The Annals of Statistics
1997, Vol. 25, No. 2, 818–850

ASYMPTOTIC EXPANSION OF *M*-ESTIMATORS WITH LONG-MEMORY ERRORS

BY HIRA L. KOUL[1] AND DONATAS SURGAILIS

Michigan State University and Institute of Mathematics and Informatics

This paper obtains a higher-order asymptotic expansion of a class of M-estimators of the one-sample location parameter when the errors form a *long-memory* moving average. A suitably standardized difference between an M-estimator and the sample mean is shown to have a limiting distribution. The nature of the limiting distribution depends on the range of the dependence parameter θ. If, for example, $1/3 < \theta < 1$, then a suitably standardized difference between the sample median and the sample mean converges weakly to a normal distribution provided the common error distribution is symmetric. If $0 < \theta < 1/3$, then the corresponding limiting distribution is nonnormal. This paper thus goes beyond that of Beran who observed, in the case of long-memory Gaussian errors, that M-estimators T_n of the one-sample location parameter are asymptotically equivalent to the sample mean in the sense that $\mathrm{Var}(T_n)/\mathrm{Var}(\bar{X}_n) \to 1$ and $T_n = \bar{X}_n + o_P\left(\sqrt{\mathrm{Var}(\bar{X}_n)}\right)$.

1. Introduction. This paper discusses the higher-order asymptotic behavior of a class of M-estimators of the one-sample location parameter with *long-memory* errors. Beran (1991) observed, in the case of long-memory Gaussian errors, M-estimators T_n of the one-sample location parameter are asymptotically equivalent to the sample mean in the sense that $\mathrm{Var}(T_n)/\mathrm{Var}(\bar{X}_n) \to 1$ and

$$T_n = \bar{X}_n + o_P\left(\sqrt{\mathrm{Var}(\bar{X}_n)}\right). \tag{1.1}$$

A similar fact was established in Koul (1992) and Koul and Mukherjee (1993) for certain classes of M- and R-estimators of the slope parameter in multiple linear regression models. Giraitis and Surgailis (1996) and Giraitis, Koul and Surgailis (1996) extended these results to non-Gaussian moving average errors, in particular, to fractional ARIMA processes.

The natural question arises: is there any difference between any of the above-mentioned estimators asymptotically at any stage? In particular, is there any difference between the asymptotics of the sample median and the sample mean at any stage?

Received January 1996; revised August 1996.
[1] Research partially supported by NSF Grant DMS-94-02904.
AMS 1991 *subject classifications*. Primary 62M10; secondary 65G30.
Key words and phrases. Moving average errors, Appell polynomials, second order efficiency.

The aim of the present paper is to answer these questions in the one-sample location model where one observes $\{X_j\}$ obeying the relation

$$X_j = m + \varepsilon_j, \qquad 1 \le j \le n, \tag{1.2}$$

for some $m \in \mathbf{R}$. Here the errors ε_j, $j \in \mathbf{Z} := \{0, \pm 1, \pm 2, \dots\}$, are assumed to follow the moving average process

$$\varepsilon_j = \sum_{k \le j} b_{j-k}\zeta_k, \qquad b_j = L(j) j^{-(1+\theta)/2}, \qquad j \ge 1, \tag{1.3}$$

for some $\theta \in (0,1)$, where $L(\cdot)$ is a positive, slowly varying function at ∞ and where the random variables ζ_j, $j \in \mathbf{Z}$, are i.i.d., not necessarily Gaussian, with mean 0 and variance 1. We assume that the fourth moment of the "noise" ζ_0 is finite although higher moments may be needed, depending on the value of θ. We also require a rather weak smoothness condition of the distribution function $G(x) = P\{\zeta_0 \le x\}$, namely, there are constants $C, \delta > 0$ such that

$$|E\exp(iu\zeta_0)| \le C(1 + |u|)^{-\delta}, \qquad u \in \mathbf{R}. \tag{1.4}$$

Conditions (1.3) and (1.4) imply that the marginal distribution function $F(x) = P(\varepsilon_0 \le x)$ is infinitely differentiable; see Giraitis, Koul and Surgailis (1994) and also Lemma 4.1 below.

Now, to define M-estimators, let $\psi(x)$, $x \in \mathbf{R}$, be a real-valued function of bounded variation such that $\lambda(x) = E\psi(X_0 - x)$ is well defined, strictly decreasing on $\mathbf{R}$ and

$$\lambda(m) = 0. \tag{1.5}$$

The corresponding M-estimator T_n of the unknown location parameter m based on the observations from the long-memory process of (1.2) and (1.3) is defined by

$$T_n = \operatorname{argmin}\left\{\left|\sum_{j=1}^{n} \psi(X_j - x)\right| : x \in \mathbf{R}\right\}. \tag{1.6}$$

We shall also assume that λ is infinitely differentiable with its kth derivative $\lambda^{(k)}$ satisfying the relation

$$\lambda^{(k)}(x) = \int_{\mathbf{R}} \psi(y - x) f^{(k)}(y - m)\, dy, \qquad k = 0, 1, \dots, \tag{1.7}$$

where $g^{(k)}$ denotes the kth derivative of any function g from $\mathbf{R}$ to $\mathbf{R}$ with the convention that $g^{(0)} = g$. Note that

$$\lambda^{(k)}(x) = -\int_{\mathbf{R}} F^{(k)}(y - m + x)\, d\psi(y), \qquad k = 1, 2, \dots.$$

Often we shall write λ_k for $\lambda^{(k)}(m)$. Note that

$$\lambda_k = \int_{\mathbf{R}} \psi(y) f^{(k)}(y)\, dy = -\int_{\mathbf{R}} F^{(k)}(y)\, d\psi(y), \qquad k = 1, 2, \dots.$$

To formulate our main result, we introduce *Appell polynomials* $A_k(\eta)$, $k = 0, 1, 2, \dots,$ of a real random variable η, by the formal power series

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} A_k(\eta) = \frac{e^{z\eta}}{Ee^{z\eta}}. \tag{1.8}$$

For any $k \ge 0$, $A_k(\eta)$ is a polynomial in η whose coefficients are expressed in terms of moments $\mu_j = E\eta^j$, $j \le k$. In particular, $A_0(\eta) = 1$, $A_1(\eta) = \eta - \mu_1$, $A_2(\eta) = \eta^2 - 2\mu_1\eta + 2\mu_1^2 - \mu_2$, $A_3(\eta) = \eta^3 - 3\mu_1\eta^2 + 3\eta(2\mu_1^2 - \mu_2) + 6\mu_1\mu_2 - \mu_3 - 6\mu_1^3$, and so on. Note that

$$\begin{aligned} &EA_k(\eta) = 1, \quad k=0, \qquad EA_k(\eta)=0, \quad k \ne 0, \\ &A_k(\eta - x) = A_k(\eta), \qquad k \ge 1,\ x \in \mathbf{R}. \end{aligned} \tag{1.9}$$

For more on these polynomials, see, for example, Giraitis and Surgailis (1986) and Avram and Taqqu (1987).

The formal Appell expansion of a nonlinear function $g(\eta)$ is given by

$$g(\eta) = \sum_{k=0}^{\infty} \frac{c_k}{k!} A_k(\eta),$$

provided the coefficients $c_k = (Eg(\eta + y))^{(k)}|_{y=0}$, $k = 0, 1, \dots$ are well defined; see Section 5 in Giraitis and Surgailis (1986) and Giraitis and Surgailis (1994). In particular, for any $x \in \mathbf{R}$, one obtains the following (formal) Appell series expansions:

$$\mathbf{1}(X_j - m \le x) = \sum_{i=0}^{\infty} \frac{(-1)^i F^{(i)}(x)}{i!} A_i(X_j - m), \tag{1.10}$$

$$\psi(X_j - x) = \sum_{i=0}^{\infty} \frac{(-1)^i \lambda^{(i)}(x)}{i!} A_i(X_j - m), \qquad j \ge 1. \tag{1.11}$$

Put

$$\begin{aligned} R_j(\psi; k) &= \psi(X_j - m) - \sum_{i=1}^{k} \frac{(-1)^i \lambda_i}{i!} A_i(X_j - m), \qquad 1 \le j \le n, \\ S_{k,n} &= n^{k\theta/2-1} L^{-k}(n) \sum_{j=1}^{n} A_k(X_j - m), \qquad k = 1, 2, \dots, \\ k^* &\equiv k^*(\theta) = [1/\theta], \end{aligned} \tag{1.12}$$

where $[\cdot]$ is the integer part. Observe that, with $\bar{X}_n := n^{-1} \sum_{j=1}^{n} X_j$,

$$S_{1,n} = n^{\theta/2-1} L^{-1}(n) \sum_{j=1}^{n} (X_j - m) = n^{\theta/2} L^{-1}(n)(\bar{X}_n - m).$$

Introduce the kth-order *Hermite process*

$$Z_k(t) = \int_{\mathbf{R}^k} \left\{ \int_0^t \prod_{i=1}^{k} (v - u_i)_+^{-(1+\theta)/2} \, dv \right\} W(du_1) \cdots W(du_k), \qquad 0 \le t \le 1,$$

$Z_k \equiv Z_k(1)$, where $v_+ = v \vee 0$ and $W(du)$ is the standard Gaussian white noise with mean 0 and variance du. The random variables $\{Z_k\}$ are well defined and orthogonal for $1 \le k < 1/\theta$. See Taqqu (1979) for more on these processes and random variables.

We also need to define the following polynomials. Let $z_0 := 1$, $Q_0 = 0$, $Q_1(z_1) := z_1$, and for $k \ge 2$ define Q_k by the following iterative relation:

$$Q_k(z_1, \dots, z_k) = \frac{1}{\lambda_1} \sum_{j=2}^{k} \frac{\lambda_j}{j!} \sum_{r=0}^{j-1} (-1)^{r-1} \binom{j}{r} \frac{k!}{(k-r)!} z_r \times \mathbf{P}_{k-r}^{(j-r)}(z_1, \dots, z_{k-j+1}) + (-1)^{k-1} \frac{\lambda_k}{\lambda_1} z_k, \tag{1.13}$$

with

$$\mathbf{P}_k^{(p)}(z_1, \dots, z_{k-p+1}) = \sum \binom{k}{j_1, \dots, j_p} \prod_{s=1}^{p} Q_{j_s}(z_1, \dots, z_{j_s}),$$

where the sum is taken over all integers $j_1, \dots, j_p \ge 1$ such that $j_1 + \cdots + j_p = k$, $k \ge p \ge 1$.

Note that $Q_k(z_1, \dots, z_k)$ is a polynomial in the variables $z_1, \dots, z_k \in \mathbf{R}$ with coefficients given by λ_i, $1 \le i \le k$. All polynomials $\mathbf{P}_{k-r}^{(j-r)}$ on the right-hand side of (1.13) are expressed in terms of Q_j, $1 \le j \le k-1$. Also, note that $\mathbf{P}_k^{(1)} = Q_k$ and $\mathbf{P}_k^{(k)} = k!\, Q_1^k$, $\forall\ k \ge 1$. We also give the first few product polynomials $\mathbf{P}_k^{(p)}$ for convenience:

$$\mathbf{P}_2^{(2)} = 2Q_1^2, \qquad \mathbf{P}_3^{(2)} = 6Q_1Q_2,$$

$$\mathbf{P}_4^{(2)} = 4!\left(\frac{2}{1!\,3!} Q_1Q_3 + \frac{1}{4}Q_2^2\right), \qquad \mathbf{P}_4^{(3)} = 4!\,\frac{3}{1!\,1!\,2!} Q_1^2Q_2.$$

Using these, one obtains

$$\begin{aligned} Q_2(z_1, z_2) &= \frac{\lambda_2}{\lambda_1}(z_1^2 - z_2), \\ Q_3(z_1, z_2, z_3) &= \frac{\lambda_3}{\lambda_1}(2z_1^3 - 3z_1z_2 + z_3), \\ Q_4(z_1, z_2, z_3, z_4) &= \frac{3\lambda_2(2\lambda_1\lambda_3 - \lambda_2^2)}{\lambda_1^3}(z_1^2 - z_2)^2 \\ &\quad + \frac{\lambda_4}{\lambda_1}\{3z_1^4 - 6z_1^2z_2 + 4z_1z_3 - z_4\}. \end{aligned} \tag{1.14}$$

In the sequel, $\Rightarrow$ stands for the convergence in distribution of random variables and vectors. We are now ready to state our first result.

THEOREM 1.1. *In addition to the above conditions on ψ and ζ_0, assume that*

$$E\zeta_0^{4 \vee 2k^*(\theta)} < \infty \tag{1.15}$$

and $1/\theta$ *is not an integer. Then the following asymptotic expansion holds if* $\lambda_1 \neq 0$:

$$(1.16)\qquad \begin{aligned} T_n - m = \bar{X}_n - m + \sum_{2\le k\le k^*} \frac{1}{k!} L^k(n) n^{-k\theta/2} Q_k(S_{1,n},\dots,S_{k,n}) \\ + n^{-1/2}\lambda_1^{-1} W_n(\psi). \end{aligned}$$

Furthermore, $\forall\ 1 \le k \le k^*$,

$$(1.17)\qquad Q_k(S_{1,n},\dots,S_{k,n}) \Rightarrow Q_k(Z_1,\dots,Z_k),$$

while $W_n(\psi)$ *is asymptotically normal with mean* 0 *and variance*

$$(1.18)\qquad \sigma_\psi^2 = \sum_{j\in\mathbf{Z}} \mathrm{Cov}\big(R_0(\psi;k^*), R_j(\psi;k^*)\big).$$

Observe that the coefficient of the first nonzero Q_k in (1.16) will dominate the rest of the terms in the expansion. This leads us to the following definitions.

