Statistica Sinica **12**(2002), 1171-1190

ON FUNCTIONALS OF LINEAR PROCESSES WITH ESTIMATED PARAMETERS

Hwai-Chung Ho

Academia Sinica

Abstract: For a stationary linear process with independent and identically distributed innovations, the paper addresses asymptotic properties of partial sums of nonlinear functional applied to the process when an unknown parameter is estimated. General representations are established under the condition that the innovation coefficients are either summable or regularly varying with index in $(-1, -1/2)$. The usefulness of the representations is demonstrated through the derivation of limiting distributions for several common examples such as kurtosis, the sign test, and the absolute deviation from the mean.

Key words and phrases: Absolute deviation, empirical distribution, kurtosis, linear process, long memory, M-estimator, short memory, sign test.

1. Introduction

Let $\{X_t, t \geq 1\}$ be a linear process defined by $X_t = \mu + \sum_{i=1}^{\infty} a_i \varepsilon_{t-i}$, where the ε_i are independent and identically distributed random variables having mean 0 and at least finite second moment, and where the coefficients a_i are either summable or regularly varying $i^{-\beta} L(i)$ with index $-\beta$ in $(-1, -1/2)$. The process $\{X_t\}$ is said to be short-memory in the former case and long-memory in the latter; in both cases $\{X_t\}$ may violate the strong mixing property (Andrew (1984); Rosenblatt (1961)) and may not belong to the class of weakly dependent stationary sequences (Bradley (1986)). The present paper investigates the asymptotic behavior of partial sums of non-linear functions applied to X_t when there is another estimated parameter involved. For a large class of Borel functions $K(x, y)$, including polynomials and indicator functions $I(x \leq y)$, we develop general representations for $T_n(\theta_n) \equiv n^{-1} \sum_{t=1}^{n} K(X_t, \theta_n)$, where θ_n is an estimator for some unknown parameter θ of X_t, and then use the representations to derive limiting laws. Two typical examples of $T_n(\cdot)$ are the sample variance $n^{-1}\sum_{t=1}^{n}(X_t-\mu_n)^2$ and the empirical distribution evaluated at the sample mean, $F_n(\mu_n)$, where $F_n(\cdot)=n^{-1}\sum_{t=1}^{n}I(X_t\le\cdot)$ and $\mu_n=n^{-1}\sum_{t=1}^{n}X_t$. The quantity $F_n(\mu_n)$ is often used as a modified sign test statistic in testing the hypothesis that the distribution is symmetric about its unknown mean (Gastwirth (1971)).

Let $F(\cdot)$ be the marginal distribution function of the stationary linear process $\{X_t, t\ge 1\}$. For the function $K(\cdot\,,\cdot)$, define $T_n(\cdot)=n^{-1}\sum_{t=1}^{n}K(X_t,\cdot)$ and $K_\infty(x,y)=\int K(x+u,y)dF(u)$. When X_t is short-memory, the representation we obtain is similar to that of i.i.d. or weakly dependent sequences (Ralescu and Puri (1984), Sen (1972)), which basically leads to the Central Limit Theorem. For long-memory X_t, however, the representation has richer structures than in the short-memory case and the class of limiting distributions derived from the representation is also broader. It turns out that the form of $K_\infty(\cdot\,,\cdot)$ plays a major role in studying $T_n(\theta_n)$ and determining what type of limit theorem emerges.

We use the two examples above to illustrate some interesting features of the representation in the long-memory setting. Choose $K(x,y)=(x-y)^2$. Denote by $T_n(\mu_n)$ the sample variance and $K_\infty^{(i,j)}(\cdot\,,\cdot)$ the (i,j)-th partial derivative of $K_\infty(\cdot\,,\cdot)$. We let $\bar{Y}_{n,1}=\mu_n-\mu$. Since $K_\infty^{(0,1)}(0,\mu)=0$ and $K_\infty^{(0,2)}(0,\mu)=-K_\infty^{(1,1)}(0,\mu)=2$, we can rewrite the sample variance as

$$T_n(\mu_n)=T_n(\mu)+\sum_{j=1}^{2}K_\infty^{(0,j)}(0,\mu)\frac{(\mu_n-\mu)^j}{j!}+K_\infty^{(1,1)}(0,\mu)\bar{Y}_{n,1}(\mu_n-\mu). \quad (1.1)$$

In the third term on the right-hand side of (1.1), we deliberately use $\bar{Y}_{n,1}(\mu_n-\mu)$ instead of $(\mu_n-\mu)^2$ to emphasis that $\bar{Y}_{n,1}$ is a quantity which arises from expanding $T_n(\mu)$ and need not in general be equal to the deviation of the parameter estimation from its true value, $\mu_n-\mu$ in the sample variance case. Unlike weakly dependent or short-memory sequences, both the second and third term on the right-hand side of (1.1) contribute to the limiting distribution as does the first term $T_n(\mu)$. Each of the three terms plays a different role in the asymptotic behavior of $T_n(\mu_n)$. In fact, the message of relation (1.1) is that $T_n(\mu_n)$ consists of three parts: the functional $T_n(\mu)$ with the true parameter μ, the Taylor expansion of $K_\infty(0,\mu_n)$ at μ that concerns the estimation deviation $\mu_n-\mu$, and the cross-product term $\bar{Y}_{n,1}(\mu_n-\mu)$ describing the interrelation between the preceding two. It is this observation that leads to the representation (3.1) in Section

3. The cross-product term in (1.1) is of crucial importance in the presence of long memory. To shed more light on this we turn to the example of the empirical distribution at the sample mean, $F_n(\mu_n)$. Under the assumption of weak dependence, e.g., strong mixing, one has the standard form of representation

$$F_n(\mu_n) = F_n(\mu) + f(\mu)(\mu_n - \mu) + R_n \tag{1.2}$$

with an almost sure bound specified for the remainder term R_n (Ralescu and Puri (1984), Sen (1972)) and where $f(\cdot)$ is the probability density function of X_t. The situation is more subtle if the underlying sequence $\{X_i\}$ is of long memory. As shown in Ho and Hsing (1996), $F_n(\mu) = F(\mu) - f(\mu)(\mu_n - \mu) + R'_n$ where $R'_n = o((\mathrm{Var}\,(\mu_n))^{1/2})$ with probability one. When this expression for $F_n(\mu)$ is plugged into (1.2), there is cancellation in the linear term $(\mu_n - \mu)$. Thus the remainder terms R_n and R'_n need further expansion. Relation (1.1) hints that higher order expansion of $F(\mu_n)$ and $F_n(\mu)$ about μ only provides partial information about the asymptotic behavior of $F_n(\mu_n)$. The cross-product term has to be taken into consideration in order to draw a complete picture of $F_n(\mu_n)$. An immediate implication of the cancellation effect is that the testing power of $F_n(\mu_n)$ dominates that of $F_n(\mu)$ (μ known), which strikingly contrasts with the traditional weak dependence cases. Further details will be elaborated in Example B of Section 4.

The rest of the paper is summarized as follows. Section 2 presents notation and technical conditions. In Section 3, the representations for the long-memory and short-memory cases are discussed in Theorems 3.1 and 3.2, respectively. In Section 4, limiting distributions are derived for three specific examples of kurtosis, sign test, and the absolute deviation from the mean to demonstrate the usefulness of the representations. In the last example, the case where the robust M-estimator is used for the location estimates is also discussed. Proofs are given in the Appendix.

