_0^\infty |(1+u)u|^{-(D+1)/2} du,$$

and

$$d_N \;=\; d^{1/2} L(N) N^{1-\frac{D}{2}}.$$

Then it is well-known that

$$\sum_{t,s=1}^{N} r(t-s) \;=\; (c_D + o(1))\, d_N^2$$

and

$$\sum_{t,s=1}^{N} |r(t-s)| \;=\; O(d_N^2).$$

Theorem 1. *Assume that the characteristic function $\phi(u) = Ee^{iu\xi_0}$ of the "noise" in (1.1) satisfies the following condition: there exist constants $C > 0, \delta > 0$ such that*

$$|\phi(u)| \le C(1+|u|)^{-\delta}, \qquad u \in \mathbf{R}. \tag{1.6}$$

Moreover, assume that moments $E|\xi_0|^n < \infty$, for all $n \le n^*(D)$, where $9 \le n^*(D) < \infty$ is estimated in Remark 1 below. Then

$$\Big\{d_N^{-1}\,[Nt]\, F_N(x,t);\; (x,t) \in [-\infty,+\infty]\times[0,1]\Big\} \stackrel{\mathcal{D}}{\Longrightarrow} \big\{\sqrt{c_D} f(x) Z(t);\; (x,t) \in [-\infty,+\infty]\times[0,1]\big\}, \tag{1.7}$$

where $f(x) = F'(x)$ is the marginal probability density.

Theorem 1 can be applied to obtain the limit distribution of various estimators of parameters of the marginal distribution $F(x)$ of long memory moving average sequences (1.1) (see Sect. 2). Condition (1.6) is a rather weak smoothness condition on the distribution of the "noise" ; hovewer, it guarantees the existence and infinite differentiability of the density $f(x)$.

The empirical process (1.4) is a particular case of sums of more general "instanteneous" functionals of the form:

$$S_N(t;H) \;=\; \sum_{s=1}^{[Nt]} H(X_s), \tag{1.8}$$

$H(\cdot) \in L^2(F)$, $EH(X_0) = 0$. Limit distribution of (1.8) for long memory linear processes $X_t, t \in \mathbf{Z}$ of (1.1)-(1.2) was studied by Surgailis (1981, 1983), Avram and Taqqu (1987), Giraitis and Surgailis (1989). There, it was found that, at least for "nice" analytic functions $H(\cdot)$, this distribution is basically the same as if

$X_t, t \in \mathbf{Z}$ were Gaussian, with the only difference that the Hermite rank of $H(\cdot)$ has to be replaced by the *Appell rank* corresponding to the expansion

$$H(x) = \sum_{k=0}^{\infty} a_k A_k(x)/k! \tag{1.9}$$

in *Appell polynomials* $A_k(x) = A_k(x; F)$, $k \geq 0$, with the generating function

$$\sum_{k=0}^{\infty} z^k A_k(x)/k! = e^{zx}/Ee^{zX_0}. \tag{1.10}$$

Hovewer, the fact that Appell polynomials are *not* orthogonal in general makes the convergence of (1.9) a very difficult problem and restricts the use of it basically to analytic functions only (see Giraitis and Surgailis (1986)). In the latter case, the coefficients a_k are given by

$$a_k = EH^{(k)}(X_0) = \int_{\mathbf{R}} H^{(k)}(y)dF(y),$$

which can be formally integrated by parts, giving

$$a_k = (-1)^k EH(X_0)Q_k(X_0) = (-1)^k \int_{\mathbf{R}} H(y)Q_k(y)f(y)dy, \tag{1.11}$$

where the functions

$$Q_k(y) := f^{(k)}(y)/f(y); \quad k = 0, 1, ... \tag{1.12}$$

form a *biorthogonal system* in $L^2(F)$ to the Appell system $A_k(y; F); k \geq 0$ (see Giraitis and Surgailis (1986), Daletskii (1991)). The above facts helped us recently to extend the Dobrushin-Major-Taqqu theory to non-smooth functionals $H(\cdot)$ of the linear long memory process (1.1)-(1.2), of arbitrary Appell rank (see Giraitis and Surgailis (1994)). In the case of the (centered) indicator function

$$H(y; x) = \mathbf{1}(y \leq x) - F(x), \quad y \in \mathbf{R},$$

the corresponding Appell coefficient $a_0(x) = 0$, while

$$a_1(x) = -\int_{\mathbf{R}} H(y; x)f'(y)dy = -f(x), \tag{1.13}$$

which is not identically zero. This argument explains the leading term of the (Appell) expansion

$$\mathbf{1}(X_t \leq x) - F(x) = -f(x)X_t + ... \tag{1.14}$$

and the form of the limit empirical process in (1.7), as

$$\{d_N^{-1} \sum_{s=1}^{[Nt]} X_s; t \in [0, 1]\} \overset{\mathcal{D}([0,1])}{\Longrightarrow} \{\sqrt{c_D} Z(t); t \in [0, 1]\}; \tag{1.15}$$

see Davydov (1970) and Gorodetskii (1976).

The main step in the proof of Theorem 1 is the following "weak uniform reduction principle" (c.f. Theorem 3.1 of Dehling and Taqqu (1989)).