The *second-order Appell rank* $r_A(2) = r_A(2;\psi(\cdot - m))$ of $\psi(\cdot - m)$ is the index of the first nonzero λ_i in the Appell expansion (1.11) of the function $\psi(x-m) + \lambda_1 A_1(x) = \psi(x-m) + \lambda_1(x-m)$. More precisely, if $k^* = 1$, $r_A(2) := 2$, and if $k^* > 1$, then

$$(1.19)\qquad r_A(2) := \begin{cases} k^*+1, & \text{if } \lambda_2 = \cdots = \lambda_{k^*} = 0,\\ \min\{k\ge 2: \lambda_k \ne 0\}, & \text{otherwise.}\end{cases}$$

Theorem 1.2 below shows that the second-order Appell rank determines the limiting distribution of the random variables

$$\sum_{j=1}^{n} \big(\psi(X_j - m) + \lambda_1(X_j - m)\big),$$

which arise in the second-order approximation of T_n.

We also define the *second-order M-rank* $r_M(2) = r_M(2;\psi)$ of an *M*-estimator T_n as follows. If $k^* = 1$, $r_M(2) := 2$, and if $k^* > 1$, then

$$(1.20)\qquad r_M(2) := \begin{cases} k^*+1, & \text{if } Q_2 \equiv \cdots \equiv Q_{k^*} \equiv 0,\\ \min\{k\ge 2: Q_k \not\equiv 0\}, & \text{otherwise,}\end{cases}$$

so that $2 \le r_M(2) \le k^* + 1$ by definition. Using the recurrent formula (1.13) for the polynomials Q_k, Lemma 2.2 below proves the remarkable equality: $r_M(2) = r_A(2)$.

The following corollary shows the importance of $r_M(2)$ in determining the limiting distribution of $T_n - \bar{X}_n$.

COROLLARY 1.1. *Under the conditions of Theorem* 1.1,

$$L^{-r_M(2)}(n) n^{\theta r_M(2)/2}(T_n - \bar{X}_n) \Rightarrow \frac{1}{r_M(2)!} Q_{r_M(2)}(Z_1, \dots, Z_{r_M(2)}),$$

$$2 \le r_M(2) \le k^*,\ k^* > 1,$$

$$\lambda_1 n^{1/2}(T_n - \bar{X}_n) \Rightarrow \mathscr{N}(0, \sigma_\psi^2), \qquad r_M(2) = k^* + 1.$$

Before discussing the following application, we need to mention the following facts about the r.v.'s $\{Z_k\}$. Let

$$c_k^2(\theta) \equiv EZ_k^2 = \frac{k!}{(1-(k\theta/2))(1-k\theta)}\left(\int_0^\infty (u+u^2)^{-(1+\theta)/2}\,du\right)^k, \qquad k \ge 1,\ k\theta < 1.$$

Using some of the results from Taqqu (1979), Surgailis (1982), Giraitis and Surgailis (1986) and Avram and Taqqu (1987), it can be shown that for all integers $k \ge 1$ with $k < 1/\theta$, $E\zeta_0^{2k} < \infty$,

$$(S_{1,n}, \dots, S_{k,n}) \Rightarrow (Z_1, \dots, Z_k) \tag{1.21}$$

and

$$ES_{j,n}S_{l,n} \to \begin{cases} c_j^2(\theta), & j = l, \\ 0, & j \ne l, \end{cases} \qquad 1 \le j, l \le k. \tag{1.22}$$

An application. We shall now discuss a statistical application of Theorem 1.1 and . with a focus on the second-order asymptotic variance comparisons of these estimators. To that effect, suppose, additionally, that f is symmetric around 0, ψ is skew symmetric, that is, $\psi(-x) \equiv -\psi(x)$, and $m = 0$. Then $f^{(k)}(y) \equiv (-1)^k f^{(k)}(-y)$ and integration by parts shows that $\lambda_k = \int_0^\infty \psi(y)[1 - (-1)^k] f^{(k)}(y)\,dy$ so that

$$\lambda_k = \begin{cases} 0, & k = \text{positive even integer}, \\ 2\int_0^\infty \psi(y) f^{(k)}(y)\,dy, & k = \text{positive odd integer}. \end{cases} \tag{1.23}$$

Thus, $\lambda_2 = 0$ and $\lambda_1 \ne 0$. Now, $k^* = 1, 2$ or ≥ 3, depending on if $1/2 < \theta < 1$, $1/3 < \theta < 1/2$ or $0 < \theta < 1/3$, and the second-order M-rank of the corresponding M-estimator T_n is

$$\begin{aligned} r_m(2) &= 2, && 1/2 < \theta < 1, \\ &= 3, && 1/3 < \theta < 1/2, \\ &\ge 3, && 0 < \theta < 1/3. \end{aligned}$$

In particular, if $\lambda_3 \neq 0$, then $r_M(2) = 3$ for $0 < \theta < 1/3$, and from Corollary 1.1 we obtain

$$
\begin{aligned}
&L^{-3}(n) n^{3\theta/2}\big(T_n - \bar{X}_n\big) \Rightarrow \frac{\lambda_3}{3!\,\lambda_1}\big(2Z_1^3 - 3Z_1Z_2 + Z_3\big), && 0 < \theta < 1/3,\\
&n^{1/2}\big(T_n - \bar{X}_n\big) \Rightarrow \mathcal{N}\left(0, \left(\frac{\sigma_\psi}{\lambda_1}\right)^2\right), && 1/3 < \theta < 1,\ \theta \neq 1/2.
\end{aligned} \tag{1.24}
$$

For the remaining discussion here, write $T_{n,\psi}$, $\lambda_{k,\psi}$ and so forth, to emphasize the dependence on ψ.

Now focus on the case $0 < \theta < 1/3$. This case may be thought of as the case of *very long memory*. Note that here the above asymptotic distribution of $T_{n,\psi}$ depends on ψ only through $C_\psi := \lambda_{3,\psi}/\lambda_{1,\psi}$. It is thus natural to define, for any two score functions ψ_1, ψ_2 of the above type, the *second-order asymptotic relative efficiency* (ARE) of T_{n,ψ_1} relative to T_{n,ψ_2} in the present case as the ratio

$$e_{\psi_1,\psi_2} := \big\{C_{\psi_1}/C_{\psi_2}\big\}^2.$$

The three interesting ψ functions are $\psi_I(x) = (1/2)\mathrm{sign}(x)$, $\psi_{II}(x) = 2F(x) - 1$ and the Huber(h) score $\psi_h(x) := h\,\mathrm{sign}(x)\mathbf{1}(|x| > h) + x\mathbf{1}(|x| \le h)$, $h > 0$. The estimator T_{n,ψ_I} is the sample median. In the i.i.d. error case, it is well known that this estimator is asymptotically first-order optimal at the double exponential errors, while $T_{n,\psi_{II}}$ has the same property at the logistic errors. All three scores yield robust estimators against heavy tail errors.

Now consider the case of Gaussian errors ε_j, $j \in \mathbf{Z}$, with mean 0 and covariance $r_t = E\varepsilon_0\varepsilon_t$, $r_0 = \tau^2$. Then $F(x) \equiv \Phi(x/\tau)$, where Φ denotes the d.f. of the standard normal random variable. With φ denoting its density, one verifies that

$$C_{\psi_I} = -1/\tau^2, \qquad C_{\psi_{II}} = -1/2\tau^2, \qquad C_{\psi_h} = \frac{-(h/\tau^3)\varphi(h/\tau)}{(\Phi(h/\tau) - \Phi(0))},$$

and hence

$$
\begin{aligned}
e_{\psi_I,\psi_{II}} &= 4,\\
e_{\psi_h,\psi_I} &= \left(\frac{(h/\tau)\varphi(h/\tau)}{\Phi(h/\tau) - \Phi(0)}\right)^2,\\
e_{\psi_h,\psi_{II}} &= 4\left(\frac{(h/\tau)\varphi(h/\tau)}{\Phi(h/\tau) - \Phi(0)}\right)^2.
\end{aligned}
$$

Thus, in terms of the second-order ARE, $T_{n,\psi_{II}}$ is four times as efficient as the sample median. Also note that $e_{\psi_h,\psi_{II}} \to 4$ as $\tau \to \infty$, while $e_{\psi_h,\psi_{II}} \to 0$ as $\tau \to 0$. Thus, for an arbitrarily large error variance τ^2, $T_{n,\psi_{II}}$ is preferable to T_{n,ψ_h} for all $h > 0$. Moreover, using the fact that the function $\Phi(x) - 1/2 - 2x\varphi(x)$, $x \in \mathbf{R}$, is positive for $x \ge 1.4$, one can say that the estimator T_{n,ψ_h} is preferable to $T_{n,\psi_{II}}$ for all those h for which $h \geq 1.4\tau$, while the opposite is true otherwise.

Observe that in all of the above three cases $\lambda_3 \neq 0$. For a general ψ and the Gaussian errors, (1.11) becomes the Hermite expansion

$$\psi(\varepsilon_j) = \sum_{i=0}^{\infty} \frac{(-1)^i \lambda_i}{i!} H_i(\varepsilon_j),$$

where $H_i(x)$, $i = 0, 1, \ldots,$ are Hermite polynomials, and

$$\lambda_i = (-1)^i E\big(\psi(\varepsilon_j) H_i(\varepsilon_j)\big).$$

Observe that $\lambda_3 = 0$ if and only if

$$\int_{|x|<\tau} (\tau^2 - x^2)\varphi(x/\tau)\, d\psi(x) = \int_{|x|>\tau} (x^2 - \tau^2)\varphi(x/\tau)\, d\psi(x).$$

In particular, $\lambda_3 = 0$ if the measure $d\psi$ is purely atomic with the only atoms at $\pm\tau$.

Next, focus on the case $1/3 < \theta < 1$. In this case, the asymptotic distribution of $T_{n,\psi}$ at (1.24) depends on ψ only through $K_\psi := (\sigma_\psi/\lambda_{1,\psi})^2$, where now

$$\sigma_\psi^2 = \sum_{j\in\mathbf{Z}} E\big(\psi(\varepsilon_0) + \lambda_{1,\psi}\varepsilon_0\big)\big(\psi(\varepsilon_j) + \lambda_{1,\psi}\varepsilon_j\big).$$

In general, it is hard to obtain a closed expression for this. However, some simplification is possible in the case of the above type of Gaussian errors, where we shall now take $\tau = 1$, for convenience. Let $\rho_k := \sum_{j\in\mathbf{Z}} r_j^k$. Because $r_j \propto L^2(|j|)|j|^{-\theta}$, $|j| \to \infty$, ρ_k converges for $k \geq 3$ in the present case. Now, using (1.23) and the orthogonality of Hermite polynomials: $EH_i(\varepsilon_0)H_{i'}(\varepsilon_j) = i!\, r_j^i$, $i = i' = 0, 1, \ldots;\ = 0$, $i \neq i'$; for all j, one obtains

$$\sigma_\psi^2 = \sum_{k=1}^{\infty} \frac{\lambda_{2k+1,\psi}^2 \rho_{2k+1}}{(2k+1)!}, \qquad K_\psi = \sum_{k=1}^{\infty} \left(\frac{\lambda_{2k+1,\psi}}{\lambda_{1,\psi}}\right)^2 \frac{\rho_{2k+1}}{(2k+1)!}.$$

This expression for K_ψ can be used to minimize it over a given class of functions ψ. For example, consider the class of M-estimators corresponding to the class of functions $\{\psi_h,\ h > 0\}$. This class includes the median and the sample mean as the two limiting cases $h \to 0$ and $h \to \infty$, respectively. In this case, the Hermite coefficients $\lambda_k \equiv \lambda_{k,h}$ can be explicitly found: $\lambda_{k,h} = -2H_{k-2}(h)\varphi(h)$, $k = 3, 5, \ldots,$ $\lambda_{1,h} = 1 - 2\Phi(h)$, resulting in

$$K_{\psi_h} = \sum_{k=1}^{\infty} \left(\frac{H_{2k-1}(h)\varphi(h)}{\Phi(h) - \Phi(0)}\right)^2 \frac{\rho_{2k+1}}{(2k+1)!}.$$

The function $h \mapsto K_{\psi_h}$ is continuous and strictly positive on $(0, \infty)$, $\lim_{h\to\infty} K_{\psi_h} = 0$ and $\lim_{h\to 0} K_{\psi_h} = K_{\psi_I}$ is the corresponding constant for the median. It follows that K_{ψ_h} has a well-defined minimum on each compact interval which depends on the covariance function $\{r_t, t \in \mathbf{Z}\}$. By assuming some specific form of the latter, for example, taking it to be the covariance of fractional Gaussian noise or fractional ARIMA errors, the corresponding minimization problem for K_{ψ_h} can be dealt with numerically.

Finally, we state the following conjecture. First, note that because $\lambda_2 = 0$, (1.16) implies that

$$(T_n - m) = (\bar{X}_n - m) + \frac{\lambda_3}{3!\,\lambda_1} \frac{L^3(n)}{n^{3\theta/2}} (2S_{1,n}^3 - 3S_{1,n}S_{2,n} + S_{3,n})$$

$$+ \sum_{4 \le k \le k^*} \frac{1}{k!} L^k(n) n^{-k\theta/2} Q_k(S_{1,n}, \dots, S_{k,n}) \tag{1.25}$$

$$+ n^{-1/2}\lambda_1^{-1} W_n.$$

Now, in the very long memory region, that is, when $0 < \theta < 1/3$, the last two terms in (1.25) are $o_P(n^{-3\theta/2}L^3(n))$. This together with (1.22), which implies that $ES_{1,n}S_{3,n} \to 0$, leads us to the following conjecture even without the assumption of Gaussianity:

$$E(T_n - m)^2 = E(\bar{X}_n - m)^2 + L^4(n)n^{-2\theta}(\lambda_3/\lambda_1)c_{13}(\theta)(1 + o(1)),$$

where

$$c_{13}(\theta) = EZ_1(2Z_1^3 - 3Z_1Z_2 + Z_3)/3 = 2(EZ_1^2)^2 - EZ_1^2Z_2.$$

One way to prove this conjecture would be to show that $EW_n^2(\psi) \to \sigma_\psi^2$, where $W_n(\psi)$ is as in Theorem 1.1.