2. Notation and Preliminaries

In the sequel we assume $E\varepsilon_1^2 = 1$. Define $Y_{n,0} = n$ and, for $r \geq 1$,

$$Y_{n,r} = \sum_{t=1}^{n} \sum_{1 \leq j_1 < \cdots < j_r < \infty} \prod_{s=1}^{r} a_{j_s} \varepsilon_{t-j_s} \text{ and } \bar{Y}_{n,r} = n^{-1} Y_{n,r}.$$

If $a_i = i^{-\beta}L(i)$ for some $\beta \in (1/2, 1)$ and r is any positive integer such that $r(2\beta - 1) < 1$, then the process $\{\sum_{1\le j_1<\cdots<j_r<\infty}\prod_{s=1}^{r} a_{j_s}\varepsilon_{t-j_s} : t \ge 1\}$ also has long memory, and

$$\sigma_{n,r}^2 \equiv \mathrm{Var}\,(Y_{n,r}) \approx n^{2-r(2\beta-1)}L^{2r}(n). \tag{2.1}$$

(The notation $\approx$ signifies asymptotic proportionality.) It is also known that if $r(2\beta-1) < 1$ and $E\varepsilon_1^{2r} < \infty$, then $n^{r(\beta-1/2)-1}L^{-r}(n)Y_{n,r}$ converges as $n \to \infty$ in distribution to Z_r which can be represented by the multiple Wiener-Ito integral

$$Z_r = \int_{-\infty<u_1<\cdots<u_r<1}\int \Big\{\int_0^1 \prod_{j=1}^{r}[(v-u_j)^+]^{-\beta}dv\Big\}dB(u_1)\cdots dB(u_r), \tag{2.2}$$

with B denoting standard Brownian motion (Surgalis (1983)). On the other hand, if $r(2\beta - 1) > 1$ then $\sigma_{n,r}^2 \approx n$ and $n^{-1/2}Y_{n,r}$ obeys a central limit theorem (Giraitis (1985) and Ho and Hsing (1997)). Define $X_{t,0} = \mu$ and $X_{t,j} = \mu + \sum_{1\le i\le j} a_i\varepsilon_{t-i}$, and $\tilde{X}_{t,j} = X_t - X_{t,j}$, $1 \le j \le \infty$. Here $X_{t,\infty} = X_t$ and $\tilde{X}_{t,\infty} = 0$. Let F_j, $\tilde{F}_j$ and G_j be the distribution functions of $X_{1,j}$, $\tilde{X}_{1,j}$ and $a_j\varepsilon_1$, respectively. For $j \ge 0$, define $K_j(x,y) = \int K(x+u,y)dF_j(u)$, and $K_\infty^{(i,j)}(x,y) = \frac{d^{(i,j)}}{dx^i dy^j}\int K(x+u,y)dF(u)$ whenever they are well-defined. Note that $K_0(\tilde{X}_{t,0},\cdot) = K(X_t,\cdot)$ and $K_\infty^{(0,0)}(0,\cdot) = EK(X_t,\cdot)$.

3. Representations and Limit Laws

Let θ_n be a consistent estimate of some unknown parameter θ associated with X_t. We establish for both long-memory and short-memory cases that, for some $J \ge 1$,

$$\begin{aligned} T_n(\theta_n) = T_n(\theta) &+ \sum_{i=1}^{J} K_\infty^{(0,i)}(0,\theta)\frac{(\theta_n-\theta)^i}{i!} \\ &+ \sum_{j=1}^{J-1}\bar{Y}_{n,j}\left[\sum_{j'=1}^{J-j} K_\infty^{(j,j')}(0,\theta)\frac{(\theta_n-\theta)^{j'}}{j'!}\right] + R_{n,J}, \end{aligned} \tag{3.1}$$

and provide an upper bound (in probability) for $|R_{n,J}|$ (see Theorems 3.1 and 3.2). Note that (1.1) is a special case of (3.1) with $J = 2$ and, because $K(x,y) = (x-y)^2$, $R_{n,2} = 0$. To make use of (3.1), it is necessary that a bound for $|\theta_n - \theta|$ be specified. For this purpose, we assume

(**A1**) $|\theta_n - \theta| = O_p(\sqrt{\mathrm{Var}\,(\bar{Y}_{n,1})})$ and θ belongs to a compact set of $\mathcal{R}$.

Remark 1. When the linear process $X_t = \sum_{i=0}^{\infty} a_i\varepsilon_{t-i}$ is short-memory, i.e., $\sum |a_i| < \infty$, (**A1**) holds for most estimators in practice since $\mathrm{Var}\,(\bar{Y}_{n,1}) = O(1/n)$. For long-memory $\{X_t\}$, it is well known that the variances of many common estimators, such as maximum likelihood estimators, moment estimators and some robust estimators (see Beran (1994)), are dominated by that of $\bar{Y}_{n,1}$, which is asymptotically proportional to $n^{-(2\beta-1)}L^2(n)$ (see (2.1)).

The next assumption provides, in addition to some regularity conditions, two types of functions the paper focuses on, indicator functions and functions that behave like polynomials after being smoothed.

(**A2**) $E\varepsilon_1^8 < \infty$, $EK^2(X_1, \theta) < \infty$, and $K(x, y)$ satisfies one of the following two conditions.

1. $K(x, y)$ is the indicator function $I(x \le y)$. For some positive integer J the distribution function of ε_1 is $J+3$ times differentiable with bounded, continuous and integrable derivatives.
2. For a positive integer J, the partial derivatives $K_1^{(t,s)}(x, y)$ of $K_1(x, y)$ of order $0 \le t+s \le J+3$ are continuous on $\mathcal{R} \times (\theta-\varepsilon, \theta+\varepsilon)$ for some $\varepsilon > 0$, and $\sum_{p=0}^{J} |K_\infty^{(p+1,0)}(0, y)| \ne 0$ for all y in $(\theta-\varepsilon, \theta+\varepsilon)$. For all $0 \le t+s \le J+3$, $|K_1^{(t,s)}(x, y)| \le U(x, y)$ where $U(x, y)$ is a two-variable polynomial of degree M and $E\varepsilon_1^{\max\{8,4M\}} < \infty$. For certain $\alpha \in [1, 2]$ and a positive function $V(x, y)$ such that $EV^2(X_1, \theta) < \infty$,
$$|K(x, y_1) - K(x, y_2)| \le |y_1 - y_2|^{\alpha/2} V(x, \theta) \tag{3.3}$$
for all (x, y_i) in $\mathcal{R} \times (\theta-\varepsilon, \theta+\varepsilon), i = 1, 2$.

Remark 2. (1) The two types of functions in **(A2)** cover most of the interesting cases in the literature. Later in Section 4 we consider three examples: a polynomial $a(x-y)^4 + b(x-y)^2 + c$ of degree 4; the absolute value function $|x-y|$; the indicator function $I(x \le y)$. The first two are of type 2 in **(A2)**. (2) The functions described in **(A2)** all satisfy the following technical properties. Let $\mathcal{B}$ be the set of all sequences $\tilde{b} = \{b_1, b_2, \ldots, \}$, where $b_i = a_i$ except for finitely many i's in which b_i equals one of the three values: $0, a_{i+1}$, or $a_i + a_{i+1}$. For each $\tilde{b} \in \mathcal{B}$, define $X_{\tilde{b}} = \sum_{i=1}^{\infty} b_i\varepsilon_i$. For $0 \le t+s \le J+3$ and $\lambda \ge 0$, define

$K_{1,\lambda}^{(t,s)}(x, y_1, y_2) = \sup_{|u|\le\lambda} |K_1^{(t,s)}(x+u, y_1) - K_1^{(t,s)}(x+u, y_2)|$, y_1 and y_2 in I. Then

$$E[K_{1,\lambda}^{(t,s)}(X_{\tilde{b}}, y_1, y_2)]^4 \le C|y_1 - y_2|^4. \tag{3.4}$$