Theorem 2. *There are constants $C, \gamma > 0$ such that for any $0 < \epsilon < 1$*

$$P\big[\max_{n \le N} \sup_{x \in \mathbf{R}} d_N^{-1} \big| \sum_{j=1}^{n} (\mathbf{1}(X_j \le x) - F(x) + f(x) X_j) \big| > \epsilon\big] \;\le\; C N^{-\gamma}(1 + \epsilon^{-3}). \tag{1.16}$$

The proof of Theorem 2 uses the chaining argument of Dehling and Taqqu (1989), together with the following asymptotics of the joint (bivariate) probability density $f_t(x_1, x_2)$ of (X_1, X_t):

$$f_t(x_1, x_2) \;=\; f(x_1) f(x_2) + r(t) f'(x_1) f'(x_2) + o(r(t)) \qquad (t \to \infty) \tag{1.17}$$

uniformly in $x_1, x_2 \in \mathbf{R}$ (see Lemma C below). In turn, the proof of (1.17) uses the factorization of the corresponding bivariate characteristic function due to the independence of the noise variables, and the asymptotics (1.2).

The rest of the paper is organized as follows. Sect. 2 is given to the proof of Theorems 1 and 2 and the auxiliary Lemmas A, B, C, and D. Sect. 3 discusses applications to the limit distribution of some statistical functionals based on the empirical distribution function.

2. Proofs of Theorems 1 and 2

Write $\widehat{g}(u) = \int_{\mathbf{R}^d} e^{iu \cdot x} g(x) dx$ for the Fourier transform of a real function $g = g(x)$, $x \in \mathbf{R}^{\mathbf{d}}$ $(d \ge 1)$, whenever it is well-defined.

Lemma A. *For any $k \ge 0$,*

$$\int_{\mathbf{R}} |u|^k |\widehat{f}(u)| du \;<\; \infty \tag{2.1}$$

and

$$\limsup_{|t| \to \infty} \int_{\mathbf{R}^2} |u_1|^k |\widehat{f}_t(u_1, u_2)| du_1 du_2 \;<\; \infty. \tag{2.2}$$

In particular, the marginal probability density $f(x)$ is infinitely differentiable, and the bivariate probability density $f_t(x_1, x_2)$ exists and is jointly continuous in $\mathbf{R}^2$ for any sufficiently large t .

Proof. The characteristic function

$$\widehat{f}(u) = E e^{iuX_0} = \prod_{j \ge 0} \phi(u b_j), \qquad u \in \mathbf{R}, \tag{2.3}$$

where $\phi(u) = E e^{iu\xi_0}$ satisfies

$$|\phi(u)| \;<\; C/(1 + |u|)^{\delta}; \tag{2.4}$$

see (1.6). Choose $J \subset \{j \in \mathbf{Z} : b_j \ne 0\}$ such that $k + 1 < \delta |J| < \infty$. Then

$$|\widehat{f}(u)| \;\le\; \prod_J |\phi(u b_j)| \;\le\; C \prod_J (1 + |u b_j|)^{-\delta} \;\le\; C\,(1 + |u|)^{-\delta |J|},$$

which proves (2.1).

To prove (2.2), let $\{b_{1j}\}, \{b_{2j}\} \in l_2$ be two real sequences such that

$$\delta |J_i| \;>\; k_i + 1, \qquad i = 1, 2, \tag{2.5}$$

where

$$J_1 = \{j \in \mathbf{Z} : |b_{1j}| > |b_{2j}|\}, \tag{2.6}$$

$$J_2 = \{j \in \mathbf{Z} : |b_{1j}| < |b_{2j}|\}, \tag{2.7}$$

and $k_i \geq 0$, $i = 1, 2$. Then

$$\Phi(u_1, u_2) := \sum_{i=1}^{2} \int_{\mathbf{R}^2} |u_i|^{k_i} |\prod_{j \in \mathbf{Z}} \phi(u_1 b_{1j} + u_2 b_{2j})| du_1 du_2 < \infty. \tag{2.8}$$

Indeed, put $\Phi_i(u_1, u_2) = \prod_{j \in J_i} \phi(u_1 b_{1j} + u_2 b_{2j})$, and assume J_i, $i = 1, 2$ are finite. Then, similarly as in the proof of (2.1),

$$|\Phi_1(u_1, u_2)| \leq C \prod_{J_1} (1 + |u_1 b_{1j} + u_2 b_{2j}|)^{-\delta} \leq C \prod_{J_1} (1 + |u_1 + \beta_j u_2|)^{-\delta},$$

where $\beta_j = b_{2j}/b_{1j}$ and $\max|\beta_j| < 1 - \epsilon$, for some $\epsilon > 0$. Therefore

$$|\Phi_1(u_1, u_2)| \leq C \prod_{J_1} (1 + \epsilon|u_1|)^{-\delta} = C(1 + \epsilon|u_1|)^{-\delta|J_1|},$$

and a similar estimate holds for $\Phi_2(u_1, u_2)$. Hence

$$|\Phi(u_1, u_2)| \leq \prod_{i=1}^{2} |\Phi_i(u_1, u_2)| \leq C \prod_{i=1}^{2} (1 + \epsilon|u_i|)^{-\delta|J_i|},$$

which proves (2.8).