Assuming the above approximation to be true, it can be used to compare the mean square errors of T_n and $\bar{X}_n$ at the second-order level as follows. First, using some of the results from Taqqu (1978) or directly from the definition of Z_k, one obtains

$$c_{13}(\theta) = 4\{I(\theta)/(1-\theta)\}^2 l(\theta),$$

where

$$I(\theta) = \int_0^\infty (u + u^2)^{-(1+\theta)/2}\,du,$$

$$l(\theta) := \frac{2}{(2-\theta)^2} - \frac{1}{3 - 2\theta} - \int_0^1 (v - v^2)^{1-\theta}\,dv.$$

The function l is negative on $[0, 1]$ with the values ranging from 0 to -0.004994776 so that $c_{13}(\theta) < 0$, $\forall\ 0 < \theta < 1$.

Thus, if $\lambda_{3,\psi}/\lambda_{1,\psi} > 0\ (<0)$, then the mean square error of $T_{n,\psi}$ is smaller (larger) than that of $\bar{X}_n$ in the second-order sense, that is, in the sense that

$$\lim_n L^{-4}(n)n^{2\theta}\left\{E(T_n - m)^2 - E(\bar{X}_n - m)^2\right\} < 0\ (>0).$$

For example, for the Gaussian $X_j \sim \mathcal{N}(0, 1)$,

$$\frac{\lambda_{3,\psi}}{\lambda_{1,\psi}} = \frac{\int (x^2 - 1)\varphi(x)\,d\psi(x)}{\int \varphi(x)\,d\psi(x)} > 0$$

if the support of $d\psi \subset (-\infty, -1) \cup (1, +\infty)$, while $\lambda_{3,\psi}/\lambda_{1,\psi} < 0$ if the support of $d\psi \subset (-1, +1)$. As an example, the estimator corresponding to

$$\psi(x) := -\mathbf{1}(x \le -2) + (x+1)\mathbf{1}(-2 \le x \le -1) + (x-1)\mathbf{1}(1 \le x \le 2) + \mathbf{1}(x \ge 2)$$

would have smaller mean square error in the second-order sense than the sample mean at the long-memory Gaussian errors!

The proof of Theorem 1.1 is based on the expansion of the empirical functionals $\Psi_n(m)$ given in Theorem 1.2 below, where

$$\Psi_n(x) := \frac{1}{n} \sum_{j=1}^{n} \psi(X_j - x), \qquad x \in \mathbf{R}.$$

Also, let $F_n(x) = n^{-1} \sum_{j=1}^{n} \mathbf{1}(X_j - m \le x)$, $x \in \mathbf{R}$, denote the empirical distribution function of the errors.

THEOREM 1.2. *Under the conditions of Theorem* 1.1, *for any function* ψ *of bounded variation*,

$$(1.26) \quad \Psi_n(m) = \sum_{1 \le k < 1/\theta} \frac{(-1)^k}{k!} \lambda_k L^k(n) n^{-k\theta/2} S_{k,n} + n^{-1/2} H_n(\psi),$$

where

$$(1.27) \qquad H_n(\psi) = n^{-1/2} \sum_{j=1}^{n} R_j(\psi; k^*) \Rightarrow \mathcal{N}\big(0, \sigma_\psi^2\big).$$

Moreover,

$$(1.28) \qquad \sup_{x \in \mathbf{R}} |F_n(x) - F(x)| = O_P\big(|\bar{X}_n - m|\big) = O_P\big(L(n) n^{-\theta/2}\big).$$

REMARK 1.1. Recently, Ho and Hsing (1996) established the following result. Assume $E\zeta_0^4 < \infty$ and the distribution function of ζ_0 is $(k^* + 3)$ times differentiable. Let

$$S_{k,n}^0 = n^{k\theta/2-1} L^{-k}(n) \sum_{j=1}^{n} \sum_{s_1,\dots,s_k \ge 0}^{(0)} b_{s_1} \cdots b_{s_k} \zeta_{j-s_1} \cdots \zeta_{j-s_k}, \qquad k \ge 1,$$

where the second sum is taken over all $s_1 \ge 0, \dots, s_k \ge 0$ except for the diagonals $s_i = s_j$, $i \ne j$. Then Ho and Hsing proved that

$$F_n(x) = \sum_{0 \le k < 1/\theta} \frac{(-1)^k F^{(k)}(x)}{k!} L^k(n) n^{-k\theta/2} S_{k,n}^0 + n^{-1/2} H_n^0(x),$$

with $\sup_x n^{-\delta} |H_n^0(x)| \to 0$, almost surely, for every $\delta > 0$. They did not address the question of the weak convergence of H_n^0.

Using the "multinomial formula" for generalized powers [Avram and Taqqu (1987)], one can write the difference $S_{k,n} - S_{k,n}^0$ as a sum over the diagonals of products of Appell polynomials of the noise variables. Here, both $S_{k,n}$ and

$S^0_{k,n}$ converge in distribution to the multiple integral Z_k, while $H_n(x) - H^0_n(x)$ ($=$ a linear combination of the "diagonal" sums $S_{k,n} - S^0_{k,n}$, $2 \le k < 1/\theta$) can be shown to be asymptotically normal under the moment condition (1.15).

Moreover, note that F_n corresponds to the empirical functional $\Psi_n(m)$ with ψ given by an indicator function. Theorem 1.2 thus gives an analogous expansion for a class of ψ functions of bounded variation with the difference that the remainder is shown to converge in distribution. Our assumptions are somewhat stronger than those of Ho and Hsing but the expansion result is valid for a larger class of empirical functionals.

The proof of Theorem 1.2 uses some ideas of Ho and Hsing (1996) and Giraitis, Koul and Surgailis (1996).

2. Proof of Theorem 1.1. The function Ψ_n has bounded variation (as ψ does) and therefore has right and left limits at every point $x \in \mathbf{R}$. In order to avoid ambiguity with the definition of the M-estimator, we define T_n as the upper limit of the set $\operatorname{argmin}\{|\Psi_n(x)|_-: x \in \mathbf{R}\} = \{x \in \mathbf{R}: |\Psi_n(x)|_- = \min_{y \in \mathbf{R}} |\Psi_n(y)|_-\}$, where $|\Psi_n|_-(x) = \min(|\Psi_n(x+)|, |\Psi_n(x-)|)$.

The following two lemmas are needed in the proof of Theorem 1.1. The first one shows that the magnitude of T_n is the same as that of $\bar{X}_n$.

LEMMA 2.1. *Under the conditions of Theorem* 1.1,

$$T_n - m = O_P(\bar{X}_n - m) \tag{2.1}$$

and

$$\rho_n := \Psi_n(T_n) = O_P(n^{-1}). \tag{2.2}$$

PROOF. Assume, without loss of generality, that $m = 0$. By the ergodic theorem, for any $x \in \mathbf{R}$,

$$\Psi_n(x) \to \lambda(x), \qquad n \to \infty \text{ a.s.} \tag{2.3}$$

As Ψ_n and λ have bounded variation, the convergence (2.3) is uniform on every compact interval [Feller (1971), Section 8.10, Problem 8]. Together with the assumption of monotonicity of λ, this implies, in particular, that T_n is strongly consistent: $T_n \to 0$, $n \to \infty$, a.s.

As $X_i \ne X_j$, $i \ne j$, a.s., the jumps $\Delta F_n(x) = F_n(x) - F_n(x-) \le 1/n$ and therefore $\Delta\Psi_n(x) = O(1/n)$ a.s.; indeed,

$$|\Delta\Psi_n(x)| \le \int_{\mathbf{R}} |\Delta F_n(y+x)|\,|d\psi(y)| \le |\psi|/n, \tag{2.4}$$

where $|\psi|$ is the variation of ψ. Hence and by the argument above, including $\lambda(0) = 0$, the relation (2.2) and even the stronger one,

$$|\rho_n| \le |\psi|/n \qquad \forall\, n > n_0(\omega),$$

easily follow. [By (2.4), for sufficiently large n, almost surely, the graph of Ψ_n crosses the x-axis in a neighborhood of 0 at some point T_n with $\Psi_n(T_n+) \le 0$, $\Psi_n(T_n-) \ge 0$ and hence $|\Psi_n(T_n+) - \Psi_n(T_n-)| = |\Psi_n(T_n+)| + |\Psi_n(T_n-)| \le |\psi|/n$.]

To prove (2.1), rewrite

$$
\begin{aligned}
\rho_n &= \int_{\mathbf{R}} \psi(x - T_n)\, dF_n(x) - \int_{\mathbf{R}} \psi(x)\, dF(x) \\
&= \int_{\mathbf{R}} \psi(x - T_n)\, d\{F_n(x) - F(x)\} + \int_{\mathbf{R}} [\psi(x - T_n) - \psi(x)]\, dF(x) \qquad (2.5)\\
&= \rho_n' + \lambda(T_n) - \lambda(0).
\end{aligned}
$$

According to Theorem 1.2, (1.28),

$$
\begin{aligned}
|\rho_n'| &= \left| \int_{\mathbf{R}} (F_n - F)(x + T_n)\, d\psi(x) \right| \qquad (2.6)\\
&\le \sup_x |(F_n - F)(x)| |\psi| = O_P(\overline{X}_n).
\end{aligned}
$$

On the other hand, by the mean value theorem, $\lambda(T_n) - \lambda(0) = T_n \lambda^{(1)}(\tilde{T}_n)$, with $|\tilde{T}_n| \le |T_n| = o_P(1)$ and $(\lambda^{(1)}(\tilde{T}_n))^{-1} = O_P(1)$, due to $\lim_{x \to 0} \lambda^{(1)}(x) = \lambda_1 \neq 0$. Thus (2.1) follows from (2.2), (2.5) and (2.6), thereby completing the proof of Lemma 2.1. □

From now on, for convenience, write $T \equiv T_n - m$ and $S_k \equiv S_{k,n}$. Also, for $x \in \mathbf{R}$, let

$$
\begin{aligned}
R_j(x;k) &= \mathbf{1}(X_j - m \le x) \\
&\quad - \sum_{i=0}^{k} \frac{(-1)^i F^{(i)}(x)}{i!} A_i(X_j - m), \qquad 1 \le j \le n,\ k \ge 0, \\
H_n(x) &= n^{-1/2} \sum_{j=1}^{n} R_j(x;k) \qquad (2.7a)\\
&= n^{1/2} \left(F_n(x) - \sum_{i=0}^{k^*} \frac{(-1)^i F^{(i)}(x) L^i(n)}{i!} n^{-i\theta/2} S_i \right).
\end{aligned}
$$

The relations

$$
\begin{aligned}
R_j(\psi;k) &= \int \psi(x)\, dR_j(x;k), \qquad j, k \ge 0; \qquad (2.7b)\\
H_n(\psi) &= \int \psi(x)\, dH_n(x).
\end{aligned}
$$

are often used in the proofs of Theorems 1.1 and 1.2 below.

The $W_n(\psi)$ in (1.16) involves the r.v. $\int H_n(x + T)\, d\psi(x)$. In view of Lemma 1.1, the following lemma is useful in concluding that it is close to $\int H_n(x)\, d\psi(x)$ in probability.

LEMMA 2.2. *For any* $0 < \gamma < 1$,

$$E\left\{\sup_{|y|<n^{-\gamma}}\left|\int_{\mathbf{R}}(H_n(z+y)-H_n(z))\,d\psi(z)\right|\right\}^2 = O\big(n^{-\gamma}\log_2^2 n\big).$$

PROOF. Clearly, it suffices to prove the lemma with $|y| < n^{-\gamma}$ replaced by $0 \le y < n^{-\gamma}$ and ψ nondecreasing. The proof uses a chaining argument and Lemma 3.1, (3.8), below. Put $y_n = n^{-\gamma}$ and let

$$K = \big[\log_2(ny_n)\big].$$

Consider the sequence of partitions

$$x_{i,k} = y_n i 2^{-k}, \qquad 0 \le i \le 2^k,\ k = 0,1,\dots,K,$$

of the interval $[0, y_n]$. For a $y \in [0, y_n]$ and a $k = 0,1,\dots,K$, define i_k^y by

$$x_{i_k^y,k} \le y < x_{i_k^y+1,k}.$$

Define a chain linking 0 to a given point $y \in [0, y_n]$ by

$$0 = x_{i_0^y,0} \le x_{i_1^y,1} \le \cdots \le x_{i_K^y,K} \le y < x_{i_K^y+1,K}.$$

Let $\Gamma_n(y) = \int_{\mathbf{R}} H_n(z+y)\,d\psi(z)$ and write $g(x,y) = g(y) - g(x)$ for any real function g on $\mathbf{R}$. Note that

$$\Gamma_n(0,y) = \int_{\mathbf{R}}\{H_n(z+y) - H_n(z)\}\,d\psi(z).$$

Now rewrite

$$\begin{aligned}\Gamma_n(0,y) &= \Gamma_n(x_{i_0^y,0}, x_{i_1^y,1}) + \Gamma_n(x_{i_1^y,1}, x_{i_2^y,2}) + \cdots\\ &\quad + \Gamma_n(x_{i_{K-1}^y,K-1}, x_{i_K^y,K}) + \Gamma_n(x_{i_K^y,K}, y),\end{aligned}$$

so that

$$\begin{aligned}\sup_{y\in[0,y_n]}\Gamma_n^2(0,y) &\le 2\left(\sum_{k=0}^{K-1}\sup_{y\in[0,y_n]}\left|\Gamma_n(x_{i_k^y,k}, x_{i_{k+1}^y,k+1})\right|\right)^2\\ &\quad + 2\sup_{y\in[0,y_n]}\Gamma_n^2(x_{i_K^y,K}, y).\end{aligned}$$

Hence, by the Cauchy–Schwarz inequality,

$$\begin{aligned}E\sup_{y\in[0,y_n]}\Gamma_n^2(0,y) &\le 2K\sum_{k=0}^{K-1}E\sup_{y\in[0,y_n]}\Gamma_n^2(x_{i_k^y,k}, x_{i_{k+1}^y,k+1})\\ &\quad + 2E\sup_{y\in[0,y_n]}\Gamma_n^2(x_{i_K^y,K}, y).\end{aligned}$$

We shall first estimate the last term. By the mononicity of F_n, the boundedness of ψ and Lemma 4.1, (4.7), below,

$$
\begin{aligned}
&\left|\Gamma_n(x_{i_K^y, K}, y)\right| \\
&\quad = n^{1/2}\left|\int_{\mathbf{R}} F_n(z + x_{i_K^y, K}, z + y)\, d\psi(z)\right. \\
&\qquad \left. - \sum_{0 \le k < 1/\theta} \frac{(-1)^k L^k(n)}{k!} n^{-k\theta/2} S_k \int_{\mathbf{R}} F^{(k)}(z + x_{i_K^y, K}, z + y)\, d\psi(z)\right| \\
&\quad \le b^{1/2} \int_{\mathbf{R}} F_n(z + x_{i_K^y, K}, z + x_{i_K^y + 1, K})\, d\psi(z) \\
&\qquad + C n^{1/2} y_n 2^{-K} \sum_{0 \le k < 1/\theta} n^{-k\theta/2} L^k(n) |S_k| \\
&\quad \le \left|\Gamma_n(x_{i_K^y, K}, x_{i_K^y + 1, K})\right| + C_1 n^{1/2} y_n 2^{-K} \sum_{0 \le k < 1/\theta} L^k(n) n^{-k\theta/2} |S_k|,
\end{aligned}
$$

where the constant C_1 is independent of y, n, K.