Moreover, for all $0 \le t+s \le J+3$ and y in $I = (\theta-\varepsilon, \theta+\varepsilon)$, we have from (3.4) that

$$K_{j+1}^{(t,s)}(x, y) = \int K_j^{(t,s)}(x+u, y)dG_{j+1}(u). \tag{3.5}$$

This can be proved by the argument used in Lemma 2.1 of Ho and Hsing (1997). For all y_1 and y_2 in I,

$$E[K(X_1, y_1) - K(X_1, y_2)]^2 \le C \cdot |y_1 - y_2|^\alpha, \tag{3.6}$$

where the α values for the three examples mentioned above are 2, 2, and 1, respectively. In addition, the random variable $\sup_{y\in I} |\sum_{t=1}^n [K(X_t, y) - K(X_t, \theta)]|$ can be approximated by $\max_{y\in\mathcal{P}_n} |\sum_{t=1}^n [K(X_t, y) - K(X_t, \theta)]|$ for a set of properly designed partitions $\{\mathcal{P}_n\}$ of I. This is guaranteed for functions of type 2 by (3.3) and for the indicator functions by the simple fact that $I(x \le y_1) - I(x \le y_2) = I(y_2 < x \le y_1), y_2 < y_1$. (3) From $K_\infty(x, y) = \int K_1(x+u, y)d\tilde{F}_1(u)$ and $K_1(\cdot\,,\cdot)$ being smooth by (3.5), it follows that $K_\infty(\cdot\,,\cdot)$ has continuous partial derivatives of order $J+3$ in $\mathcal{R} \times I$. Thus $K_\infty^{(t,s)}(0, y)$ is well-defined for $1 \le t+s) \le J+3$. (4) A trivial extension of the functions in **(A2)** is their linear combinations.

We now formulate our two main theorems as follows.

Theorem 3.1. *Let $a_i = i^{-\beta}L(i)$ with $\beta \in (1/2, 1)$ and $L^*(n) = \max\{L(n), 1\}$. Assume* (A1) *and* (A2). *Then* (3.1) *holds and the remainder term $R_{n,J}$ satisfies, as $n \to \infty$,*

(i) $|R_{n,J}| = o_p(n^{-J(\beta-1/2)-\tau})$ *for any* $\tau < \beta - 1/2$ *if* $2\beta - 1 \le (J+1)^{-1}$, *or* $\tau < -J(\beta-1/2) + \min\{(J+1)(\beta-1/2), 1/2 + \alpha(\beta-1/2)/2\}$ *if* $(J+1)^{-1} < 2\beta - 1 < J^{-1}$, *or*

(ii) $|R_{n,J}| = o_p(n^{-(1/2+\alpha(\beta-1/2)/2)+\delta})$ *for any* $\delta > 0$ *if* $J^{-1} \le 2\beta - 1$.

Theorem 3.2. *Suppose $\sum_i |a_i| < \infty$ and $E\varepsilon_1^4 < \infty$. Assume* (A1) *and* (A2) *with $J = 1$. Then* (3.1) *holds with $J = 1$ and $R_{n,1} = o_p(n^{-(1/2+\alpha/4)+\delta})$ for any $\delta > 0$.*

Remark 3. (1) Suppose in Theorem 3.2 that $K(x, y) = I(x \le y)$, $\theta = \mu$ and $\theta_n = n^{-1}\sum_{t=1}^n X_t$. Let $F(\cdot)$ be the distribution function of X_t with density $f(\cdot)$. Define $U_t = I(X_t \le \mu) + f(\mu)(X_t - \mu)$. It follows by Theorem 4.1 of Ho and Hsing (1997) that $\sqrt{n}(T_n(\theta_n) - F(\theta))$ is asymptotically $N(0, c^2)$ with $c^2 = \sum_{m=-\infty}^{\infty} r(m)$, where $r(m) = \operatorname{cov}(U_t, U_{t+m})$. As pointed out in Remark 2-(1), α is 1. Hence the main factor appearing in $R_{n,1}$ is $n^{-3/4}$, which is consistent with the previous findings in the literature (see, for example, Sen (1972) and Ralescu and Puri (1984)). The novelty of Theorem 3.2 lies in that the linearity of X_t is fully exploited so that the representation and its resulting central limit theorem can be achieved without imposing conventional types of mixing conditions. With the same $K(\cdot,\cdot)$, θ and θ_n, a broader class of limiting laws will be obtained later in Example B of Section 4 for long-memory X_t as an application of Theorem 3.1. (2) The expression of $R_{n,J}$ in Theorem 3.1 suggests the following. When $\{X_t\}$ is long-memory, the influence the regularity index α of (3.6) has on the rate of $R_{n,J}$ decreases as the dependence increases (i.e., β is closer to $1/2$). In fact, the rate becomes independent of α for two out of the three cases in Theorem 3.1−(i), $2\beta - 1 \le (J+1)^{-1}$ or $2\beta - 1 > (J+1)^{-1}$ and $(J+1)(\beta - 1/2) - 1/2 \le \alpha(\beta - 1/2)/2$. If $2\beta - 1 > J^{-1}$ as in Theorem 3.1−(ii), the influence of α on the error rate is always present. This interesting phenomenon is supported by the following observation made in Ho and Hsing (1997). There are two possibilities for the asymptotic distribution of the partial sums $T_n(y,\theta) \equiv \sum_{t=1}^{n}\{K(X_t, y) - K(X_t, \theta)\}$, central limit theorem ($\sqrt{n}$ rate) and the non-central limit theorem (non-$\sqrt{n}$ rate). In the former case, the limiting variance involves the quantity $E\{K(X_1, y) - K(X_1, \theta)\}^2 = C|y-\theta|^\alpha$ and thus the regularity index α has influence on $T_n(y,\theta)$ as y approaches θ; in the latter case, $T_n(y,\theta)$ is asymptotically equivalent to $\{K_\infty^{(J,0)}(0,y) - K_\infty^{(J,0)}(0,\theta)\}\bar{Y}_{n,J}$ for a certain positive integer J, and the smooth function $K_\infty^{(J,0)}(0,y)$ instead of the function $K(\cdot,\cdot)$ itself determines the behavior of $T_n(y,\theta)$ when y is very close to θ. One can of course find some extreme examples without the above property. Theorems 3.1 and 3.2 of the paper will collapse if $K_\infty(\cdot,\cdot)$ has some irregularity properties such as discontinuity in the second coordinate at θ. In that case, there is probably not much one can say about the asymptotics of $T_n(\theta_n)$.

The bounds that we obtain for $|R_{n,J}|$ in Theorems 3.1 and 3.2 are far from sharp but they serve our main goal sufficiently well as to understand the structure of the representation and to use the representation to obtain limiting distributions.

4. Examples

When $\{X_t\}$ is long-memory, there are three steps in using Theorem 3.1 to derive the asymptotic distribution of $T_n(\theta_n) - EK(X_t, \theta)$. We first replace $T_n(\theta)$ in (3.1) by its martingale expansion (Theorem 3.1, Ho and Hsing (1997); see (A.1) in the Appendix) and obtain

$$T_n(\theta_n) = \sum_{r=0}^{J} K_\infty^{(r,0)}(0,\theta)\bar{Y}_{n,r} + \sum_{i=1}^{J} K_\infty^{(0,i)}(0,\theta)\frac{(\theta_n-\theta)^i}{i!} + \sum_{j=1}^{J-1}\bar{Y}_{n,j}\left[\sum_{j'=1}^{J-j} K_\infty^{(j,j')}(0,\theta)\frac{(\theta_n-\theta)^{j'}}{j'!}\right] + S_{n,J}(\theta) + R_{n,J}, \quad (4.1)$$