Now, take $b_{1j} = b_j$ and $b_{2j} = b_{j-t}$. As $b_j \to 0$ ($|t| \to \infty$), for any $k \geq 0$ there exists $t_0 > 0$ such that J_1, J_2 defined in (2.6), (2.7), respectively, satisfy $\delta|J_i| > k + 1$, $i = 1, 2$, for any $|t| > t_0$. Moreover, the corresponding estimates hold uniformly in $|t| > t_0$, hence (2.2) follows from (2.8).

Lemma B. *Assume that, for some $n \in \mathbf{Z}_+$,*

$$E|\xi_0|^{n+2} < \infty.$$

Then

$$\int_{\mathbf{R}} (1 + |x|)^n |f'(x)| dx < \infty. \tag{2.9}$$

Proof. As $f'(\cdot) \in C^\infty(\mathbf{R})$ (see Lemma A), and

$$\left(u\widehat{f}(u)\right)^{(n+2)} = -\left[x^{n+2} f'(x)\right]^{\widehat{}}(u) \tag{2.10}$$

($\widehat{g}(\cdot)$ stands for the Fourier transform of $g(\cdot)$), it suffices to show that the left hand side of (2.10) is in $L^1(\mathbf{R})$. Indeed, this implies the boundedness of $|x^{n+2} f'(x)|$ and therefore the integrability of $(1 + |x|)^n f'(x)$.

The m-th derivative of the characteristic function (2.3) can be written as

$$\widehat{f}^{(m)}(u) = \sum_{1 \leq |J| \leq m} \sum_{|\mathbf{k}| = m} \binom{m}{\mathbf{k}} \prod_{j \in J} \phi^{(k_j)}(u b_j)\, b_j^{k_j}\, \Phi_{J^c}(u), \tag{2.11}$$

where the first sum is taken over all subsets $J \subset \mathbf{Z}_+$ of cardinality $1 \le |J| \le m$, the second sum is taken over all vectors $\mathbf{k} = (k_j : j \in J)$, $k_j \in \mathbf{Z}_+$, $|\mathbf{k}| = \sum k_j = m$, $\binom{m}{\mathbf{k}} = m!/\prod_{j\in J} k_j!$; finally,

$$\Phi_{J^c}(u) \;=\; \prod_{j\in \mathbf{Z}_+\setminus J} \phi(ub_j).$$

Similarly as in the proof of Lemma A, for any $n \in \mathbf{Z}_+$ and any $m \in \mathbf{Z}_+$ one can find a constant $C = C_{n,m}$ such that, uniformly in $J \subset \mathbf{Z}_+$, $|J| = m$,

$$|\Phi_{J^c}(u)| \;\le\; C\,(1+|u|)^{-n-3}. \tag{2.12}$$

Furthermore, as $|\phi^{(k)}(u)| \le E|\xi_0|^k \quad (k \ge 2)$, $|\phi^{(1)}(u)| \le |u|E\xi_0^2$, hence, according to (2.11), (2.12),

$$\begin{aligned} |\widehat{f}^{(n+2)}(u)| \;&\le\; C(1+|u|)^{-n-3} \sum_{1\le |J|\le n+2}\; \sum_{|\mathbf{k}|=n+2}\; \prod_{j:k_j\ge 2} |b_j|^{k_j} \prod_{j:k_j=1} |u|b_j^2 \\ &\le\; C(1+|u|)^{-3}. \end{aligned}$$

Consequently,

$$\begin{aligned} |(u\widehat{f}(u))^{(n+2)}| \;&=\; |(n+2)\widehat{f}^{(n+1)}(u) + u\widehat{f}^{(n+2)}(u)| \\ &\le\; C(1+|u|)^{-2}. \end{aligned}$$

This proves the lemma, according to (2.10) and the argument above.

Lemma C. *There exists $\delta > 0$ such that, uniformly in $x_1, x_2 \in \mathbf{R}$,*

$$f_t(x_1,x_2) - f(x_1)f(x_2) + r(t)f'(x_1)f'(x_2) \;=\; O(t^{-D-\delta}). \tag{2.13.}$$

Proof. Write $p_t(x_1,x_2)$ for the left hand side of (2.13). Then

$$p_t(x_1,x_2) \;=\; (2\pi)^{-2}\int_{\mathbf{R}^2} e^{-ix\cdot u}\widehat{p}_t(u)du, \tag{2.14}$$

where

$$\widehat{p}_t(u_1,u_2) \;=\; \widehat{f}_t(u_1,u_2) - \widehat{f}(u_1)\widehat{f}(u_2) - r(t)u_1u_2\widehat{f}(u_1)\widehat{f}(u_2). \tag{2.15}$$

According to Lemma A, for any $\delta > 0$,

$$\int_{\mathbf{R}^2} |\widehat{p}_t(u)|\mathbf{1}(|u| > t^\delta)du \;=\; O(t^{-D-\delta}). \tag{2.16}$$

Indeed, choose $k \ge (D+\delta)/\delta$, then

$$\int_{\mathbf{R}^2} |\widehat{f}_t(u)|\mathbf{1}(|u| > t^\delta)du \;\le\; t^{-k\delta}\int_{\mathbf{R}^2} |u|^k|\widehat{f}_t(u)|du \;\le\; Ct^{-k\delta} \;=O(t^{-D-\delta}) \qquad (t\to\infty)$$

by (2.2), and a similar estimate is valid for the two other terms on the right hand side of (2.15).