Next, observe that $\forall\ k = 0, 1, \dots, K - 1$,

$$
\begin{aligned}
&\sup_{y \in [0, y_n]} \left|\Gamma_n(x_{i_k^y, k}, x_{i_{k+1}^y, k+1})\right| \\
&\quad = \max_{0 \le j \le 2^{k+1} - 1} \sup_{y \in [x_{j, k+1}, x_{j+1, k+1})} \left|\Gamma_n(x_{i_k^y, k}, x_{i_{k+1}^y, k+1})\right| \\
&\quad \le \max_{0 \le j \le 2^{k+1} - 1} \left|\Gamma_n(x_{j, k+1}, x_{j+1, k+1})\right|.
\end{aligned}
$$

Hence, in view of (3.8) below, we obtain that, for any $0 \le k \le K - 1$,

$$
E \sup_{y \in [0, y_n]} \Gamma_n^2(x_{i_k^y, k}, x_{i_{k+1}^y, k+1}) \le \sum_{i=0}^{2^{k+1}-1} E\Gamma_n^2(x_{i, k+1}, x_{i+1, k+1}) \le C y_n,
$$

and similarly,

$$
E \sup_{y \in [0, y_n]} \Gamma_n^2(x_{i_K^y, K}, x_{i_K^y + 1, K}) \le \sum_{i=0}^{2^K - 1} E\Gamma_n^2(x_{i, K}, x_{i+1, K}) \le C y_n.
$$

Consequently,

$$
E \sup_{y \in [0, y_n]} \Gamma_n^2(0, y) \le C y_n K^2 + C_1 n y_n^2 2^{-2K} E\left(\sum_{0 \le k < 1/\theta} |S_k|\right)^2.
$$

Now, from the definition of K, we obtain $2^{-2K} = O(n^{-2(1-\gamma)})$ and

$$
\begin{aligned}
n y_n^2 2^{-2K} &= O(n^{1 - 2\gamma - 2 + 2\gamma}) = O(n^{-1}), \\
K^2 y_n &= O\left(n^{-\gamma}\left(\log_2 n^{1-\gamma}\right)^2\right) = O\left(n^{-\gamma} \log_2^2 n\right).
\end{aligned}
$$

Thus the proof is complete in view of the fact

$$E\left(\sum_{0\le k<1/\theta}|S_k|\right)^2 = O(1),$$

which follows from (1.22). □

REMARK 2.1. The above proof has roots in the paper of Dehling and Taqqu (1989). The major difference is that we are taking the supremum over already small intervals, while in their paper the supremum is being taken over **R**. Our chain is thus different from the one used in their paper while some of the arguments are similar.

PROOF OF THEOREM 1.1. Again, assume, without loss of generality, that $m = 0$. The first term on the right-hand side of (2.5) can be rewritten as

$$\begin{aligned}-\rho_n' &= \int_{\mathbf{R}} (F_n - F)(x+T)\,d\psi(x)\\ &= \sum_{1\le k\le k^*} \frac{(-1)^k}{k!} L^k(n) n^{-k\theta/2} S_k \int_{\mathbf{R}} F^{(k)}(x+T)\,d\psi(x)\\ &\quad + n^{-1/2}\int_{\mathbf{R}} H_n(x+T)\,d\psi(x),\end{aligned}$$

which together with (2.5) yields

$$\begin{aligned}\lambda(T) - \lambda(0) &= \sum_{1\le k\le k^*} \frac{(-1)^{k-1}}{k!} L^k(n) n^{-k\theta/2} S_k \lambda^{(k)}(T)\\ &\quad + n^{-1/2} V_n(\psi),\end{aligned} \tag{2.8}$$

where $V_n(\psi) := \tilde{V}_n(\psi) + n^{1/2}\rho_n$, with

$$\tilde{V}_n(\psi) = \int_{\mathbf{R}} H_n(x+T)\,d\psi(x). \tag{2.9}$$

The idea of the rest of the proof is to use Taylor's expansion of $\lambda(T)$ and $\lambda^{(k)}(T)$, and then to use an iterative procedure to solve the resulting algebraic equation in T. Before proceeding further, consider the "normal" term $V_n(\psi)$; that is, we claim that $V_n(\psi) \Rightarrow \mathscr{N}(0, \sigma_\psi^2)$. In view of (2.2), this will follow if

$$\tilde{V}_n(\psi) \Rightarrow \mathscr{N}\left(0, \sigma_\psi^2\right). \tag{2.10}$$

Indeed, by (1.27) of Theorem 1.2, (2.10) in turn will follow from

$$h_n \equiv \int_{\mathbf{R}} (H_n(x+T) - H_n(x))\,d\psi(x) = o_P(1). \tag{2.11}$$

As

$$\begin{aligned}P\{|h_n| > \delta\} &\le P\left\{\sup_{|y|<n^{-\gamma}} \left|\int (H_n(x+y) - H_n(x))\,d\psi(x)\right| > \delta\right\}\\ &\quad + P\{|T| > n^{-\gamma}\},\end{aligned}$$

(2.11) follows from (2.1) and Lemma 2.2 upon taking $0 < \gamma < \theta/2$.

Next, applying Taylor's formula to (2.8), we obtain

$$\lambda(T) - \lambda(0) = \sum_{j=1}^{k^*} \frac{\lambda_j}{j!} T^j + O_P(T^{k^*+1})$$

$$(2.12) \qquad = \sum_{k=1}^{k^*} \frac{(-1)^{k-1}}{k!} \left(\frac{L(n)}{n^{\theta/2}} \right)^k \times S_k \left(\sum_{j=k}^{k^*} \frac{\lambda_j}{(j-k)!} T^{j-k} + O_P(T^{k^*-k+1}) \right) + n^{-1/2} V_n(\psi).$$

Note that, with $\chi \equiv \chi_n = L(n) n^{-\theta/2}$ $(= \overline{X}/S_1)$,

$$(2.13) \qquad |T^{k^*+1}| + |\chi^k T^{k^*-k+1}| = o_P(n^{-1/2}), \qquad k = 1, \dots, k^*,$$

according to (2.1). Hence (2.12) can be rewritten as

$$(2.14) \qquad \sum_{j=1}^{k^*} \frac{\lambda_j T^j}{j!} = \sum_{j=1}^{k^*} \frac{\lambda_j T^j}{j!} \sum_{k=1}^{j} (-1)^{k-1} \binom{j}{k} \left(\frac{\chi}{T} \right)^k S_k + n^{-1/2} V_n(\psi) + o_P(n^{-1/2}).$$

From (2.14), we obtain the first iteration:

$$(2.15) \qquad T = \chi S_1 + \begin{cases} n^{-1/2} \lambda_1^{-1} (V_n(\psi) + o_P(1)), & k^* = 1, \\ O_P(\chi^2), & k^* > 1, \end{cases}$$

hence (1.16) for $k^* = 1$.

Let $k^* > 1$. By (2.15), $T^2 = \chi^2 S_1^2 + O_P(\chi^3)$, $T^j = O_P(\chi^3)$, $j \geq 3$. Substituting these relations into (2.14), we obtain

$$\lambda_1 T + \frac{\lambda_2}{2!} (\chi^2 S_1^2 + O_P(\chi^3)) + O_P(\chi^3)$$
$$= \lambda_1 \chi S_1 + \frac{\lambda_2}{2!} \big(2 \chi S_1 (\chi S_1 + O_P(\chi^2)) - \chi^2 S_2 \big) + O_P(\chi^3) + n^{-1/2} V_n(\psi) + o_P(n^{-1/2})$$

or

$$(2.16) \qquad T = \chi S_1 + \frac{1}{2!} \chi^2 Q_2(S_1, S_2) + \begin{cases} n^{-1/2} \lambda_1^{-1} (V_n(\psi) + o_P(1)), & k^* = 2, \\ O_P(\chi^3), & k^* > 2, \end{cases}$$

with $Q_2(z_1, z_2)$ given by (1.14), thereby proving (1.16) for $k^* = 2$.

To prove (1.16) for a general $k^* > 2$, now assume

$$T = \sum_{j=1}^{k-1} \frac{\chi^j}{j!} Q_j + O_P(\chi^k) \tag{2.17}$$

for some $2 \le k \le k^*$. We shall prove

$$T = \sum_{j=1}^{k} \frac{\chi^j}{j!} Q_j + \begin{cases} n^{-1/2}\lambda_1^{-1}(V_n(\psi) + o_P(1)), & k = k^*, \\ O_P(\chi^{k+1}), & k < k^*. \end{cases} \tag{2.18}$$

Assume $k < k^*$; the case $k = k^*$ can be considered in a similar way. From (2.17) one obtains

$$\begin{aligned} T^j &= \begin{cases} \displaystyle\sum_{r=j}^{k} \frac{\chi^r}{r!} \mathbf{P}_r^{(j)} + O_P(\chi^{k+1}), & 2 \le j \le k, \\ O_P(\chi^{k+1}), & j > k, \end{cases} \\ \chi^p T^{j-p} &= \sum_{r=j-p}^{k-p} \frac{\chi^{r+p}}{r!} \mathbf{P}_r^{(j-p)} + O_P(\chi^{k+1}), \qquad 1 \le p < j \le k. \end{aligned} \tag{2.19}$$

Substitute (2.19) into (2.14), that is, into the right-hand side of

$$\lambda_1 T = -\sum_{j=2}^{k} \frac{\lambda_j}{j!} T^j + \sum_{j=1}^{k} \frac{\lambda_j}{j!} \sum_{p=1}^{j} (-1)^{p-1} \binom{j}{p} \chi^p T^{j-p} S_p + O_P(\chi^{k+1})$$

to obtain

$$\begin{aligned} &\lambda_1(T - \chi S_1) \\ &\quad = -\sum_{j=2}^{k} \frac{\lambda_j}{j!} \sum_{r=j}^{k} \frac{\chi^r}{r!} \mathbf{P}_r^{(j)} + \sum_{j=2}^{k} \frac{\lambda_j}{j!} \sum_{p=1}^{j-1} (-1)^{p-1} \binom{j}{p} S_p \sum_{r=j-p}^{k-p} \frac{\chi^{r+p}}{r!} \mathbf{P}_r^{(j-p)} \\ &\qquad + \sum_{j=2}^{k} \frac{\lambda_j}{j!} (-1)^{j-1} S_j \chi^j + O_P(\chi^{k+1}) \\ &\quad = \sum_{j=2}^{k} \frac{\lambda_j}{j!} \sum_{p=0}^{j-1} (-1)^{p-1} \binom{j}{p} S_p \sum_{r=j-p}^{k-p} \frac{\chi^{r+p}}{r!} \mathbf{P}_r^{(j-p)} \\ &\qquad + \sum_{j=2}^{k} \frac{\lambda_j}{j!} S_j (-1)^{j-1} \chi^j + O_P(\chi^{k+1}) \\ &\quad = \sum_{j=2}^{k} \frac{\lambda_j}{j!} \sum_{p=0}^{j-1} (-1)^{p-1} \binom{j}{p} S_p \sum_{q=j}^{k} \frac{\chi^q}{(q-p)!} \mathbf{P}_{q-p}^{(j-p)} \\ &\qquad + \sum_{q=2}^{k} \frac{\lambda_q}{q!} S_q (-1)^{q-1} \chi^q + O_P(\chi^{k+1}) \end{aligned}$$

$$= \sum_{q=2}^{k} \frac{\chi^q}{q!} \left\{ \sum_{j=2}^{q} \frac{\lambda_j}{j!} \sum_{p=0}^{j-1} (-1)^{p-1} \frac{q!}{(q-p)!} \binom{j}{p} S_p \mathbf{P}_{q-p}^{(j-p)} + \lambda_q S_q (-1)^{q-1} \right\}$$
$$+ O_P(\chi^{k+1})$$
$$= \lambda_1 \sum_{q=2}^{k} \frac{\chi^q}{q!} Q_q + O_P(\chi^{k+1}),$$

proving (2.18) and thereby also completing the proof of (1.16) with $W_n(\psi) = V_n(\psi) + o_P(1)$.

The claim (1.17) follows from (1.21). This also completes the proof of Theorem 1.1. □

REMARK 2.2. The recurrent formula for the polynomials Q_k given at (1.13) is obtained from the above argument.

LEMMA 2.3. $r_M(2; \psi) = r_A(2; \psi)$.