where $S_{n,J}(\theta) = T_n(\theta) - \sum_{r=0}^{J} K_\infty^{(r,0)}(0,\theta)\bar{Y}_{n,r}$ and satisfies $\mathrm{Var}\,(S_{n,j}(\theta)) \le C(n^{-1} \vee n^{-(J+1)(2\beta-1)+\zeta})$ for some universal constant C and any $\zeta > 0$. Second, we need to figure out the joint limiting distribution of $(\bar{Y}_{n,j}, (\theta_n - \theta)^i)$ for all positive integers $1 \le i,\ j \le J$. Note here that the two components, $\bar{Y}_{n,j}$ and $(\theta_n - \theta)^i$, may marginally obey different types of limit theorems. The third step is to compute the values of $K_\infty^{(i,j)}(0,\theta)$ for all $1 \le i, j \le J$ and, bearing in mind the possibility of cancellation of terms on the right-hand side of (4.1), then decide the normalization factor for $T_n(\theta_n)$. In the second step, the asymptotic distribution of $(\bar{Y}_{n,j}, (\theta_n - \theta)^i)$ can be characterized for two important cases: either θ_n can be written as a smooth function (such as a polynomial) plus a negligible residual term $H(\bar{Y}_{n,1}, \ldots, \bar{Y}_{n,j}) + o_p(1)$ in terms of $\bar{Y}_{n,j}$'s, or it can be sufficiently well approximated by an estimator having such an expression. Examples that have been widely discussed in the literature are sample average estimators $\theta_n = n^{-1}\sum_{t=1}^{n} G(X_t)$ (Ho and Hsing (1997)) and U-statistics (Ho and Hsing (1996)) for the former case, and some robust estimators for the latter (Koul and Surgalis (1997)). There are circumstances in which precise knowledge of the asymptotic distribution of $(\bar{Y}_{n,j}, (\theta_n - \theta)^i)$ is not absolutely necessary. One such example of interest is the maximal likelihood estimator θ_n of parameters θ (such as the memory parameter or the innovation variance) of the spectral density of $\{X_t\}$ (Fox and Taqqu (1987), and Giraitis and Surgalis (1990)). It is known that the convergence rate of the MLE θ_n is $\theta_n - \theta = O_p(n^{-1/2})$. Then all terms on the righthand side of (4.1) involving the power of $(\theta_n - \theta)$ become negligible if the normalization factor of $T_n(\theta)$ is $\{n^{\beta-1/2}L^{-1}(n)\}^J$ with $J(2\beta - 1) < 1$, meaning that the deviation $\theta_n - \theta$ plays no role in the limiting distribution of $T_n(\theta)$. If $J(2\beta-1) > 1$, then we still have to deal with $(\bar{Y}_{n,j}, (\theta_n - \theta)^i)$ to fulfill the second step. Finding the asymptotic joint distribution of $\bar{Y}_{n,j}$ and the MLE θ_n is itself a challenging problem of independent interest. Its statistical relevance however is not clear especially under our setting based on nonlinear functions. For the third step, it can be seen that there are numerous cases involved in determining precisely which terms contribute to the limiting distribution. This and the discussion above on the second step both indicate that in order to apply Theorem 3.1 to deriving limits, the task would very much be of case-by-case nature. The present paper thus will not seek for a general rule derived from (4.1) to classify the limiting distributions. We instead demonstrate the usefulness of the representation (4.1) through a more direct approach by developing the asymptotic distributions for three very practical examples. The first example, Example A, treats sample kurtosis as a representative case of general moment estimators. The second example, Example B, deals with a modified sign test a common nonparametric testing method. The case of estimated scale parameter is discussed in a remark to Example B. In Examples A and B only sample averages are considered for the estimated parameters. Example C focuses on estimation of the absolute deviation from the mean and allows for the robust M-estimation of location.

The following lemma collects some known results that are needed for later discussion.

Lemma 4.1. *Assume* $a_i = i^{-\beta}L(i)$, $E\varepsilon_1^{(2p)\vee 8} < \infty$, *and* $p^{-1} > (2\beta-1) > (p+1)^{-1}$. *Set* $a_{n,j} = n^{j(\beta-1/2)}L^{-j}(n)$.

(i) *Let* $j_1, \ldots, j_m$ *be distinctive positive integers that are not greater than* p. *Then* $(a_{n,j_1}\bar{Y}_{n,j_1}, \ldots, a_{n,j_m}\bar{Y}_{n,j_m}) \xrightarrow{d} (Z_{j_1}, \ldots, Z_{j_m})$, *where the* Z_i*'s are defined in* (2.2).

(ii) *Let* $K(\cdot\,,\cdot)$ *belong to the class of functions defined in* **(A2)**. *Then* $S_{n,q}(\theta) = O_p(n^{-q(\beta-1/2)}L^q(n))$ *for* $1 \le q \le p$, *and* $n^{1/2}S_{n,p}(\theta) \xrightarrow{d} N(0, w^2)$, *where* $w^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{cov}(X_t^{(p)},\ X_{t+\ell}^{(p)})$ *with* $X_t^{(p)} = K(X_t, \theta) - EK(X_t\,, \theta) - \sum_{1\le j_1 < \cdots < j_r < \infty} \prod_{s=1}^{p} a_{j_s}\varepsilon_{t-j_s}$.

Part (i) of Lemma 4.1 is contained in Avram and Taqqu (1987), and part (ii) in Theorems 3.1 and 3.2 of Ho and Hsing (1997). Note here that in order to have the central limit theorem in Lemma 4.1-(ii) one needs the ℓ-approximation property, $E\{K(X_1,\theta) - K(X_{1,\ell},\theta)\}^2 \to 0$ as $\ell \to \infty$, which is satisfied by the functions in **(A2)**.

In Examples A, B and C, we let $X_t = \mu + \sum_{i=1}^{\infty} a_i \varepsilon_{t-i}$ and the coefficient $a_i = i^{-\beta} L(i)$ with $1/2 < \beta < 1$, and assume that $E(\varepsilon_t^8) < \infty$ and the probability density functions $f(\cdot)$ and $g(\cdot)$ of X_t and ε_t, respectively, are sufficiently smooth so that (**A2**) is assured for some J. Recall that $\mu_n = n^{-1}\sum_{t=1}^{n} X_t$ and $\mu_n - \mu = \bar{Y}_{n,1}$.

Example A. Denote the i-th central moment $\mu^{(i)}$ and their usual estimates by $\mu^{(i)} = E(X_1-\mu)^i, i \geq 2$, and $\mu_n^{(i)} = n^{-1}\sum_{t=1}^{n}(X_t-\mu_n)^i$, respectively. We now apply Theorem 3.1 to derive the limiting distribution of the kurtosis estimator $\hat{\kappa} \equiv \mu_n^{(4)}/(\mu_n^{(2)})^2$. Write

$$\hat{\kappa}-\kappa = n^{-1}\sum_{t=1}^{n}\left\{\left[(X_t-\mu_n)^4-\mu^{(4)}\right]/(\mu^{(2)})^2-(\mu^{(4)}/(\mu^{(2)})^3)\left[(X_t-\mu_n)^2-\mu^{(2)}\right]\right\}$$
$$+O_p\left(|\mu_n^{(2)}-\mu^{(2)}|\left[|\mu_n^{(2)}-\mu^{(2)}|+|\mu_n^{(4)}-\mu^{(4)}|\right]\right).$$