It remains to show that there exists $\delta > 0$ such that

$$\sup_{|u|\le t^\delta} |\widehat{p}_t(u)| \;=\; O(t^{-D-3\delta}); \tag{2.17}$$

indeed,

$$\Big|\int_{\mathbf{R}^2} e^{ix\cdot u}\widehat{p}_t(u)\mathbf{1}(|u|\le t^\delta)du\Big| \;\le\; \pi t^{2\delta}t^{-D-3\delta} \;=\; O(t^{-D-\delta}).$$

To prove (2.17), write

$$\widehat{f}_t(u_1, u_2) \;=\; \prod_{j\in\mathbf{Z}} \phi(u_1 b_{-j} + u_2 b_{t-j}) \;=\; \prod_{J_1}\cdots\prod_{J_2}\cdots\prod_{J_3}\cdots \;=:\; a_1\cdot a_2\cdot a_3,$$

and, similarly,

$$\widehat{f}(u_1)\widehat{f}(u_2) \;=\; \prod_{j\in\mathbf{Z}} \phi(u_1 b_{-j})\phi(u_2 b_{t-j}) \;=\; \prod_{J_1}\cdots\prod_{J_2}\cdots\prod_{J_3}\cdots \;=:\; a_1'\cdot a_2'\cdot a_3',$$

where

$$\begin{aligned} J_1 &= \{j\in\mathbf{Z} : |j|\le t^{2\delta}\},\\ J_2 &= \{j\in\mathbf{Z} : |j-t|\le t^{2\delta}\},\\ J_3 &= \mathbf{Z}\backslash J_1\cup J_2. \end{aligned}$$

We have

$$\begin{aligned} \widehat{f}_t(u_1,u_2) - \widehat{f}(u_1)\widehat{f}(u_2) &= a_1a_2a_3 - a_1'a_2'a_3'\\ &= (a_1-a_1')a_2a_3 + a_1'(a_2-a_2')a_3 + a_1'a_2'(a_3-a_3'). \end{aligned}$$

Hence, (2.17) follows from $|a_i|\le 1$, $|a_i'|\le 1$, $i=1,2,3$, and

$$a_i - a_i' \;=\; O(t^{-D-3\delta}), \qquad i=1,2, \tag{2.18}$$

$$a_3 - a_3' \;=\; -a_3'\cdot u_1u_2\,r(t) + O(t^{-D-3\delta}), \tag{2.19}$$

uniformly in $|u|\le t^\delta$.

Using the inequality

$$|\prod b_i - \prod b_i'| \;\le\; \sum |b_i - b_i'|,$$

$|b_i|\le 1$, $|b_i'|\le 1$, one has

$$|a_1 - a_1'| \;\le\; \sum_{|j|\le t^{2\delta}} |\phi(u_1b_{-j}+u_2b_{t-j}) - \phi(u_1b_{-j})\phi(u_2b_{t-j})|,$$

where, for $|u_2|\le t^\delta$, $|j|\le t^{2\delta}$, and sufficiently small $\delta>0$,

$$|u_2 b_{t-j}| \;\le\; C|t|^{-D-5\delta}$$

in view of the asymptotics (1.2) of b_t , and the inequality $(D+1)/2 > D$. Therefore, as $|\phi(y_1+y_2) - \phi(y_1)\phi(y_2)| \;\le\; C|y_2|$ with $C = 2\sup_x|\phi'(x)|\le 2E|\xi_0|$, we obtain

$$|a_1 - a_1'| \;=\; O(\sum_{|j|\le t^{2\delta}} |t|^{-D-5\delta}) \;=\; O(|t|^{-D-3\delta}),$$

or eq. (2.18) for $i=1$; the case $i=2$ is analogous.

It remains to prove (2.19). Note that $|u|\le t^\delta$ and $j\in J_3$ imply

$$|u_1b_{-j}| + |u_2b_{t-j}| \;=\; O(t^{-\delta'}),$$

where $0<\delta'<\delta D$. Therefore, for sufficiently large t , the left hand side of (2.19) can be represented as

$$a_3 - a_3' \;=\; a_3'\big[\exp\{Q_t(u_1,u_2)\} - 1\big], \tag{2.20}$$

where

$$Q_t(u_1,u_2) \;:=\; \sum_{|j|>t^{2\delta},|t-j|>t^{2\delta}} \psi(u_1b_j, u_2b_{t-j})$$

and

$$\psi(x,y) \;:=\; \log \frac{\phi(x+y)}{\phi(x)\phi(y)}$$

is well-defined and twice continuosly differentiable in a neighborhood of $(x,y) = 0 \in \mathbf{R^2}$. Note that

$$\psi(x,0) = \psi(0,y) = \psi_x(x,0) = \psi_y(0,y) = 0,$$

$\psi_{xy}(0,0) = \phi''(0) = -1$, and

$$\psi_{xy}(x+y) \;=\; (\log\phi)''(x+y).$$

Hence

$$\psi(v_1,v_2) \;=\; \int_0^{v_1}\int_0^{v_2} (\log\phi)''(x+y)dxdy$$

and, as $(\log\phi)^{(3)}(z)$ is bounded in a neighborhood of $z=0$,

$$\begin{aligned}\psi(u_1b_{-j},u_2b_{t-j}) \;&=\; \int_0^{u_1b_{-j}}\int_0^{u_2b_{t-j}} \big[(\log\phi)''(0) + (x+y)(\log\phi)^{(3)}(z)\big]\,dxdy \\ &=\; -u_1u_2b_{-j}b_{t-j} \;+\; O\big((u_1b_{-j})^2|u_2b_{t-j}| + |u_1b_{-j}|(u_2b_{t-j})^2\big).\end{aligned}$$