PROOF. Obviously, $\lambda_2 = \cdots = \lambda_k = 0$ implies $Q_2 \equiv \cdots \equiv Q_k \equiv 0$, hence the inequality $r_M(2) \geq r_A(2)$. Conversely, let $\lambda_2 = \cdots = \lambda_{k-1} = 0$, $\lambda_k \neq 0$, or $r_A(2) = k$; we need to prove $Q_k \not\equiv 0$. Indeed, according to (1.13),

$$Q_k(z_1, \dots, z_k) = \frac{\lambda_k}{\lambda_1} \sum_{r=0}^{k-1} \binom{k}{r} \frac{1}{(k-r)!} z_r \mathbf{P}_{k-r}^{(k-r)}(z_1) + (-1)^{k-1} \frac{\lambda_k}{\lambda_1} z_k \not\equiv 0,$$

as the last sum does not contain the variable z_k. Hence the lemma is proved. □

3. Proof of Theorem 1.2. As in the proof of Theorem 1.1, we assume $m = 0$, so that $X_j = \varepsilon_j = \sum_{s \leq j} b_{j-s} \zeta_s = \sum_{s \geq 0} b_s \zeta_{j-s}$. We start with the expansion [cf. Ho and Hsing (1996)]

$$\psi(X_j) = \sum_{s \geq 0} \left(E\{\psi(X_j) \mid \mathcal{F}_{j-s}\} - E\{\psi(X_j) \mid \mathcal{F}_{j-s-1}\} \right), \tag{3.1}$$

where $\mathcal{F}_j = \sigma\{\zeta_s : s \leq j\}$ is the σ-algebra generated by the "noise" in the "past" $s \leq j$ so that X_j is $\mathcal{F}_j$-measurable. The summands on the right-hand side of (3.1) are orthogonal in $L^2(\Omega)$ and the series converges in this space. Put

$$X_{j,s} = \sum_{1 \leq i \leq s} b_i \zeta_{j-i}, \qquad \tilde{X}_{j,s} = \sum_{i > s} b_i \zeta_{j-i}.$$

Let $F_s(x) := P\{X_{j,s} \leq x\}$, $f_s(x) := F_s^{(1)}(x)$, $x \in \mathbf{R}$. Note that

$$E\{\psi(X_j) \mid \mathcal{F}_{j-s-1}\} = \int_{\mathbf{R}} \psi(x)\, dF_s(x - \tilde{X}_{j,s}) = \langle \psi(\cdot + \tilde{X}_{j,s}), F_s \rangle. \tag{3.2}$$

Next, by the multinomial formula for Appell polynomials, for any $k \geq 1$,

$$A_k(X_j) = \sum_{r=1}^{k} \binom{k}{r} \sum_{s \geq 0} b_s^r A_r(\zeta_{j-s}) A_{k-r}(\tilde{X}_{j,s}), \tag{3.3}$$

$A_0(\tilde{X}_{j,s}) \equiv 1$. From (3.1)–(3.3) we obtain the following expansion of the empirical functional:

$$R_j(\psi; k) = \psi(X_j) - \sum_{i=1}^{k} \frac{(-1)^i}{i!} \lambda_i A_i(X_j) = \sum_{s \geq 0} U_{j,s}(\psi, k), \tag{3.4}$$

where

$$\begin{aligned} U_{j,s}(\psi; k) &= E\{\psi(X_j) \mid \mathcal{F}_{j-s}\} - E\{\psi(X_j) \mid \mathcal{F}_{j-s-1}\} \\ &\quad - \sum_{i=1}^{k} \frac{(-1)^i}{i!} \lambda_i \sum_{r=1}^{i} \binom{i}{r} b_s^r A_r(\zeta_{j-s}) A_{i-r}(\tilde{X}_{j,s}) \\ &= \left\langle \psi\left(\cdot + \tilde{X}_{j,s-1}\right), F_{s-1} \right\rangle - \psi\left\langle\left(\cdot + \tilde{X}_{j,s}\right), F_s \right\rangle \\ &\quad - \sum_{i=1}^{k} \frac{(-1)^i}{i!} \lambda_i \sum_{r=1}^{i} \binom{i}{r} b_s^r A_r(\zeta_{j-s}) A_{i-r}(\tilde{X}_{j,s}). \end{aligned} \tag{3.5}$$

Recall that $A_r(\zeta_{j-s})$ and $A_{i-r}(\tilde{X}_{j,s})$ are polynomials of respective degrees r and $i - r$. Thus (1.9), (1.15), the boundedness of ψ and the independence of ζ_{j-s} from $\tilde{X}_{j,s}$ for all j, s imply that

$$EU_{j,s}(\psi; k) = 0, \qquad EU_{j,s}^2(\psi; k) < \infty, \tag{3.6}$$

for all $s \geq 0$, $j \in \mathbf{Z}$ and $k \leq k^*$.

Now, according to (3.4) and (3.5),

$$H_n(\psi) = n^{-1/2} \sum_{j=1}^{n} R_j(\psi; k^*) = n^{-1/2} \sum_{s \geq 0} \sum_{j=1}^{n} U_{j,s}(\psi; k^*).$$

Put also

$$R_{j,t}(\psi; k) = \sum_{0 \leq s < t} U_{j,s}(\psi; k), \qquad \tilde{R}_{j,t}(\psi; k) = \sum_{s \geq t} U_{j,s}(\psi; k),$$

so that

$$H_n(\psi) = H_{n,t}(\psi) + \tilde{H}_{n,t}(\psi),$$

where

$$H_{n,t}(\psi) = n^{-1/2} \sum_{j=1}^{n} R_{j,t}(\psi; k^*), \qquad \tilde{H}_{n,t}(\psi) = n^{-1/2} \sum_{j=1}^{n} \tilde{R}_{j,t}(\psi; k^*).$$

The following lemma is the central technical lemma of the paper, with the help of which we prove Theorem 1.2. It is useful to show that Var $\tilde{H}_{n,t}(\psi) \to 0$ as $t \to \infty$, uniformly in n. The proof of the lemma itself will be postponed until Section 4. To state the lemma, recall (2.7b) and define, for $y \in \mathbf{R}$,

$$\Delta_y R_j(\psi; k) = R_j(\Delta_j \psi; k) = \int_{\mathbf{R}} (\psi(x+y) - \psi(x))\, dR_j(x; k),$$

where $\Delta_y \psi(x) = \psi(x+y) - \psi(x)$.

LEMMA 3.1. (i) *For any integers* $1 \le k < 1/\theta$, $t \ge 0$ *and any* $\delta_0 > 0$, *one can find a* $0 < \delta < \delta_0$ *and a constant* $C(t) \to 0$, $t \to \infty$, *such that*

$$\left|\operatorname{Cov}\left(\tilde{R}_{j,t}(\psi;k), \tilde{R}_{j',t}(\psi;k)\right)\right| < C(t/(|j'-j| \vee 1))(1 \vee |j'-j|)^{-\gamma(k+1)}, \qquad j, j' \in \mathbf{Z},$$

where

$$\gamma(k) = \gamma(k;\theta,\delta) = \begin{cases} k\theta - \delta, & \text{if } k\theta < 1, \\ 1+\delta, & \text{if } k\theta > 1. \end{cases} \tag{3.7}$$

(ii) *Moreover*, *for any* $|y| < 1$,

$$\left|\operatorname{Cov}\left(\Delta_y R_j(\psi;k), \Delta_y R_{j'}(\psi;k)\right)\right| < C\,|y|(1 \vee |j'-j|)^{-\gamma(k+1)}, \qquad j, j' \in \mathbf{Z},$$

where the constant C does not depend on y, j, j'.

As a consequence of Lemma 3.1(ii), we obtain

$$\begin{aligned} &E\left(\int_{\mathbf{R}} (H_n(x+y) - H_n(x))\, d\psi(x)\right)^2 \\ &\quad = E\left(\int_{\mathbf{R}} (\psi(x-y) - \psi(x))\, dH_n(x)\right)^2 \\ &\quad = n^{-1} \sum_{j,j'=1}^{n} E\left(\Delta_{-y} R_j(\psi;k^*)\, \Delta_{-y} R_{j'}(\psi;k^*)\right) \\ &\quad \le C\,|y|\, n^{-1} \sum_{j,j'=1}^{n} (1 \vee |j'-j|)^{-1-\delta} \le C\,|y| \end{aligned} \tag{3.8}$$

as $(k^*+1)\theta > 1$.

By Lemma 3.1(i), similarly as in (3.8),

$$\begin{aligned} \operatorname{Var}\left(\tilde{H}_{n,t}(\psi)\right) &\le n^{-1} \sum_{j,j'=1}^{n} C(t/(1 \vee |j'-j|))(1 \vee |j'-j|)^{-1-\delta} \\ &\le \sum_{j=0}^{\infty} C(t/(1 \vee j))(1 \vee j)^{-1-\delta} = \tilde{C}(t), \end{aligned}$$

where $\tilde{C}(t)$ does not depend on n and tends to 0 as $t \to \infty$. Therefore, it suffices to prove the asymptotic normality of the truncated sum $H_{n,t}(\psi)$, $\forall\ t$.

Now, for convenience, write $U_{j,s}$ for $U_{j,s}(\psi, k^*)$. Then $H_{n,t}(\psi)$ can be rewritten as

$$H_{n,t}(\psi) = n^{-1/2} \sum_{j=1}^{n} \sum_{s=0}^{t-1} U_{j+s,s} + N_n,$$

where

$$N_n = n^{-1/2} \sum_{s=0}^{t-1} \left(\sum_{j=1}^{s} U_{j,s} - \sum_{j=n+1}^{n+s} U_{j,s} \right).$$

Observe that, for every fixed $t < \infty$, $n^{1/2} N_n$ is a sum of a finite number of stationary r.v.'s whose second moment is finite by (3.6). Thus, for every fixed $t < \infty$,

$$N_n = O_P(n^{-1/2}).$$

Furthermore, for any $0 \le s \le t < \infty$,

$$\begin{aligned} U_{j+s,s} &= E\{\psi(X_{j+s}) \mid \mathscr{F}_j\} - E\{\psi(X_{j+s}) \mid \mathscr{F}_{j-1}\} \\ &\quad - \sum_{k=1}^{k^*} \frac{(-1)^k}{k!} \lambda_k \sum_{r=1}^{k} \binom{k}{r} b_s^r A_r(\zeta_j) A_{k-r}(\tilde{X}_{j+s,s}), \end{aligned}$$

and $M_j(t) = \sum_{s=0}^{t-1} U_{j+s,s}$ are square-integrable martingale difference sequences, satisfying

$$E\{U_{j+s,s} \mid \mathscr{F}_{j-1}\} = E\{M_j(t) \mid \mathscr{F}_{j-1}\} = 0,$$

which follows from the fact that $A_{k-r}(\tilde{X}_{j+s,s})$ is measurable with respect to the σ-algebra $\mathscr{F}_{j-1}$ and that $A_r(\zeta_j)$ has mean 0 and is independent of $\mathscr{F}_{j-1}$ for each j. Moreover, the above martingale difference sequences are stationary and ergodic, due to the ergodicity of the i.i.d. sequence ζ_j, $j \in \mathbf{Z}$. Therefore, the classical martingale central limit theorem applies [see, e.g., Billingsley (1968)], according to which

$$n^{-1/2} \sum_{j=1}^{n} M_j(t) \Rightarrow \mathscr{N}(0, \sigma_\psi^2(t)),$$

where, by stationarity,

$$\begin{aligned} \sigma_\psi^2(t) &= \operatorname{Var}(M_0(t)) \\ &= \sum_{s,s'=1}^{t-1} E(U_{s,s} U_{s',s'}) = \sum_{s,s'=1}^{t-1} E(U_{0,s} U_{s'-s,s'}) \\ &= \sum_{j \in \mathbf{Z}} \sum_{s,s'=1}^{t-1} E(U_{0,s} U_{j,s'}) \quad \text{by orthogonality; see Lemma 4.4 below} \\ &= \sum_{j \in \mathbf{Z}} \operatorname{Cov}(R_{0,t}(\psi; k^*), R_{j,t}(\psi; k^*)). \end{aligned}$$

Hence it easily follows that $\sigma_\psi(t) \to \sigma_\psi$, $t \to \infty$, with the limit given by (1.20).

Finally, the relation (1.28) follows from

$$\sup_{x\in\mathbf{R}} \left|F_n(x) - F(x) + f(x)\left(\bar{X}_n - m\right)\right| = o_P\left(\sqrt{\operatorname{Var}}\left(\bar{X}_n\right)\right) \tag{3.9}$$

and Lemma 4.1 below. The result (3.9) was proved under somewhat different assumptions on the distribution function of ζ_0 in Giraitis and Surgailis (1994) and Ho and Hsing (1996). Under the present setup, its proof is obtained in a similar way, using Lemma 3.1 above. This also completes the proof of Theorem 1.2. □

4. Some general results and proof of Lemma 3.1. This section contains some general results, which may be of independent interest. Lemma 3.1 is a special case of Lemma 4.3 below. Accordingly, let $\{X\}$ denote the class of all moving averages

$$X \equiv X_j = \sum_{s\geq 0} b_s \zeta_{j-s}, \qquad j \in \mathbf{Z}, \tag{4.1}$$

whose coefficients b_s, $s \geq 0$, satisfy

$$|b_s| \leq L(s)(1 \vee s)^{-(1+\theta)/2} \quad \text{and} \quad b_s \sim L(s)s^{-(1+\theta)/2}, \qquad s \to \infty,$$

where $\theta \in (0,1)$ and the slowly varying function L is fixed. For any $X \in \{X\}$ of (4.1), put

$$s_0 \equiv s_0(X) = \min\{s \geq 0;\, b_s \neq 0\}.$$

The quantities $s_0(X)$ and $s_1(X,k)$ appearing in Lemma 4.1 below are nonrandom. They depend on the underlying process only through its distribution. For later use, we note that, for any $s \geq 0$, the transformation $X \to \tilde{X}_{\bullet,s}$ maps elements of $\{X\}$ into $\{X\}$ and

$$s_0\left(\tilde{X}_{\bullet,s}\right) \geq s. \tag{4.2}$$

We also need to define the r.v.'s

$${}_sX_j = X_j - b_s\zeta_{j-s} = \sum_{i\geq 0,\, i\neq s} b_i \zeta_{j-i}, \qquad j \in \mathbf{Z},\ s \geq 0,$$

and the constants

$$b_s^{(2)} = E\tilde{X}_{j,s}^2 = \sum_{i>s} b_i^2 = O\left(L^2(s)s^{-\theta}\right). \tag{4.3}$$

Let ${}_sF(x) = P\{{}_sX_0 \leq x\}$ and ${}_sf(x) = d({}_sF(x))/dx$ denote the marginal distribution function and density of the stationary process ${}_sX_j$. The following lemma gives the behavior of the derivatives of the distribution functions F, F_s and ${}_sF$ of X_0, $X_{0,s}$ and ${}_sX_0$, respectively. It is an extension of Lemma 2.1 of Giraitis, Koul and Surgailis (1996).