Thus the functional $K(\cdot\,,\cdot)$ to be considered is, by neglecting $O_p(|\mu_n^{(2)}-\mu^{(2)}|(|\mu_n^{(2)}-\mu^{(2)}|+|\mu_n^{(4)}-\mu^{(4)}|)$,

$$K(x,y) = [(x-y)^4-\mu^{(4)}]/(\mu^{(2)})^2 - (\mu^{(4)}/(\mu^{(2)})^3)[(x-y)^2-\mu^{(2)}]$$

with the sample mean μ_n being the estimated parameter, i.e., the representation (4.1) is applied to $T_n(\mu_n) = n^{-1}\sum_{t=1}^{n} K(X_t,\mu_n)$ instead of to $\hat{\kappa}-\kappa$ directly. By straight forward computations, $K_\infty^{(1,0)}(0,\mu) = -K_\infty^{(0,1)}(0,\mu) = 4\mu^{(3)}/(\mu^{(2)})^2$ and $K_\infty^{(0,2)}(0,\mu) = K_\infty^{(2,0)}(0,\mu) = -K_\infty^{(1,1)}(0,\mu) = 12(\mu^{(2)})^{-1} - 2\mu^{(4)}(\mu^{(2)})^{-3}$. Then (4.1) (with $J=2$) is, after canceling the linear terms,

$$\begin{aligned}
&T_n(\mu_n)\\
&= S_{n,1}(\mu) + \left\{\left[12(\mu^{(2)})^{-1} - 2\mu^{(4)}(\mu^{(2)})^{-3}\right](\mu_n-\mu)^2/2\right\}\\
&\quad - \left\{12(\mu^{(2)})^{-1} - 2\mu^{(4)}(\mu^{(2)})^{-3}\right\}\bar{Y}_{n,1}(\mu_n-\mu) + R_{n,2}\\
&= \left\{2(\mu^{(2)})^{-1}(6-\mu^{(4)}/(\mu_n^{(2)})^2)\bar{Y}_{n,2}\right\} + \left\{(\mu^{(2)})^{-1}(6-\mu^{(4)}/(\mu_n^{(2)})^2)(\mu_n-\mu)^2\right\}\\
&\quad - \left\{2\mu^{(2)})^{-1}(6-\mu^{(4)}/(\mu_n^{(2)})^2)\bar{Y}_{n,1}\cdot(\mu_n-\mu)\right\} + S_{n,2}(\mu) + R_{n,2}. \qquad (4.2)
\end{aligned}$$

Suppose $2(2\beta-1) < 1$. In view of the right side of the second identity in (4.2), it follows by Theorem 3.1−(i) and Lemma 4.1 that

$$n^{2\beta-1}L^{-2}(n)\,(\hat{\kappa}-\kappa) \xrightarrow{d} (\mu^{(2)})^{-1}\left(6-\mu^{(4)}/(\mu^{(2)})^2\right)\left(2Z_2-Z_1^2\right),$$

since $n^{2\beta-1}L^{-2}(n)(S_{n,2}(\mu)+R_{n,2})=o_p(1)$ and $n^{2\beta-1}L^{-2}(n)\bar{Y}_{n,2}$ and $n^{2\beta-1}L^{-2}(n)$ $\times\bar{Y}_{n,1}(\mu_n-\mu)$ converge in distribution to Z_2 and Z_1^2, respectively. If $2(2\beta-1)>1$, then a central limit theorem holds. In fact, on the right side of the first identity in (4.2), only $S_{n,1}(\mu)=T_n(\mu)-4(\mu^{(3)}/(\mu^{(2)})^2)\bar{Y}_{n,1}$ contributes to the limit as all the terms $R_{n,2}, (\mu_n-\mu)^2$ and $\bar{Y}_{n,}(\mu_n-\mu)$ are negligible. Therefore, Theorem 3.1−(ii) and Lemma 4.1−(ii) imply $n^{1/2}\,(\hat{\kappa}-\kappa) = n^{-1/2}\sum_{t=1}^n([(X_t-\mu)^4-\mu^{(4)}]/(\mu^{(2)})^2$ $-(\mu^{(4)}/(\mu_n^{(2)})^3)[(X_t-\mu)^2-\mu^{(2)}].-4(\mu^{(3)}/(\mu^{(2)})^2)(X_t-\mu))+o_p(1)$, which converges in distribution to $N(0,w_1^2)$, where $w_1^2=\sum_{j=-\infty}^{\infty}\mathrm{cov}(X_t',X_{t+j}')$ with

$$X_t' \equiv (X_t-\mu)^4/(\mu^{(2)})^2-(\mu^{(4)}/(\mu_n^{(2)})^3)[(X_t-\mu)^2-\mu^{(2)}]-4(\mu^{(3)}/(\mu^{(2)})^2)(X_t-\mu).$$

Example B. We focus on the modified sign test statistics described in the Introduction. The functional is $K(x,y)=I(x\le y)$ and the estimated parameter is the sample mean μ_n. Assume that $f'(\mu)=0$ and $f''(\mu)\neq 0$. Following Example A, we first obtain $K_\infty^{(0,1)}(0,\mu) = -K_\infty^{(1,0)}(0,\mu) = f(\mu)$, $K_\infty^{(i,j)}(0,\mu) = f'(\mu) = 0$ if $i+j=2$, and $K_\infty^{(0,3)}(0,\mu) = -K_\infty^{(1,2)}(0,\mu) = K_\infty^{(2,1)}(0,\mu) = f''(\mu)$. From (4.1) with $J=3$, we have

$$\begin{aligned} F_n(\mu_n)-F(\mu) &= f''(\mu)\{\frac{-(\mu_n-\mu)^3}{3}-\bar{Y}_{n,3}+\bar{Y}_{n,2}(\mu_n-\mu)\}+S_{n,3}(\mu)+R_{n,3}\\ &= S_{n,2}(\mu)-f''(\mu)(\mu_n-\mu)^3/3+f''(\mu)\bar{Y}_{n,2}(\mu_n-\mu)+R_{n,3}\end{aligned}$$

As a result of Theorem 3.1 and Lemma 4.1,

$$\begin{aligned} &n^{3(\beta-1/2)}L^{-3}(n)\{F_n(\mu_n)-F(\mu)\}\\ &= n^{3(\beta-1/2)}L^{-3}(n)f''(\mu)\{\frac{-(\mu_n-\mu)^3}{3}-\bar{Y}_{n,3}+\bar{Y}_{n,2}(\mu_n-\mu)\}+o_p(1)\\ &\quad\xrightarrow{d}\ f''(\mu)(-Z_1^3/3-Z_3+Z_1Z_2)\end{aligned}$$

if $3(2\beta-1)<1$, and $n^{1/2}\{F_n(\mu_n)-F(\mu)\} = n^{1/2}S_{n,2}(\mu)+o_p(1) \xrightarrow{d} N(0,w_2^2)$ if $3(2\beta-1)>1$, where $w_2^2=\sum_{j=-\infty}^{\infty}\mathrm{cov}(X_t'',X_{t+j}'')$ with $X_t''=I(X_t\le\mu)+f(\mu)(X_t-\mu)$. The purpose of assuming $f'(\mu)=0$ above is merely to include the Gaussian case. When $f'(\mu)\neq 0$, a similar procedure also applies and the limiting distribution could of course be different. If $\mu_n = \mu$, the norming factor for $F_n(\mu) - F(\mu)$ is $n^{\beta-1/2}L^{-1}(n)$ instead of $n^{3(\beta-1/2)}L^{-3}(n)$ or $n^{1/2}$ (Corollary 3.3, Ho and Hsing (1996)), implying that the testing power of $F_n(\mu_n)$ dominates that of $F_n(\mu)$.