Consequently,

$$Q_t(u_1,u_2) \;=\; -r(t)u_1u_2 + i_1 + i_2,$$

where

$$\begin{aligned} i_1 \;&=\; O\Big(|u_1u_2|\big(\sum_{|j|\le t^{2\delta}} |b_{-j}b_{t-j}| + \sum_{|t-j|\le t^{2\delta}} |b_{-j}b_{t-j}|\big)\Big), \\ i_2 \;&=\; O\Big((u_1^2|u_2| + |u_1|u_2^2)\big(\sum_j b_{-j}^2|b_{t-j}| + |b_{-j}|b_{t-j}^2\big)\Big).\end{aligned}$$

Here, as $|u_i| \le t^\delta$, $i=1,2$, so

$$i_1 \;=\; O(t^{2\delta}\max_{|j|\le t^{2\delta}} |b_{t-j}|) \;=\; O(t^{-D-3\delta}),$$

and

$$i_2 \;=\; O(t^{3\delta}\sum_j b_{-j}^2|b_{t-j}|) \;=\; O(t^{-D-3\delta})$$

provided $\delta > 0$ is sufficiently small. In particular,

$$Q_t(u_1,u_2) \;=\; O(t^{2\delta}r(t)) \;=\; O(t^{-D+3\delta})$$

uniformly in $|u_1|,|u_2| \le t^\delta$, which implies

$$\begin{aligned} e^{Q_t(u_1,u_2)} - 1 \;&=\; Q_t(u_1,u_2) + O(Q_t^2(u_1,u_2)) \\ &=\; -u_1u_2r(t) + O(t^{-D-3\delta}).\end{aligned}$$

By (2.20), this proves (2.19) and the lemma, too.

Put

$$S_N(n;x) \;=\; \sum_{j=1}^n \big(\mathbf{1}(X_j \le x) - F(x) + f(x)X_j\big). \tag{2.21}$$

Lemma D. *There is a constant* $\gamma > 0$ *such that, for any* $n \le N$,

$$d_N^{-2}\, ES_N^2(n;x,y) \;\le\; \frac{n}{N^{1+\gamma}}\mu[x,y],$$

where $\mu(\cdot)$ *is a finite measure on* $\mathbf{R}$.

Proof. We have

$$\begin{aligned} S_N(n;x,y) &= S_N(n;x) - S_N(n;y) \\ &= \sum_{j=1}^{n} H(X_j), \end{aligned}$$

where

$$H(z) = H(z;x,y) = \mathbf{1}(y < z \le x) - F(x,y) + (f(x) - f(y))z.$$

Therefore

$$ES_N^2(n;x,y) \;=\; \sum_{t,s=1}^{n} \rho(t-s),$$

where

$$\rho(t) = \rho(t;x,y) = EH(X_0)H(X_t).$$

Using the relations

$$\begin{aligned} \int_{\mathbf{R}} f(z)H(z)dz &= 0, \\ \int_{\mathbf{R}} f'(z)H(z)dz &= 0, \end{aligned}$$

one can represent the covariance $\rho(t)$ as

$$\rho(t) \;=\; \int_{\mathbf{R}}\int_{\mathbf{R}} H(z_1)H(z_2)p_t(z_1,z_2)dz_1dz_2, \tag{2.22}$$

where

$$p_t(z_1,z_2) = f_t(z_1,z_2) - f(z_1)f(z_2) + r(t)f'(z_1)f'(z_2);$$

c.f. (2.13). Assume t is large enough ($t > t_0$). Then, according to Lemma C and (2.22),

$$|\rho(t)| \;\le\; Ct^{-D-\gamma}\int\int |H(z_1)H(z_2)||p_t(z_1,z_2)|^{\theta}dz_1dz_2, \tag{2.23}$$

where $\gamma = (D+\delta)(1-\theta) - D > 0$, provided $\theta > 0$ is chosen sufficiently small.