LEMMA 4.1. *For any moving average $X \in \{X\}$ and any $k \geq 0$, one can find $s_1 = s_1(X,k) > 0$ such that for any $s > s_1$, the distribution functions $F(x)$, $F_s(x)$ and ${}_sF(x)$ are k times continuously differentiable and*

$$\left|F^{(k)}(x)\right| + \left|F_s^{(k)}(x)\right| + \left|{}_sF^{(k)}(x)\right| \leq C. \tag{4.4}$$

Moreover,

$$|F^{(k)}(x) - F_s^{(k)}(x)| \le Cb_s^{(2)}, \tag{4.5}$$

$$|F^{(k)}(x) - {}_sF^{(k)}(x)| \le Cb_s^2, \tag{4.6}$$

and, for any $y \in \mathbf{R}$,

$$|\Delta_y F^{(k)}(x)| + |\Delta_y F_s^{(k)}(x)| + |\Delta_y({}_sF^{(k)}(x))| \le C\,|y|, \tag{4.7}$$

$$|\Delta_y(F^{(k)}(x) - F_s^{(k)}(x))| \le Cb_s^{(2)}\,|y|, \tag{4.8}$$

$$|\Delta_y(F^{(k)}(x) - {}_sF^{(k)}(x))| \le Cb_s^2\,|y|. \tag{4.9}$$

The constant C *in* (4.4)–(4.9) *does not depend on* s, x, y.

The following lemma is an elementary result needed later.

LEMMA 4.2.

$$\begin{aligned}\Sigma(t) &\equiv \sum_{s>t} (1 \vee |s|)^{-\alpha}(1 \vee |j-s|)^{-\alpha} \\ &\le C(t/(1 \vee |j|))\begin{cases}|j|^{-2\alpha+1}, & \text{if } 1/2 < \alpha < 1,\\ |j|^{-\alpha}, & \text{if } \alpha > 1,\end{cases}\end{aligned}$$

where $C(t) \to 0$, $t \to \infty$.

Proofs of both of the above lemmas appear at the end of this section.
We now turn to the proof of Lemma 3.1.

PROOF OF LEMMA 3.1. Our proof uses induction in k. Recall the decomposition (3.4)–(3.5):

$$R_j(\psi, X_0; k) = \sum_{s \ge 0} U_{j,s}(\psi, X_0; k). \tag{4.10}$$

(We shall now indicate the dependence on the sequence X_j, $j \in \mathbf{Z}$, or its marginal X_0, too, as it will vary in the sequel.) Let

$$\langle \psi(\cdot + a), F_i \rangle := \int_{\mathbf{R}} \psi(x+a)\, dF_i(x), \qquad a \in \mathbf{R},\ i \ge 0,$$

and rewrite

$$U_{j,s} = U_{j,s}^{(0)} + U_{j,s}^{(1)}, \tag{4.11}$$

where

$$\begin{aligned}U_{j,s}^{(0)} &= \langle \psi(\cdot + \tilde{X}_{j,s-1}), F_{s-1}\rangle - \langle \psi(\cdot + \tilde{X}_{j,s}), F_s\rangle \\ &\quad - b_s\zeta_{j-s}\left\{\sum_{i=1}^{k} \frac{(-1)^i}{(i-1)!}\lambda_i A_{i-1}(\tilde{X}_{j,s})\right\},\end{aligned} \tag{4.12}$$

$$
\begin{aligned}
U_{j,s}^{(1)} &= \sum_{i=1}^{k} \frac{(-1)^{i+1}}{i!} \lambda_i \sum_{r=2}^{i} \binom{i}{r} A_{i-r}\big(\tilde{X}_{j,s}\big) b_s^r A_r(\zeta_{j-s}) \\
&= \sum_{r=2}^{k} b_s^r A_r(\zeta_{j-s}) \left\{ \sum_{i=r}^{k} \frac{(-1)^{(i+1)}}{i!} \binom{i}{r} \lambda_i A_{i-r}\big(\tilde{X}_{j,s}\big) \right\} \\
&\equiv \sum_{r=2}^{k} U_{j,s}^{(1,r)}.
\end{aligned} \tag{4.13}
$$

Note that $U_{j,s}^{(1)}$ contains the summands that decrease to 0 fast as $s \to \infty$, due to the fact that b_s^r, $s \geq 0$, are summable for $r \geq 2$.

Now let s_1 be as in Lemma 4.1 and rewrite

$$
U_{j,s}^{(0)} = \sum_{p=0}^{k+1} U_{j,s}^{(0,p)}, \tag{4.14}
$$

where

$$
\begin{aligned}
U_{j,s}^{(0,0)} &= \big\langle \psi\big(\cdot + \tilde{X}_{j,s-1}\big), F_{s-1} \big\rangle - \big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_s \big\rangle \\
&\quad + b_s \zeta_{j-s} \big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_{s-1}^{(1)} \big\rangle \mathbf{1}(s > s_1),
\end{aligned} \tag{4.15}
$$

$$
\begin{aligned}
U_{j,s}^{(0,p)} &= \frac{(-1)^{p-1}}{(p-1)!} b_s \zeta_{j-s} \Big(\lambda_p - E\big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_{s-1}^{(p)} \big\rangle \mathbf{1}(s > s_1) \Big) \\
&\quad \times A_{p-1}\big(\tilde{X}_{j,s}\big), \qquad p = 1, \dots, k,
\end{aligned} \tag{4.16}
$$

$$
\begin{aligned}
U_{j,s}^{(0,k+1)} &= -b_s \zeta_{j-s} \Bigg\{ \big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_{s-1}^{(1)} \big\rangle - \sum_{i=1}^{k} \frac{(-1)^{(i-1)}}{(i-1)!} \\
&\qquad \times A_{i-1}\big(\tilde{X}_{j,s}\big) E\big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_{s-1}^{(i)} \big\rangle \Bigg\} \mathbf{1}(s > s_1).
\end{aligned} \tag{4.17}
$$

We shall show below that the terms $U_{j,s}^{(0,p)}$, $p = 0, 1, \dots, k$, similarly to $U_{j,s}^{(1,r)}$, $r = 2, \dots, k$, are of the "summable" order $O_P(b_s^2)$, while to the main term, $U_{j,s}^{(0,k+1)}$, a suitable inductive hypothesis can be applied.

Indeed, for fixed $s \geq 0$, consider the stationary sequence $\tilde{X}_{j,s}$, $j \in \mathbf{Z}$. Let $\psi_s(x) = \langle \psi(\cdot + x), F_{s-1}^{(1)} \rangle$. Integration by parts shows that the coefficients $\lambda_i(\psi_s, \tilde{X}_{0,s})$ in the formal Appell series

$$
\psi_s\big(\tilde{X}_{0,s}\big) = \sum_{i=0}^{\infty} \frac{(-1)^i}{i!} \lambda_i\big(\psi_s, \tilde{X}_{0,s}\big) A_i\big(\tilde{X}_{0,s}\big)
$$

are given by

$$
\lambda_i\big(\psi_s, \tilde{X}_{0,s}\big) = d^i E \psi_s\big(\tilde{X}_{0,s} - x\big)/dx^i|_{x=0} = E\big\langle \psi\big(\cdot + \tilde{X}_{0,s}\big), F_{s-1}^{(i+1)} \big\rangle; \tag{4.18}
$$

cf. (1.10) and (1.11). Therefore, (4.17) can be rewritten as

$$U_{j,s}^{(0,k+1)} = -b_s \zeta_{j-s} R_j\big(\psi_s, \tilde{X}_{0,s}; k-1\big)\mathbf{1}(s > s_1). \tag{4.19}$$

We need to apply the inductive hypothesis to $R_j(\psi_s, \tilde{X}_{0,s}; k-1)$. However, as ψ_s and $\tilde{X}_{0,s}$ are different from the original ψ and X, the hypotheses and statements of Lemma 3.1 have to be correspondingly strengthened. This is done in the following lemma.

Next, let $\{\psi\}$ be the class of all functions ψ of bounded variation. The transformation $\psi \to \psi_s$ is well defined for $s > s_1$ and maps $\{\psi\}$ into itself, which follows from Lemma 4.1 and the inequality

$$|\psi_s| \le |\psi|\,|F_{s-1}^{(1)}|. \tag{4.20}$$

LEMMA 4.3. *For any processes* $X, X' \in \{X\}$, *any functions* $\psi, \psi' \in \{\psi\}$, *any integers* $0 \le t$, $1 \le k < 1/\theta$ *and any* $\delta_0 > 0$, *one can find* $0 < \delta < \delta_0$ *and constants* $C(t)$, $C < \infty$, *independent of* $j, j' \in \mathbf{Z}$, $|y| < 1$, *such that* $C(t) \to 0$, $t \to \infty$, *and such that*, *for all* $j, j' \in \mathbf{Z}$,

$$\begin{aligned} &\Big|\mathrm{Cov}\Big(\tilde{R}_{j,t}(\psi, X; k), \tilde{R}_{j',t}(\psi', X'; k)\Big)\Big| \\ &\qquad \le C\big(t/(1 \vee |j-j'|)\big)(1 \vee |j-j'|)^{-\gamma(k+1)}, \end{aligned} \tag{4.21}$$

$$\begin{aligned} &\Big|\mathrm{Cov}\Big(\Delta_y R_j(\psi, X; k), \Delta_y R_{j'}(\psi', X'; k)\Big)\Big| \\ &\qquad \le C\,|y|\,(1 \vee |j-j'|)^{-\gamma(k+1)}, \end{aligned} \tag{4.22}$$

$\gamma(k) = \gamma(k; \theta, \delta)$ *being given by* (3.7). *Moreover*, *for all* $|y| < 1$,

$$\mathrm{Var}\Big(R_j(\psi, X; k)\Big) \le C s_0^{-\gamma(k+1)}, \tag{4.23}$$

$$\mathrm{Var}\Big(\Delta_y R_j(\psi, X; k)\Big) \le C\,|y|\,s_0^{-\gamma(k+1)}. \tag{4.24}$$

Lemma 4.3 easily follows from Lemma 4.5 below. Put

$$V_j^{(i,p)} \equiv V_j^{(i,p)}(\psi, X; k) = \sum_{s \ge 0} U_{j,s}^{(i,p)}(\psi, X; k), \tag{4.25}$$

$$\tilde{V}_{j,t}^{(i,p)} = \sum_{s \ge t} U_{j,s}^{(i,p)} \tag{4.26}$$

and

$$\Delta_y V_j^{(i,p)} \equiv V_j^{(i,p)}\big(\Delta_y \psi, X; k\big) = \sum_{s \ge 0} \Delta_y U_{j,s}^{(i,p)}, \tag{4.27}$$

$i = 0, 1$, $j = 0, 1, \dots, k+1$, $y \in \mathbf{R}$, where, for $p = 0, 1$ and $p = k+1$, we put $U_{j,s}^{(1,p)} = \Delta_y U_{j,s}^{(1,p)} = 0$. Clearly,

$$R_j(\psi, X; k) = \sum_{i=0}^{1} \sum_{p=0}^{k+1} V_j^{(i,p)}(\psi, X; k), \tag{4.28}$$

and similar relations hold for $\tilde{R}_{j,t}$ and $\Delta_y R_j$, too.

We shall also need the following orthogonality relation of random variables $U_{j,s}^{(i,p)}$, which is a consequence of them being martingale difference sequences.

LEMMA 4.4. *For any* ψ, ψ', X, X' *of Lemma* 4.1, *any* $i, i' = 0, 1$, $p, p' = 0, 1, \dots, k+1$ *and any* $j, j', s, s' \in \mathbf{Z}_+$ *such that* $j' - s' \neq j - s$,

$$\operatorname{Cov}\big(U_{j,s}^{(i,p)}(\psi, X; k), U_{j',s'}^{(i',p')}(\psi', X'; k)\big) = 0.$$

LEMMA 4.5. *Under the conditions and notation of Lemma* 4.3,

$$\begin{aligned}(4.29)\qquad &\Big|\operatorname{Cov}\big(\tilde{V}_{j,t}^{(i,p)}(\psi, X; k), \tilde{V}_{j',t}^{(i',p')}(\psi', X'; k)\big)\Big| \\ &\qquad \le C\big(t/(1 \vee |j'-j|)\big)(1 \vee |j'-j|)^{-\gamma(k+1)},\end{aligned}$$

$$\begin{aligned}(4.30)\qquad &\Big|\operatorname{Cov}\big(\Delta_y V_j^{(i,p)}(\psi, X; k), \Delta_y V_{j'}^{(i',p')}(\psi', X'; k)\big)\Big| \\ &\qquad \le C|y|(1 \vee |j'-j|)^{-\gamma(k+1)}.\end{aligned}$$

Moreover,

$$(4.31)\qquad \operatorname{Var}\big(V_j^{(i,p)}(\psi, X; k)\big) \le C s_0^{-\gamma(k+1)},$$

$$(4.32)\qquad \operatorname{Var}\big(\Delta_y V_j^{(i,p)}(\psi, X; k)\big) \le C|y|\, s_0^{-\gamma(k+1)}.$$

PROOF. Note that the Appell coefficients (4.18) can be rewritten as

$$(4.33)\qquad E\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_{s-1}^{(i+1)}\Big\rangle = \big\langle \psi, {}_sF^{(i+1)}\big\rangle \equiv {}_s\lambda_i.$$

Indeed,

$$EF_{s-1}\big(x - \tilde{X}_{j,s}\big) = EP\big\{{}_sX_j \le x \mid \mathscr{F}_{j-s-1}\big\} = {}_sF(x)$$

from which (4.33) easily follows.