Remark 4. From (4.1) and the two previous examples, it is clear by now that if the sample mean μ_n is the estimated parameter involved in the functional $K(\cdot\,,\cdot)$ and $K_\infty(x,\mu) = G(x-\mu)$ holds for some smooth $G(\cdot)$ such that $G'(-\mu) \neq 0$, then $K_\infty^{(1,0)}(0,\mu) = -K_\infty^{(0,1)}(0,\mu)$ causes cancellation in the linear terms. We now present a case where the linear terms will not be cancelled and the asymptotic normality with non-root n normalization can be achieved. We modify the functional considered in Example B by changing the sample mean to a scale parameter estimate. Specifically, we consider $K(X_t,\theta_n) = I(\frac{X_t}{\theta_n} \leq c)$ where $\theta_n = \{n^{-1}\sum_{t=1}^n (X_t-\mu_n)^2\}^{1/2}$. Similar to Example B, $K_\infty(x,\theta) = F(c\theta - x)$, $K_\infty^{(1,0)}(0,\theta) = -f(c\theta)$ and $K_\infty^{(0,1)}(0,\theta) = cf(c\theta)$. By (4.1) (with $J=1$) and Theorem 3.1, $T_n(\theta_n) \equiv n^{-1}\sum_{t=1}^n K(X_t,\theta_n) = n^{-1}\sum_{t=1}^n K(X_t,\theta) + cf(c\theta)(\theta_n - \theta) + R_{n,1} = -f(c\theta)\bar{Y}_{n,1} + o_p(n^{-(\beta-1/2)}L(n))$, since both $n^{-1}\sum_{t=1}^n K(X_t,\theta) + f(c\theta)\bar{Y}_{n,1}$ and $(\theta_n - \theta)$ are of order $O_p(n^{-(2\beta-1)}L^2(n))$. Hence, by Lemma 4.1−(i), $n^{\beta-1/2}L^{-1}(n)\,(T_n(\theta_n) - F(c\theta)) \xrightarrow{d} -f(c\theta)\cdot Z_1$.

Example C. We investigate asymptotic distributions of the sample absolute deviation $n^{-1}\sum_{t=1}^n |X_t - \theta_n|$. We first consider the case where the robust M-estimator is adopted to play the role of the location estimate θ_n. To define the M-estimator, let ψ be a real-valued function of bounded variation such that $\psi(x) = -\psi(-x)$ and $\lambda(x) \equiv E\psi(X_1 - x)$ is smooth, $\lambda(\mu) = 0$. The M-estimator θ_n of μ corresponding to ψ is $\theta_n = \operatorname{argmin}\{|\sum_{t=1}^n \psi(X_t - x)|\}$. Assume that the probability density function f is symmetric about its mean μ and is sufficiently smooth, so that $f^{(k)}(y+\mu) = (-1)^k f^{(k)}(-y+\mu)$ holds for a suitable positive integer k.

In order to describe the asymptotic behavior of $\theta_n - \mu$, we need a result obtained by Koul and Surgalis. Suppose $(k^*+1)^{-1} < 2\beta - 1 < (k^*)^{-1}$ and let $A_j(x)$ be the j-th Appell polynomial and $Q_j(x)$ be a the polynomial of degree j defined by their recursive relation (see (1.8) and (1.13) of Koul Surgalis (1997), respectively). Define

$$V_t = \psi(X_t - \mu) - \sum_{j=1}^{k^*} \frac{(-1)^j}{j!} A_j(X_t - \mu).$$

Koul and Surgalis ((1997) Theorem 1.1) show that under appropriate conditions on the marginal distribution of innovation ε_1,

$$\theta_n - \mu_n = \sum_{j=2}^{k^*} n^{-j(\beta-1/2)} L^j(n) Q_{n,j} + n^{-1/2} W_n, \tag{4.3}$$

where $Q_{n,j} \xrightarrow{d} Q_j(Z_1,\ldots,Z_j)$ and $W_n \xrightarrow{d} N(0,w^2)$, with $w^2 = \sum_{j=-\infty}^{\infty} \mathrm{cov}(V_t, V_{t+j})$. The functional $K(\cdot,\cdot)$ corresponding to the sample absolute deviation is $K(X_t,\theta) = |X_t - \theta|$. Assume $K_\infty(x,y) \equiv \int |u+x-y| f(u)du$ is smooth and $K_\infty^{(i,j)}(x,y) = \int |u+x-y|(-1)^i f^{(i+j)}(u)du$ holds. It is straight forward that $K_\infty^{(0,1)}(0,\mu) = K_\infty^{(1,0)}(0,\mu) = 0$ and $K_\infty^{(2,0)}(0,\mu) = K_\infty^{(0,2)}(x,y) = -K_\infty^{(1,1)}(0,\mu) = 2\int_0^\infty |u-\mu| f^{(2)}(u)du$. Therefore,

$$\begin{aligned}
&T_n(\theta_n)\\
&= T_n(\mu) + (K_\infty^{(0,2)}(0,\mu)/2)\Big\{(\theta_n-\mu_n)^2 + (\mu_n-\mu)^2 - 2(\theta_n-\mu_n)(\mu_n-\mu)\Big\}\\
&\quad + K_\infty^{(1,1)}(0,\mu)\bar{Y}_{n,1}\{(\theta_n-\mu_n)+(\mu_n-\mu)\} + R_{n,2}\\
&= \Big\{K_\infty^{(2,0)}(0,\mu)\bar{Y}_{n,2} + [(K_\infty^{(0,2)}(0,\mu)/2)(\mu_n-\mu)^2 + K_\infty^{(1,1)}(0,\mu)\bar{Y}_{n,1}(\mu_n-\mu)]\Big\}\\
&\quad + \Big\{(K_\infty^{(0,2)}(0,\mu)/2)[(\theta_n-\mu_n)^2 - 2(\theta_n-\mu_n)(\mu_n-\mu)] + K_\infty^{(1,1)}(0,\mu)\bar{Y}_{n,1}(\theta_n-\mu_n)\Big\}\\
&\quad + S_{n,2}(\mu) + R_{n,2}.
\end{aligned} \tag{4.4}$$

Suppose $2(2\beta-1) < 1$. We concentrate on the second identity of (4.4). From Lemma 4.1 and (4.3), it follows that only the first term on the right side account for the limiting distribution,

$$n^{2\beta-1}L^{-2}(n)[T_n(\theta_n) - E|X_1-\mu|)] \;\xrightarrow{d}\; K_\infty^{(0,2)}(0,\mu)(Z_2 - Z_1^2/2).$$

If $2(2\beta-1) > 1$, by examining the first identity of (4.4), Lemma 4.1 and (4.3) imply that $n^{1/2}[T_n(\theta_n) - E|X_1-\mu|] \;\xrightarrow{d}\; N(0,w_2^2)$, where $w_2^2 = \sum_{j=-\infty}^{\infty} \mathrm{cov}(|X_t - \mu|, |X_{t+j}-\mu|)$. If the sample mean μ_n is used for the estimated location θ_n, the same limit distributions will be obtained. In other words, when estimating the absolute deviation from the mean it makes no differences whether the location is estimated by an M-estimator or the sample mean. These two approaches are equally efficient in the sense of the first order approximation.

Appendix

The principal idea of the proof is to use $\{\bar{Y}_{n,r}, r \geq 0\}$ to expand $T_n(y)$ as

$$\begin{aligned} T_n(y) &= n^{-1}\sum_{t=1}^{n} K(X_t, y) \\ &= n^{-1}\sum_{t=1}^{n}\sum_{p=0}^{J} K_\infty^{(p,0)}(0,y) \sum_{1\leq j_1<\cdots<j_p<\infty}\prod_{s=1}^{p} a_{j_s}\varepsilon_{t-j_s} + S_{n,J}(y) \\ &= \sum_{p=0}^{J} K_\infty^{(p,0)}(0,y)\bar{Y}_{n,p} + S_{n,J}(y), \end{aligned} \tag{A.1}$$

so that the magnitude of $|S_{n,J}(y)|$ can be estimated effectively by using arguments of Ho and Hsing (1997) with the help of the technical properties listed in **Remark 2**-(2). If X_t is standard Gaussian then for each y,