Let us estimate the last integral, denoted by $i(t)$. Put $A_- = \{(z_1,z_2) \in \mathbf{R^2} : |p_t(z_1,z_2)|^{\theta} \le (1+|z_1|)^{-3}(1+|z_2|)^{-3}\}$; $A_+ = \mathbf{R^2}\backslash A_-$. Then

$$i(t) \;=\; \int\int_{A_-} ... + \int\int_{A_+} ... \;=:\; i_-(t) + i_+(t).$$

Using the estimate

$$|H(z;x,y)| \;\le\; \mathbf{1}(x < z \le y) + F(x,y) + |z|\int_x^y |f'(z)|dz, \tag{2.24}$$

we obtain

$$i_-(t) \le \left(\int_{\mathbf{R}} |H(z)|(1+|z|)^{-3}dz\right)^2 \;\le\; C_-\,\mu_-[x,y], \tag{2.25}$$

where $C_- := \mu_-(\mathbf{R})$ and

$$\mu_-(A) := \int_A (1+|z|)^{-3}dz + F(A)\int_{\mathbf{R}} (1+|z|)^{-3}dz + \int_A |f'(x)|dx \int_{\mathbf{R}} |z|(1+|z|)^{-3}dz$$

is a finite measure on $\mathbf{R}$, according to Lemma B.

Next, as

$$|p_t(z_1,z_2)|^{\theta-1} \le (1+|z_1|)^{n_*}(1+|z_2|)^{n_*}, \qquad (z_1,z_2)\in A_+,$$

with $n_* = [3(1-\theta)/\theta]+1$, we obtain

$$\begin{aligned} i_+(t) &\le \int\int |p_t||H(z_1)H(z_2)|(1+|z_1|)^{n_*}(1+|z_2|)^{n_*}dz_1dz_2 \\ &\le \int\int |p_t|\, H^2(z_1)(1+|z_1|)^{2n_*}dz_1dz_2 \\ &\le 2\int_{\mathbf{R}}\int_{\mathbf{R}} |p_t(z_1,z_2)||H(z_1)|(1+|z_1|)^{2n_*+1}dz_1dz_2 \\ &\le \int_{\mathbf{R}} (4f(z)+2|f'(z)|)\,|H(z)|(1+|z|)^{2n_*+1}dz\,. \end{aligned}$$

Hence, by (2.24),

$$i_+(t) \le \mu_+[x,y], \tag{2.26}$$

where

$$\begin{aligned} \mu_+(A) &:= \int_A (4f(z)+2|f'(z)|)\,(1+|z|)^{2n_*+1}dz \\ &\quad + F(A)\int_{\mathbf{R}} (4f(z)+2|f'(z)|)\,(1+|z|)^{2n_*+1}dz \\ &\quad + \int_A |f'(x)|dx \int_{\mathbf{R}} |z|(4f(z)+2|f'(z)|)\,(1+|z|)^{2n_*+1}dz \end{aligned}$$

is a finite measure, too, provided $E|\xi_0|^{2n_*+4} < \infty$ (see Lemma B). Combining (2.25) and (2.26), we obtain that for all sufficiently large $t > t_0$

$$|\rho(t)| \le t^{-D-\gamma}\mu[x,y], \tag{2.27}$$

where $\mu(\cdot)$ is a finite measure, independent of t . The same bound clearly holds for $0 \le t \le t_0$, too, which can be shown by the same argument as in (2.25),(2.26). Consequently,

$$\begin{aligned} d_N^{-2}ES_N^2(n;x,y) &\le \mu[x,y]\,d_N^{-2}\sum_{t,s=1}^{n}|t-s|^{-D-\gamma} \\ &\le \text{const}\,\mu[x,y]\,(n/N)^{2-D-\gamma}/N^\gamma L^2(N), \end{aligned}$$

which proves the required estimate, provided $\gamma > 0$ is chosen sufficiently small ($\gamma \le 1-D$).

Proof of Theorem 2. Let $\Lambda(x) = \mu(x)$, where $\mu(x) = \mu((-\infty,x])$ is the measure of Lemma D. Then

$$\Lambda(x) \ge F(x) + \int_{-\infty}^{x} |f'(y)|dy, \quad x\in\mathbf{R},$$

and the argument of Dehling and Taqqu (1989), Lemma 3.2, applies with small changes, yielding the estimate

$$P\big[\sup_x d_N^{-1}|S_N(n;x)| > \epsilon\big] < CN^{-\gamma}\big(\epsilon^{-3}(n/N) + (n/N)^{2-D}\big) \tag{2.28}$$

for some $C > 0, \gamma > 0$ and all $n \le N$. From (2.28), Theorem 2 follows similarly to the proof of Theorem 3.1 of Dehling and Taqqu (1989).

Proof of Theorem 1 follows from the "weak uniform reduction principle " (Theorem 2) and the convergence (1.8) (c.f. Dehling and Taqqu (1989)).

Remark 1. From the proofs of Lemmas C and D, one can obtain the following estimate of the order $n^*(D)$ of finite "noise" moments in Theorem 1:

$$n^*(D) \le \begin{cases} 9, & \text{if } 0 < D \le 1/3 \\ (5+3D)/(1-D), & \text{if } 1/3 < D < 1 \,. \end{cases}$$

Remark 2. Theorems 1 and 2 can be generalized for the empirical process based on "nonlinear" observations $Y_j = G(X_j)$, $1 \le j \le N$, where $G(\cdot)$ is a measurable function , and $X_j, j \in \mathbf{Z}$ is a linear moving average process of (1.1)-(1.2).