The proof of (4.29)–(4.32) uses induction in k. First, we prove the case $k=1$. In this case, we need to check (4.29)–(4.32) for $i = i' = 0$, $p, p' = 0, 1, 2$ only, as, by definition, $V_j^{(1,p)}(\psi, X; 1) = 0$. We have

$$\begin{aligned} U_{j,s}^{(0,0)} &= \Big\langle \psi\big(\cdot + \tilde{X}_{j,s-1}\big), F_{s-1}\Big\rangle - \Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), F_s\Big\rangle \\ &\quad + b_s\zeta_{j-s}\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle \mathbf{1}(s > s_1), \\ U_{j,s}^{(0,1)} &= b_s\zeta_{j-s}\big(\lambda_1 - {}_s\lambda_1 \mathbf{1}(s > s_1)\big), \\ U_{j,s}^{(0,2)} &= -b_s\zeta_{j-s}\Big(\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle - E\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle\Big)\mathbf{1}(s > s_1). \end{aligned}$$

By Lemma 4.4, for any i, p, I', p' and any $k = 1, 2, \dots,$

$$\begin{aligned}(4.34)\qquad \Big|E\Big(\tilde{V}_{j,t}^{(i,p)}\tilde{V}_{j',t}^{(i',p')}\Big)\Big| &= \Big|\sum_{s \ge t:\, j'-j+s \ge t} E\Big(U_{j,s}^{(i,p)} U_{j',j'-j+s}^{(i',p')}\Big)\Big| \\ &\le \sum_{s \ge t} E^{1/2}\Big(U_{j,s}^{(i,p)}\Big)^2 E^{1/2}\Big(U_{j',j'-j+s}^{(i',p')}\Big)^2.\end{aligned}$$

and

$$(4.35)\qquad \left|E\left(\Delta_y V_j^{(i,p)}\,\Delta_y V_{j'}^{(i',p')}\right)\right| \le \sum_{s\ge 0} E^{1/2}\left(\Delta_y U_{j,s}^{(i,p)}\right)^2 E^{1/2}\left(\Delta_y U_{j',j'-j+s}^{(i',p')}\right)^2.$$

We shall shortly prove that, for $p = 0, 1, 2$,

$$(4.36)\qquad E\left(U_{j,s}^{(0,p)}\right)^2 \le C\left(b_s^4 + b_s^2 b_s^{(2)}\right),$$

$$(4.37)\qquad E\left(\Delta_y U_{j,s}^{(0,p)}\right)^2 \le C\,|y|\left(b_s^4 + b_s^2 b_s^{(2)}\right).$$

Hence, as $b_s^2 = o(b_s^{(2)})$, by Lemma 4.2, we obtain

$$(4.38)\qquad \begin{aligned} \left|E\left(\tilde V_{j,t}^{(0,p)}\tilde V_{j',t}^{(0,p')}\right)\right| &\le C\sum_{s\ge t}|b_s|\left(b_s^{(2)}\right)^{1/2}|b_{j'-j+s}|\left(b_{j'-j+s}^{(2)}\right)^{1/2}\\ &\le C\sum_{s\ge t}(1\vee|s|)^{-1/2-\theta+\delta/2}(1\vee|j'-j+s|)^{-1/2-\theta+\delta/2}\\ &\le C(t/(1\vee|j'-j|))(1\vee|j'-j|)^{-\gamma(2)},\end{aligned}$$

$\gamma(2) = 2\theta - \delta$ if $0 < \theta \le 1/2$, $\delta < 2\theta$, and $\gamma(2) = 1 + \delta$ if $1/2 < \theta < 1$, $\delta < (2\theta - 1)/3$. This proves (4.29) for $k = 1$. The same argument implies (4.30)–(4.32), for example,

$$E\left(\Delta V_j^{(0,p)}\right)^2 \le C\,|y|\sum_{s\ge 0} b_s^2 b_s^{(2)} < C\sum_{s>s_0} s^{-1-2\theta+\delta} \le C\,|y|\,s_0^{-2\theta+\delta}.$$

Next, we shall prove (4.36). Using the definition (4.15), similarly to Ho and Hsing (1996), one obtains by Taylor's formula and the mean value theorem

$$\begin{aligned} U_{j,s}^{(0,0)} &= \int_{\mathbf R}\left\langle F_{s-1}\left(\cdot - \tilde X_{j,s} - b_s u\right) - F_{s-1}\left(\cdot - \tilde X_{j,s} - b_s\zeta_{j-s}\right), \psi\right\rangle dG(u)\\ &\quad - b_s\zeta_{j-s}\left\langle f_{s-1}\left(\cdot - \tilde X_{j,s}\right), \psi\right\rangle\\ &= \int_{\mathbf R} b_s(\zeta_{j-s} - u)\left\langle f_{s-1}\left(\cdot - \tilde X_{j,s} - b_s\zeta_{j-s}\right), \psi\right\rangle dG(u)\\ &\quad + \tfrac12\int_{\mathbf R} b_s^2(\zeta_{j-s} - u)^2\left\langle f_{s-1}^{(1)}\left(\cdot - \tilde X_{j,s} - b_s z(u)\right), \psi\right\rangle dG(u)\\ &\quad - b_s\zeta_{j-s}\left\langle f_{s-1}\left(\cdot - \tilde X_{j,s}\right), \psi\right\rangle\\ &= -b_s^2\zeta_{j-s}^2\left\langle f_{s-1}^{(1)}\left(\cdot - \tilde X_{j,s} - b_s z\right), \psi\right\rangle\\ &\quad + \tfrac12 b_s^2\int_{\mathbf R}(\zeta_{j-s} - u)^2\left\langle f_{s-1}^{(1)}\left(\cdot - \tilde X_{j,s} - b_s z(u)\right), \psi\right\rangle dG(u),\end{aligned}$$

where z, $z(u)$ are some (random) points. Therefore, according to (4.4),

$$(4.39)\qquad \left|U_{j,s}^{(0,0)}\right| \le Cb_s^2\left(1 + \zeta_{j-s}^2\right),$$

which proves (4.36) for $p = 0$.

In the case $p = 1$, as $|f(x) - {}_sf(x)| \le Cb_s^2$, by (4.6) and because ψ has bounded variation,

$$E\big(U_{j,s}^{(0,1)}\big)^2 = b_s^2 \langle f - {}_sf, \psi\rangle^2 \le Cb_s^4. \tag{4.40}$$

In the case $p = 2$, using (4.7) of Lemma 4.1, $|f_{s-1}(x) - f_{s-1}(x - \tilde{X}_{j,s})| \le C\,|\tilde{X}_{j,s}|$ and therefore

$$\begin{aligned}
&\Big|\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle - E\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle\Big| \\
&\quad \le \Big|\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big) - \psi, f_{s-1}\Big\rangle\Big| + E\,\Big|\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big) - \psi, f_{s-1}\Big\rangle\Big| \\
&\quad \le C\,\big(|\tilde{X}_{j,s}| + E\,|\tilde{X}_{j,s}|\big).
\end{aligned}$$

Hence

$$\begin{aligned}
E\big(U_{j,s}^{(0,2)}\big)^2 &= b_s^2 E\Big(\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle - E\Big\langle \psi\big(\cdot + \tilde{X}_{j,s}\big), f_{s-1}\Big\rangle\Big)^2 \\
&\le Cb_s^2 E\big(\tilde{X}_{j,s}^2\big) \le Cb_s^2 b_s^{(2)}.
\end{aligned} \tag{4.41}$$

Next, consider (4.37). Fix $s_1 > 1$ sufficiently large in order that the statement of Lemma 4.1 applies, and consider the cases $s > s_1$ and $s \le s_1$ separately. In the first case, similarly to (4.39),

$$\begin{aligned}
\Delta_y U_{j,s}^{(0,0)} = &-b_s^2 \zeta_{j-s}^2 \Big\langle \psi\big(\cdot + \tilde{X}_{j,s} + b_s z\big), \Delta_y f_{s-1}^{(1)}\Big\rangle \\
&+ \tfrac{1}{2} b_s^2 \int_{\mathbf{R}} \Big\langle \psi\big(\cdot + \tilde{X}_{j,s} + b_s z(u)\big), \Delta_y f_{s-1}^{(1)}\Big\rangle \big(\zeta_{j-s} - u\big)^2\, dG(u),
\end{aligned}$$

where $z \in (-\zeta_{j-s}, \zeta_{j-s})$ and $z(u) \in (\zeta_{j-s} - u, \zeta_{j-s} + u)$ are random points. Hence, again using (4.7), similar to (4.39), we obtain $|\Delta_y U_{j,s}^{(0,0)}| \le C\,|y|\, b_s^2(1 + \zeta_{j-s}^2)$, or (4.36) for $p = 0$, $s > s_1$. Cases $p = 1, 2$, $s > s_1$ can be considered exactly in the same way as (4.40) and (4.41), with $f, f_{s-1}, {}_sf$ replaced by $\Delta_y f, \Delta_y f_{s-1}, \Delta_y({}_sf)$, respectively, and using (4.7)–(4.9) of Lemma 4.1.

Now consider the case of $s \le s_1$. By definition,

$$\Delta_y U_{j,s}^{(0,0)} = E\big\{\Delta_y \psi(X_j) \mid \mathscr{F}_{j-s}\big\} - E\big\{\Delta_y \psi(X_j) \mid \mathscr{F}_{j-s-1}\big\}$$

and therefore

$$\begin{aligned}
E\big(\Delta_y U_{j,s}^{(0,0)}\big)^2 &\le 2\Big(E\Big(E\big\{\Delta_y \psi(X_j) \mid \mathscr{F}_{j-s}\big\}\Big)^2 + E\Big(E\big\{\Delta_y \psi(X_j) \mid \mathscr{F}_{j-s-1}\big\}\Big)^2\Big) \\
&\le 4E\big(\Delta_y \psi(X_j)\big)^2.
\end{aligned}$$

Write $\psi(x) = \psi_1(x) - \psi_2(x)$, where $\psi_i(x)$, $i = 1, 2$, have bounded variation and are *increasing*. Then $E(\Delta_y \psi(X_j))^2 \le 2E\{(\Delta_y \psi_1(X_j))^2 + (\Delta_y \psi_2(X_j))^2\}$, and, for $y > 0$,

$$\begin{aligned}
E\big(\Delta_y \psi_i(X_j)\big)^2 &= E\big(\psi_i(X_j + y) - \psi_i(X_j)\big)^2 \\
&\le |\psi_i|\, E\big(\psi_i(X_j + y) - \psi_i(X_j)\big) \\
&= |\psi_i| \int_{\mathbf{R}} (F(x) - F(x - y))\, d\psi_i(x) < Cy;
\end{aligned}$$

the last inequality follows from (4.7). Next,

$$E\left(\Delta_y U_{j,s}^{(0,1)}\right)^2 = b_s^2 \left\langle \Delta_y f, \psi \right\rangle^2 E\zeta_{j-s}^2 \le C\,|y|^2 \le C\,|y|,$$

according to (4.7). This proves (4.37), and the *first induction step* $k = 1$.

Induction step $k - 1 \to k$. For $(i, p) \le (0, k+1)$, $(i', p') \ne (0, k+1)$, the inequalities (4.29)–(4.32) follow similarly to the case $k = 1$. In fact, using (4.34) and (4.35) it suffices to check the bounds

$$E\left(U_{j,s}^{(i,p)}\right)^2 \le C b_s^4, \tag{4.42}$$

$$E\left(\Delta_y U_{j,s}^{(i,p)}\right)^2 \le C\,|y|\, b_s^4, \tag{4.43}$$

$(i, p) \ne (0, k+1)$.

For $(i, p) = (0,0)$, (4.42) and (4.43) were proved above. For $i = 0$, $p = 1, \dots, k$, their proof is similar to the case $k = 1$. Indeed, using the independence of ζ_{j-s} and $\tilde{X}_{j,s}$, from (4.16) we obtain

$$E\left(U_{j,s}^{(0,p)}\right)^2 = ((p-1)!)^{-2} b_s^2 E\left(A_{p-1}\left(\tilde{X}_{j,s}\right)\right)^2 \left\langle \psi, F^{(p)} - {}_sF^{(p)} \right\rangle^2 \le C b_s^4,$$

where we have used (4.6) and (4.33), similarly to (4.40). Next,

$$E\left(U_{j,s}^{(1,p)}\right)^2 = b_s^{2p} E\left(A_p(\zeta_{j-s})\right)^2 E\left(\sum_{i=p}^{k} \frac{(-1)^i}{i!} \binom{i}{p} \lambda_i A_{i-p}\left(\tilde{X}_{j-s}\right)\right)^2 \le C b_s^4,$$

$p = 2, \dots, k$, proving (4.42).