$$K(X_t, y) = \sum_{r=0}^{\infty} \frac{h_r(y)}{r!} H_r(X_t),$$

where $K_\infty^{(r,0)}(0,\cdot) = EK(X_t,\cdot)H_r(X_t)$ and $H_r(\cdot)$ is the r-th Hermite polynomial (Remark 2 in Ho and Hsing (1997)). Under the circumstances that X_t has long memory, the random processes $\{\sum_{1\leq j_1<\cdots<j_p<\infty}\prod_{s=1}^{r} a_{j_s}\varepsilon_{t-j_s}\}$ and $\{H_p(X_t)\}$ can be regarded as equivalent in the sense that their covariance functions are both asymptotically $O(k^{-p(2\beta-1)}\ L^{2p}(k))$, and the two partial sums $\sum_{t=1}^{n} H_p(X_t)$ and $Y_{n,p}$ converge to the same limiting distribution with the same norming factor up to a positive constant (Avram and Taqqu (1987), and Ho and Hsing (1997)). Therefore it is natural to interpret the martingale expansion in the second identity of (A.1) as the *discrete version of Wiener-Ito's expansion* of $K(X_t,\cdot)$, which provides a useful tool for dealing with partial sums $\sum_{t=1}^{n} K(X_t,\cdot)$ when X_t is not Gaussian (for Wiener-Ito expansion of functionals of Gaussian sequences, see Major (1981)). Recall some notation and definitions. Let $X_{t,0} = \mu$ and, $1 \leq j \leq \infty$, $X_{t,j} = \mu + \sum_{1\leq i\leq j} a_i\varepsilon_{t-i}$, $\tilde{X}_{t,j} = X_t - X_{t,j}$, and $K_j(x,y) = \int K(x+u,y)dF_j(u)$, where F_j is the distribution of $X_{1,j}$. When $j=\infty$, $X_{t,\infty} = X_t$, $\tilde{X}_{t,\infty} = 0$, and

$K_\infty(x,y) = EK(x+X_1,y)$. Take

$$\bar{Y}_{n,p} = n^{-1}\sum_{t=1}^{n}\sum_{1\le j_1<\cdots<j_p<\infty}\prod_{s=1}^{p} a_{j_s}\varepsilon_{t-j_s},$$

$$T_n(y) = n^{-1}\sum_{t=1}^{n} K(X_t,y) = \sum_{p=0}^{J} K_\infty^{(p,0)}(0,y)\bar{Y}_{n,p} + S_{n,J}(y),$$

where $S_{n,J}(y) = T_n(y) - \sum_{p=0}^{J} K_\infty^{(p,0)}(0,y)\bar{Y}_{n,p}$. Based on the orthogonal expansion

$$K(X_t,y) - K_\infty(0,y) = \sum_{j=1}^{\infty}\{K_{j-1}(\tilde{X}_{t,j-1},y) - K_j(\tilde{X}_{t,j},y)\},$$

we can write (cf. Ho and Hsing (1997), pp.1644-1646)

$$S_{n,J}(y) = Z_{n,1}^{(1)}(y) + \sum_{p=1}^{J-1}\{Z_{n,1}^{(p+1)}(y) - Z_{n,1}^{(p)}(y)\} + Z_{n,2}^{(J)}(y) + Z_{n,3}(y) + Z_{n,4}^{(J)}(y), \quad \text{(A.2)}$$

where one has

$$nZ_{n,1}^{(1)}(y) = \sum_{t=1}^{n}\sum_{j=2}^{\infty}[K_{j-1}(\tilde{X}_{t,j-1},y) - K_j(\tilde{X}_{t,j},y) - a_j\varepsilon_{t-j}K_{j-1}^{(1,0)}(\tilde{X}_{t,j},y)],$$

$$n\{Z_{n,1}^{(p+1)}(y) - Z_{n,1}^{(p)}(y)\} = \sum_{t=1}^{n}\sum_{2\le j_1\le\cdots\le j_{p+1}<\infty}(\prod_{s=1}^{p} a_{j_s}\varepsilon_{t-j_s})\times[K_{j_{p+1}-1}^{(p,0)}(\tilde{X}_{t,j_{p+1}-1},y) - $$
$$-K_{j_{p+1}}^{(p,0)}(\tilde{X}_{t,j_{p+1}},y) - a_{j_{p+1}}\varepsilon_{t-j_{p+1}}K_{j_{p+1}}^{(p+1,0)}(\tilde{X}_{t,j_{p+1}},y)],$$

$$nZ_{n,2}^{(J)}(y) = \sum_{t=1}^{n}\sum_{2\le j_1<\cdots<j_J<\infty}(\prod_{s=1}^{J} a_{j_s}\varepsilon_{t-j_s})\{K_{j_J}^{(J,0)}(\tilde{X}_{t,j_J},y) - K_\infty^{(J,0)}(0,y)\}, \quad \text{(A.3)}$$

$$nZ_{n,3}(y) = \sum_{t=1}^{n}\{K(X_t,y) - K_1(\tilde{X}_{t,1},y)\},$$

$$nZ_{n,4}^{(J)}(y) = -\sum_{p=1}^{J} K_\infty^{(p,0)}(0,y)\sum_{t=1}^{n}\sum_{1=j_1<j_2<\cdots<j_p<\infty+1}(\prod_{s=1}^{p} a_{j_s}\varepsilon_{t-j_s}).$$

On the righthand side of (A.2) only $Z_{n,3}$ contains the original function $K(\cdot,\cdot)$ which may not be smooth, and the functions involved in the remaining terms are all smooth.

Sketch proof of Theorem 3.1. Define

$$G_n=\{T_n(\theta_n)-K_\infty(0,\theta_n)\}-\{T_n(\theta)-K_\infty(0,\theta)\}-\sum_{j=1}^{J-1}\bar{Y}_{n,j}\Big\{\sum_{j'=1}^{J-j}K_\infty^{(j,j')}(0,\theta)\frac{(\theta_n-\theta)^{j'}}{j'!}\Big\}, \tag{A.4}$$

$$G_n(y,\theta)=\{T_n(y)-K_\infty(0,y)\}-\{T_n(\theta)-K_\infty(0,\theta)\}-\sum_{j=1}^{J-1}\bar{Y}_{n,j}\Big\{\sum_{j'=1}^{J-j}K_\infty^{(j,j')}(0,\theta)\frac{(y-\theta)^{j'}}{j'!}\Big\}.$$

In view of (A.1), we furthermore express $G_n(y,\theta)$ as

$$\begin{aligned} G_n(y,\theta) &= \sum_{j=1}^{J-1}\bar{Y}_{n,j}K_\infty^{(j,J-j+1)}(0,\theta)(y^*-\theta)^{J-j+1}/(J-j+1)! \\ &\quad +\bar{Y}_{n,J}\{K_\infty^{(J,0)}(0,y)-K_\infty^{(J,0)}(0,\theta)\}+\{S_{n,J}(y)-S_{n,J}(\theta)\} \\ &\equiv A_n(y,\theta)+B_n(y,\theta)+C_n(y,\theta) \end{aligned} \tag{A.5}$$

where $|y^*-\theta|\le|y-\theta|$. Define $I_n=[\theta-D_n,\theta+D_n]$ with a positive sequence $D_n=\{E\bar{Y}_{n,1}^2\}^{1/2}$. Clearly, $G_n\le\sup_{y\in I_n}|G_n(y,\theta)|$ conditionally on $\theta_n\in I_n$. The major part of the proof is to bound $\sup_{y\in I_n}|G_n(y,\theta)|$. We first concentrate on type 2 functions. Our plan is to bound $\sup_{y\in I_n}|C_n(y,\theta)|$ for (i) and (ii), and then to handle the two parts separately to bound $\sup_{y\in I_n}|A_n(y,\theta)|$ and $\sup_{y\in I_n}|B_n(y,\theta)|$. For the first step, we combine the results of Theorem 3.1 of Ho and Hsing(1997) and the "chaining argument" (see Dehling and Taqqu (1989) and Ho and Hsing (1996)) and sketch the proof as follows. The main part of the chaining argument is to create a dominating measure on the interval I_n and then, based upon this measure, to construct a family of appropriate partitions of the interval I_n. For any function $h(\cdot)$ on $\mathcal{R}$, let $\tilde{h}(x,y)=h(y)-h(x)$. Define

$$\Lambda(u)=\sum_{p=0}^{J}\int_{\theta-D_n}^{u}|K_\infty^{(p+1,0)}(0,v)|dv,\ \ u\in I_n.$$