Remark 3. Theorem 1 and its proof remain valid, with unimportant changes, for a two-sided moving average $X_t = \sum_{-\infty}^{\infty} b_{t-s}\xi_s$, where $b_t = L(t)|t|^{-(D+1)/2}$ and $L(\cdot)$ varies slowly at $\pm\infty$, so that there is a slowly varying function $L_0(t)$, $t > 0$ and two constants l_+, l_- such that there exist the limits

$$\lim_{t\to+\infty} L(\pm t)/L_0(t) \;=\; l_\pm.$$

In such a case, (1.7) holds with $d_N = L_0(N)N$ and

$$c_D \;=\; \int_0^1\int_0^1 |x-y|^{-D}dxdy \int_{-\infty}^{\infty} g(1+u)g(u)|^{-(D+1)/2}du,$$

where $g(u) := |u|^{-(D+1)/2}l_+$ if $u \le 0$; $= |u|^{-(D+1)/2}l_-$ if $u < 0$.

3. Applications to M-estimators and U-statistics

M-estimators. Consider the model

$$\widetilde{X}_j \;=\; X_j + \theta_0, \qquad j \in \mathbf{Z},$$

where $X_j, j \in \mathbf{Z}$ satisfy the conditions of Theorem 1, and $\theta \in \mathbf{R}$ is unknown parameter. Let $\psi(\cdot) : \mathbf{R} \to \mathbf{R}$ be a continuous function such that

$$\Gamma(\theta) := E\psi(\widetilde{X}_0 - \theta) \;=\; E\psi(X_0 + \theta_0 - \theta) = \int_{\mathbf{R}} \psi(x) f(x + \theta - \theta_0)dx$$

is well-defined and 1-1 in a neighborhood $\Theta \ni \theta_0$ and

$$\Gamma(\theta_0) \;=\; 0. \tag{3.1}$$

The M-estimate $\widehat{\theta}_N$ of θ_0 is defined by

$$\begin{aligned} 0 \;&=\; \frac{1}{N}\sum_{j=1}^N \psi(\widetilde{X}_j - \widehat{\theta}_N) \\ &= \int_{\mathbf{R}} \psi(x - \widehat{\theta}_N)d\widetilde{F}_N(x), \end{aligned} \tag{3.2}$$

where $\widetilde{F}_N(x) \;=\; N^{-1}\sum_{j=1}^N \mathbf{1}(\widetilde{X}_j \;\le\; x)$ is the empirical distribution function based on the observations $\widetilde{X}_1, ..., \widetilde{X}_N$.

Theorem 3. *Assume, in addition, that $\psi(\cdot)$ has bounded variation and $\Gamma'(\theta_0) = \int_{\mathbf{R}} \psi(x) f'(x) dx \neq 0$. Then*

$$N d_N^{-1} c_D^{-1/2} (\widehat{\theta}_N - \theta_0) \implies \mathcal{N}(0,1). \tag{3.3}$$

Remark 3. The above result implies, in particular, that the asymptotic efficiency of an M-estimate does not depend on its kernel $\psi(\cdot)$, which is in sharp contrast with the iid case. This difference was first noted by Beran (1991, 1992) in the case of Gaussian functionals.

M-estimators of the slope parameter in linear regression models with long memory moving average errors are discussed in Giraitis, Koul and Surgailis (1994).

Proof of Theorem 3. It is not difficult to check that, under the conditions of the theorem, the estimate $\widehat{\theta}_N$ exists for all sufficiently large N, and is *weakly consistent* , i.e.

$$\widehat{\theta}_N \implies \theta_0 \qquad (N \to \infty). \tag{3.4}$$

According to (3.1) and (3.2),

$$0 = \int_{\mathbf{R}} \psi(x - \widehat{\theta}_N) d\big[N d_N^{-1} c_D^{-1/2} (\widetilde{F}_N - \widetilde{F})(x)\big] + N d_N^{-1} c_D^{-1/2} \int_{\mathbf{R}} \big(\psi(x - \widehat{\theta}_N) - \psi(x - \theta_0)\big) d\widetilde{F}(x), \tag{3.5}$$

where $\widetilde{F}(x) = F(x - \theta_0)$. According to Theorem 2,

$$\begin{aligned} \int_{\mathbf{R}} \psi(x - \theta) d\big[N d_N^{-1} (\widetilde{F}_N - \widetilde{F})(x)\big] &= - \int_{\mathbf{R}} \big[N d_N^{-1} (\widetilde{F}_N - \widetilde{F})(x + \theta)\big] d\psi(x) \\ &= - \int_{\mathbf{R}} \widetilde{f}(x + \theta) d\psi(x)\ d_N^{-1} \sum_{j=1}^{N} X_j + o_P(1) \end{aligned} \tag{3.6}$$

uniformly in $\theta \in \Theta$, where $\widetilde{f}(x) = \widetilde{F}'(x) = f(x - \theta_0)$. Consequently, as $c_D^{-1/2} d_N^{-1} \sum_{j=1}^{N} X_j \implies Z_1$, from (3.4) and the continuity of the integral on the right hand side of (3.6) in θ,