In the case $(i, p) = (0, k+1)$, according to (4.19),

$$E\left(U_{j,s}^{(0,k+1)}\right)^2 = b_s^2 E\zeta_{j-s}^2 E\left(R_j\left(\psi_s, \tilde{X}_{0,s}; k-1\right)\right)^2, \tag{4.44}$$

$$E\left(\Delta_y U_{j,s}^{(0,k+1)}\right)^2 = b_s^2 E\zeta_{j-s}^2 E\left(\Delta_y R_j\left(\psi_s, \tilde{X}_{0,s}; k-1\right)\right)^2. \tag{4.45}$$

Here,

$$E\left(R_j\left(\psi_s, \tilde{X}_{0,s}; k-1\right)\right)^2 \le C s^{-k\theta+\delta}, \tag{4.46}$$

$$E\left(\Delta_y R_j\left(\psi_s, \tilde{X}_{0,s}; k-1\right)\right)^2 \le C\,|y|\, s^{-k\theta+\delta}, \tag{4.47}$$

which follow from (4.2) and the inductive hypothesis [Lemma 4.3, (4.23) and (4.24), for $k = k-1$]. Therefore, by Lemmas 4.2 and 4.3,

$$\begin{aligned}\left| E\left(\tilde{V}_{j,t}^{(0,k+1)} \tilde{V}_{j',t}^{(0,k+1)}\right)\right| &\le C \sum_{s \ge t} |b_s b_{j'-j+s}|\, s^{-(k\theta-\delta)/2} (1 \vee |j'-j+s|)^{-(k\theta-\delta)/2} \\ &\le C(t/(1 \vee |j'-j|))(1 \vee |j'-j|)^{-\gamma(k+1)}.\end{aligned}$$

Using (4.34), (4.35) and (4.42)–(4.47), the proof of the rest of the relations (4.30)–(4.32) is analogous, thereby completing the proof of Lemma 4.5. □

PROOF OF LEMMA 4.1. The claims (4.7)–(4.9) follow from (4.4)–(4.6), respectively; for example,

$$\left|\Delta_y\left(F^{(k)}(x) - F_s^{(k)}(x)\right)\right| = \left|\int_x^{x+y} \left(F^{(k+1)}(z) - F_s^{(k+1)}(z)\right) dz\right| < Cb_s^{(2)}\,|y|.$$

Consider the characteristic functions

$$\hat{f}(u) = E\exp(iuX_0) = \prod_{j\geq 0} \phi(ub_j), \tag{4.48}$$

$$\hat{f}_s(u) = E\exp(iuX_{0,s}) = \prod_{0\leq j\leq s} \phi(ub_j), \tag{4.49}$$

$${}_s\hat{f}(u) = E\exp(iu\,{}_sX_0) = \prod_{j\geq 0:\, j\neq s} \phi(ub_j), \tag{4.50}$$

where $\phi(u) = E\exp(iu\zeta_0)$ satisfies (1.4). Clearly, (4.4) follows from

$$\left|\hat{f}(u)\right| + \left|\hat{f}_s(u)\right| + \left|{}_s\hat{f}(u)\right| \leq C/(1+|u|)^{k+1}, \tag{4.51}$$

$s > s_1$. For any $k > 0$, one can find an integer $s_1 > 0$ such that the number $|J|$ of elements in the set $J = \{j = 0, 1, \ldots, s_1 : b_j \leq 0\}$ is larger than $(k+1)/\delta$, that is $k + 1 < \delta\,|J|$. Then, by (4.48)–(4.50),

$$\max\left(\left|\hat{f}(u)\right|, \left|\hat{f}_s(u)\right|, \left|{}_s\hat{f}(u)\right|\right) \leq C \prod_J \left(1 + |ub_j|\right)^{-\delta}.$$

For any $b \neq 0$, there is a constant $C = C(b, \delta)$ such that $(1 + |bu|)^{-\delta} < C(1 + |u|)^{-\delta}$. Therefore, by the argument above,

$$\max\left(\left|\hat{f}(u)\right|, \left|\hat{f}_s(u)\right|, \left|{}_s\hat{f}(u)\right|\right) \leq C(1+|u|)^{-\delta|J|} \leq C(1+|u|)^{-k-1},$$

proving (4.4).

Consider (4.5) and (4.6) for $k \geq 1$. Choose s_1 so that $\prod_{0\leq j\leq s} |\phi(ub_j)| < C(1 + |u|)^{-k-4}$, $s > s_1$, where C is independent of u, s. Then

$$\begin{aligned}\left|f^{(k)}(x) - f_s^{(k)}(x)\right| &\leq (2\pi)^{-1}\int_{\mathbf{R}} |u|^k \prod_{0\leq j\leq s}\left|\phi(ub_j)\right| \left|\prod_{j>s}\phi(ub_j) - 1\right| du\\ &\leq C\sum_{j>s}\int_{\mathbf{R}} |u|^k (1+|u|)^{-k-4}\left|\phi(ub_j) - 1\right| du \qquad (4.52)\\ &\leq C\sum_{j>s} b_s^2 \int_{\mathbf{R}} |u|^{k+2}(1+|u|)^{-k-4}\,du = Cb_s^{(2)},\end{aligned}$$

where we used the inequality $|\phi(u) - 1| \leq \frac{1}{2}|u|^2$. In a similar way,

$$\begin{aligned}\left|f^{(k)}(x) - {}_sf^{(k)}(x)\right| &\leq C\int_{\mathbf{R}} |u|^k\,|\phi(ub_s) - 1| \prod_{j\neq s}\left|\phi(ub_j)\right| du\\ &\leq Cb_s^2\int_{\mathbf{R}} |u|^{k+2}(1+|u|)^{-k-4}\,du < Cb_s^2.\end{aligned}$$

It remains to verify (4.5) and (4.6) for $k = 0$. For this, it suffices to prove the bounds

$$\left| f^{(k)}(x) - f_s^{(k)}(x) \right| \le C b_s^{(2)} (1 + |x|)^{-k-2}, \tag{4.53}$$

$$\left| f^{(k)}(x) - {}_s f^{(k)}(x) \right| \le C b_s^2 (1 + |x|)^{-k-2} \tag{4.54}$$

for $k = 1$, as they imply (4.53) and (4.54) for $k = 0$ and then, similarly, (4.5) and (4.6) for $k = 0$. Proceeding as in Giraitis, Koul and Surgailis (1996), Lemma 1(ii), (4.53) and (4.54) follow from

$$\left| \left(u \left(\hat{f}(u) - \hat{f}_s(u) \right) \right)^{(r)} \right| \le C b_s^2 (1 + |u|)^{-2},$$

$$\left| \left(u \left(\hat{f}(u) - {}_s\hat{f}(u) \right) \right)^{(r)} \right| \le C b_s^2 (1 + |u|)^{-2}$$

or

$$\left| \hat{f}^{(r)}(u) - \hat{f}_s^{(r)}(u) \right| \le C b_s^{(2)} (1 + |u|)^{-3}, \tag{4.55}$$

$$\left| \hat{f}^{(r)}(u) - {}_s\hat{f}^{(r)}(u) \right| \le C b_s^2 (1 + |u|)^{-3}, \qquad r = 0, 1, 2, 3. \tag{4.56}$$

Consider (4.55) for $r = 2$, for example, as the remaining inequalities can be similarly verified. By differentiating (4.48) and (4.49) with respect to u, one obtains

$$\begin{aligned}
\hat{f}^{(2)}(u) &= \sum_{0 \le j} \phi^{(2)}(u b_j) b_j^2 \Phi_j(u) \\
&\quad + 2 \sum_{0 \le j_1 < j_2} \phi^{(1)}(u b_{j_1}) \phi^{(1)}(u b_{j_2}) b_{j_1} b_{j_2} \Phi_{j_1, j_2}(u), \\
\hat{f}_s^{(2)}(u) &= \sum_{0 \le j \le s} \phi^{(2)}(u b_j) b_j^2 \Phi_{j;s}(u) \\
&\quad + 2 \sum_{0 \le j_1 < j_s \le s} \phi^{(1)}(u b_{j_1}) \phi^{(1)}(u b_{j_2}) b_{j_1} b_{j_2} \Phi_{j_1, j_2; s}(u),
\end{aligned}$$

where

$$\begin{aligned}
\Phi_j(u) &= \prod_{j' \ge 0 : j' \ne j} \phi(u b_{j'}), \\
\Phi_{j;s}(u) &= \prod_{0 \le j' \le s : j' \ne j} \phi(u b_{j'}), \\
\Phi_{j_1, j_2}(u) &= \prod_{j' \ge 0 : j' \ne j_1, j_2} \phi(u b_{j'}), \\
\Phi_{j_1, j_2; s} &= \prod_{0 \le j' \le s : j' \ne j_1, j_2} \phi(u b_{j'}).
\end{aligned}$$

We have

$$\left| \hat{f}^{(2)}(u) - \hat{f}_s^{(2)}(u) \right| \le \Sigma_1 + \Sigma_2 + 2\Sigma_3 + 2\Sigma_4,$$

where

$$\Sigma_1 = \sum_{0 \le j \le s} \left|\phi^{(2)}(ub_j)\right| b_j^2 \left|\Phi_j(u) - \Phi_{j;s}(u)\right|,$$

$$\Sigma_2 = \sum_{j > s} \left|\phi^{(2)}(ub_j)\right| b_j^2 \left|\Phi_j(u)\right|,$$

$$\Sigma_3 = \sum_{0 \le j_1 < j_2 \le s} \left|\phi^{(1)}(ub_{j_1})\phi^{(1)}(ub_{j_2}) b_{j_1} b_{j_2}\right| \left|\Phi_{j_1, j_2; s}(u) - \Phi_{j_1, j_2}(u)\right|,$$

$$\Sigma_4 = \sum_{0 \le j_1,\, s < j_2} \left|\phi^{(1)}(ub_{j_1})\phi^{(1)}(ub_{j_2}) b_{j_1} b_{j_2}\right| \left|\Phi_{j_1, j_2}(u)\right|.$$

We need to show

$$\Sigma_i \le C b_s^{(2)} (1 + |u|)^{-3}, \qquad s > s_1,\ i = 1, \dots, 4. \tag{4.57}$$

Consider Σ_1. As in (4.52),

$$\begin{aligned}\left|\Phi_j(u) - \Phi_{j;s}(u)\right| &\le \left|\Phi_{j;s}(u)\right| \left|\prod_{j' > s} \phi(ub_{j'}) - 1\right| \\ &\le C(1 + |u|)^{-5} \sum_{j' > s} u^2 b_{j'}^2 \le C(1 + |u|)^{-3} b_s^{(2)},\end{aligned}$$

provided $s > s_1$ is large enough, proving (4.57) for $i = 1$. The case $i = 2$ follows from the bound $|\Phi_j(u)| < C(1 + |u|)^{-3}$.

To estimate Σ_3, use $|\phi^{(1)}(ub_j)| \le |ub_j|$ and the bound

$$\begin{aligned}\left|\Phi_{j_1, j_2; s}(u) - \Phi_{j_1, j_2}(u)\right| &\le \left|\Phi_{j_1, j_2; s}(u)\right| \left|\prod_{j' > s} \phi(ub_{j'}) - 1\right| \\ &\le C(1 + |u|)^{-5} \sum_{j' > s} u^2 b_{j'}^2,\end{aligned}$$

as in the case $i = 1$. Finally, in the case $i = 4$, one uses

$$\sum_{j_2 > s} \left|\phi^{(1)}(ub_{j_2})\right| |b_{j_2}| \le |u| \sum_{j_2 > s} b_{j_2}^2 \le |u|\, b_s^{(2)}$$

together with $|\Phi_{j_1, j_2}(u)| \le C(1 + |u|)^{-5}$. This completes the proof of Lemma 4.1. □

PROOF OF LEMMA 4.2. It suffices to consider $j \ge 1$. Let $\alpha > 1$. Let $t \le j$. Then

$$\begin{aligned}\Sigma(t) &\le \sum_{s \in \mathbf{Z}} = \sum_{s \le |j|/2} + \sum_{s > |j|/2} \\ &\le 2(1 \vee |j|/2)^{-\alpha} \sum_{s \in \mathbf{Z}} (1 \vee |s|)^{-\alpha} \le C(1 \vee |j|)^{-\alpha}.\end{aligned}$$

Next, let $t > j$, $K \equiv t/j > 1$, $\alpha > 1$. Then

$$\Sigma(t) \le C_1(1 \vee t)^{-\alpha} = C_1(1 \vee j)^{-\alpha}(1 \vee K)^{-\alpha} = C(t/(1 \vee j))(1 \vee j)^{-\alpha},$$

$$C(t) \equiv C_1 t^{-\alpha} \to 0, \qquad t \to \infty.$$

Now consider the case $1/2 < \alpha < 1$, $t \le j$. Then

$$\Sigma(t) \le C_1 \int_{\mathbf{R}} |s|^{-\alpha}\, |j - s|^{-\alpha}\, ds = Cj^{-\alpha}.$$

Finally, let $1/2 < \alpha < 1$, $t > j$, $K \equiv t/j$. Then

$$\Sigma(t) \le C_1 \int_t^{\infty} (s-j)^{-2\alpha}\, ds = C_1 j^{-2\alpha+1} \int_K^{\infty} (u-1)^{-2\alpha}\, du = C(K) j^{-2\alpha+1},$$

$$C(K) = C_2 K^{-2\alpha+1} \to 0,$$

as $K \to \infty$. □

REFERENCES

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

BERAN, J. (1991). *M*-estimators of location for data with slowly decaying correlations. *J. Amer. Statist. Assoc.* **86** 704–708.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (1996). Asymptotic normality of regression estimators with long memory errors. *Statist. Probab. Lett.* **29** 317–335.

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. (1994). Asymptotics of the empirical process of a long memory linear sequence. Preprint.

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.

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

KOUL, H. L. and MUKHERJEE, K. (1993). Asymptotics of *R*-, MD- and LAD-estimators in linear regression models with long range dependent errors. *Probab. Theory Related Fields* **95** 535–553.

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

TAQQU, M. S. (1978). A representation of self-similar process. *Stochastic Process. Appl.* **7** 55–64.

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

DEPARTMENT OF STATISTICS AND PROBABILITY
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824-1027
E-MAIL: koul@assist.stt.msu.edu

INSTITUTE OF MATHEMATICS AND INFORMATICS
2600 VILNIUS
LITHUANIA