$$\begin{aligned} \int_{\mathbf{R}} \psi(x - \widehat{\theta}_N) d\big[N d_N^{-1} c_D^{-1/2} (\widetilde{F}_N(x) - \widetilde{F}(x))\big] &= - \int_{\mathbf{R}} \widetilde{f}(x + \widehat{\theta}_N) d\psi(x)\ c_D^{-1/2} d_N^{-1} \sum_{j=1}^{N} X_j + o_P(1) \\ &\implies -Z_1 \int_{\mathbf{R}} \widetilde{f}(x + \theta_0) d\psi(x) = -Z_1 \int_{\mathbf{R}} f(x) d\psi(x), \end{aligned} \tag{3.7}$$

where $Z_1 \sim \mathcal{N}(0,1)$. Next , rewrite the second integral on the right hand side of in (3.5) as

$$\int_{\mathbf{R}} \big(\psi(x - \widehat{\theta}_N) - \psi(x - \theta_0)\big) d\widetilde{F}(x) = \int_{\mathbf{R}} \psi(x) \big(f(x + \widehat{\theta}_N - \theta_0) - f(x)\big) dx = \Gamma(\widehat{\theta}_N) - \Gamma(\theta_0).$$

Using the fact that $\Gamma(\cdot)$ is continuously differentiable at θ_0 together with the mean value theorem and the convergence (3.4), we obtain

$$N d_N^{-1} c_D^{-1/2} \big(\Gamma(\widehat{\theta}_N) - \Gamma(\theta_0)\big) = N d_N^{-1} c_D^{-1/2} \big(\widehat{\theta}_N - \theta_0\big)\, \Gamma'(\theta_0) + o_P(1),$$

or the relation (3.3), in view of (3.5), (3.7).

U-statistics. Consider the U-statistic

$$U_N(h) = \sum{}' h(X_{j_1}, ..., X_{j_k}) \tag{3.8}$$

based on the observations $X_1, ..., X_N$ of a linear process satisfying the conditions of Theorem 1, where $h : \mathbf{R}^k \to \mathbf{R}$ is a measurable function, invariant under permutations of its arguments, and the sum $\sum'$ is taken over all integers $1 \leq j_p \leq N$, $p = 1, ..., k; j_p \neq j_q (p \neq q)$. Using Theorem 1 and the argument of Dehling and Taqqu (1989), Corollary 2, one has

Theorem 4. *Let $h(\cdot)$ have bounded total variation and be degenerate , i.e.*

$$\int_{\mathbf{R}} h(x_1, x_2, ..., x_n) dF(x_1) = 0 \qquad (\forall x_2, ..., x_k \in \mathbf{R}).$$

Then

$$\left\{ d_N^{-k} U_{[Nt]}(h);\ 0 \leq t \leq 1 \right\} \stackrel{\mathcal{D}[0,1]}{\Longrightarrow} \left\{ c_D^k \int_{\mathbf{R^k}} h(x_1, ..., x_k) f'(x_1) ... f'(x_k) dx_1 ... dx_k \left(Z(t)\right)^k ;\ 0 \leq t \leq 1 \right\}.$$

References

Arcones, M.A. and Yu, B. (1993). Empirical processes under long range dependence, Preprint.

Avram, F. and Taqqu, M.S. (1987). Generalized powers of strongly dependent random variables. *Ann. Probab.* **15** (1987), 767-775.

Beran, J. (1991). M-estimators for location for Gaussian and related processes with slowly decaying serial correlations. *J. Amer. Statist. Assoc.* **86**, 704-708.

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

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 functionals of Gaussian fields. *Zeitschrift Wahr. verw. Geb.* **50**, 27-52.

Daletskii, Yu. L. (1991). A biorthogonal analog of Hermite polynomials and the inverse Fourier transform with respect to a non-Gaussian measure. *Func. Anal. Appl.* **25(1)**, 136-140.

Davydov, Yu. A. (1970). An invariance principle for random processes, *Theory Probab. Appl.* **15(3)**, 487-498.

Giraitis, L. and Surgailis, D. (1986). Multivariate Appell polynomials and the central limit theorem. In: Eberlein, E. and Taqqu, M.S. (eds.), *Dependence in Probability and Statistics*, Birkháuser, Boston, pp. 27-51.

Giraitis, L. and Surgailis, D. (1989). A limit theorem for polynomials of linear process with long range dependence. *Lithuanian J. Math.* **29**, 290-311.

Giraitis, L. and Surgailis, D. (1994). Non-central limit theorem for non-smooth functionals of long memory moving averages and Appell polynomials, Preprint. L.Giraitis, H.Koul, D.Surgailis (1994). Asymptotic normality of regression estimators with long memory errors. *Preprint.*

Gorodetskii, V.V. (1977). On the convergence to semi-stable Gaussian processes, *Theory Probab. Appl.* **22**, 498-508.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long memory time series and fractional differencing. *J. Tine Series Anal.* **1**, 15-30.

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

Koul, H.L. and Mukherdjee, K. (1993). Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors, *Probab. Th. Rel. F.* **95**, 535-553.

Major, P. (1981). *Multiple Wiener- Ito integrals* , Springer, New York .

Surgailis, D. (1981), Convergence of sums of non-linear functions of moving averages to self-similar processes, *Soviet Math. Doklady* **23**, 247-250.

Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals, *Lithuanian J. Math.* **22**, 185-201.

Taqqu, M.S. (1979). Convergence of iterated processes of arbitrary Hermite rank, *Z. Wahrsch. verw. Geb.* **50**, 53-83.