For $m=1,\ldots,M$, let $u_i(m)=\inf\{u\in I_n:\Lambda(u)\ge\Lambda(\theta+D_n)i/2^m\}$, $m=0,1,\ldots,2^m$. For $y\in I_n$, let $m(y)$ be the integer between 0 to 2^m such that $u_{m(y)}(m)\le y<u_{m(y)+1}(m)$. Thus,

$$S_{n,j}(y)-S_{n,j}(\theta)=\sum_{m=0}^{M-1}\tilde{S}_{n,j}(u_{m(y)}(m),u_{(m+1)(y)}(m+1))+\tilde{S}_{n,j}(u_{m(y)}(y),y).$$

Furthermore, by (3.3),

$$\begin{aligned}
&P(\sup_{y\in I_n}|S_{n,j}(y)-S_{n,j}(\theta)|>b)\\
&\le P(\sup_{y\in I_n}|\sum_{m=0}^{M-1}\tilde S_{n,j}(u_{m(y)}(m),u_{(m+1)(y)}(m+1))|>b/2)\\
&\quad +P(\sup_{y\in I_n}C(u_{M(y)+1}(M)-u_{M(y)}(M))^{\alpha/2}[n^{-1}\sum_{t=1}^{n}V(X_t,\theta)]>b/4)\\
&\quad +P(\sup_{y\in I_n}\sum_{p=0}^{J}|K_\infty^{(p,0)}(0,y)-K_\infty^{(p,0)}(0,u_{M(y)}(M))||\bar Y_{n,p}|>b/4)\\
&\quad \equiv P_{n,1}(b)+P_{n,2}(b)+P_{n,3}(b).
\end{aligned}$$

Note that $|u_{M(y)+1}(M)-u_{M(y)}(M)|\le C\cdot D_n 2^{-M}$ since $\Lambda'(y)=\sum_{p=0}^{J}|K_\infty^{(p+1,0)}(0,y)|$ is bounded away from zero on a small enough neighborhood I_n of θ. Then, by Chebyshev's inequality, $P_{n,2}(b)+P_{n,3}(b)\le C\cdot(b^{-1}+b^{-2})\cdot 2^{-M\alpha/2}$ Similarly, $P_{n,1}(b)\le\sum_{m=0}^{M-1}b^{-2}(m+3)^4\sum_{i=0}^{2^{m+1}-1}\mathrm{Var}\,(\tilde S_{n,j}(u_i(m),u_i(m+1)))$. From expression (A.2), it follows that for each $m=0,1,\ldots,M-1$

$$\begin{aligned}
&\mathrm{Var}\,(\tilde S_{n,j}(u_i(m),u_i(m+1)))\\
&\le C\cdot[\mathrm{Var}\,(\tilde Z_{n,1}^{(1)}(u_i(m),u_i(m+1)))\\
&+\sum_{p=1}^{J-1}\mathrm{Var}\,(\tilde Z_{n,1}^{(p+1)}(u_i(m),u_i(m+1))-\tilde Z_{n,1}^{(p)}(u_i(m),u_i(m+1)))\\
&+\mathrm{Var}\,(\tilde Z_{n,2}^{(J)}(u_i(m),u_i(m+1))+\mathrm{Var}\,(\tilde Z_{n,3}(u_i(m),u_i(m+1)))\\
&+\mathrm{Var}\,(\tilde Z_{n,4}^{J}(u_i(m),u_i(m+1)))]\\
&\equiv V_{n,m,i,1}+V'_{n,m,i,1}+V_{n,m,i,2}+V_{n,m,i,3}+V_{n,m,i,4}.
\end{aligned}$$

We now evaluate these variances with a slight refinement of the argument used in Theorem 3.1 of Ho and Hsing(1997), mainly replacing $K(\cdot)$ by $K(\cdot,u_i(m))-K(\cdot,u_i(m+1))$ and retaining the increments $u_i(m+1)-u_i(m)$. Note first that both $Z_{n,3}$ and $Z_{n,4}$ are sums of orthogonal random variables. Since $0<u_i(m+1)-u_i(m)\le C\cdot D_n2^{-m}$,

$$V_{n,m,i,4}\le C\cdot n^{-1}\sum_{p=1}^{J}\{K_\infty^{(p,0)}(0,u_i(m))-K_\infty^{(p,0)}(0,u_i(m+1))\}^2$$

$$\leq C \cdot (D_n 2^{-m})^2 n^{-1}, \tag{A.6}$$

$$V_{n,m,i,3} \leq C \cdot (D_n 2^{-m})^{\alpha} n^{-1}, \tag{A.7}$$

where the presence of α is due to property (3.6). By Lemmas 6.1 and 6.2−(i) of Ho and Hsing (1997), we have

$$V_{n,m,i,1} + V'_{n,m,i,1} \leq C \cdot d_n n^{-1} \tag{A.8}$$

(cf. (6.12) and (6.13) of Ho and Hsing (1997)), where the use of the two lemmas is justified by property (3.4). The term d_n is, for appropriate $X_{\bar{b}}$ (see Remark 2−(2)), $d_n = \{E[K_{1,\lambda}^{(t,0)}(X_{\bar{b}}, u_i(m+1), u_i(m+1))]^4\}^{1/2} \leq C\cdot(u_i(m+1)-u_i(m))^2 \leq C \cdot (D_n 2^{-m})^2$. Similarly,

$$V_{n,m,i,2} \leq C \cdot (D_n 2^{-m})^2 \max\{n^{-1}, n^{-(J+1)(2\beta-1)+\zeta}\} \tag{A.9}$$

for any $\zeta > 0$ (cf. (6.15) of Ho and Hsing (1997)). Then

$$\begin{aligned} P_{n,1}(b) &\leq \sum_{m=0}^{M-1} b^{-2}(m+3)^4 \sum_{i=0}^{2^{m+1}-1} \mathrm{Var}\,(\tilde{S}_{n,j}(u_i(m), u_i(m+1))) \\ &\leq C(D_n^2 \max\{n^{-1}, n^{-(J+1)(2\beta-1)+\zeta}\} + D_n^{\alpha} n^{-1}[1+I(\alpha=1)\log_2 n]). \end{aligned} \tag{A.10}$$

Choose $M = \log_2 n^{c'}$ with sufficiently large $c' > 4\beta/\alpha$ such that $2^{-M\alpha/2} = o(n^{-1-(2\beta-1)})$ and, as a result, $P_{n,2}(b) + P_{n,3}(b)$ is dominated by the bound in the last inequality of (A.10). Note that we arrive at (A.6), (A.7), (A.8), and (A.9) without any restriction on $J(2\beta-1)$. Thus (A.10) holds for both parts (i) and (ii). We now prove part (i). By (2.1),

$$\sup_{y\in I_n} |A_n(y,\theta)| + \sup_{y\in I_n} |B_n(y,\theta)| = O_p(\{n^{-(\beta-1/2)}L(n)\}^{J+1}). \tag{A.11}$$

Comparing the two bounds on the righthand side of (A.10) and (A.11), we have $\sup_{y\in I_n} |G_n(y,\theta)| = o_p(n^{-J(\beta-1/2)-\tau})$ for any $\tau < \beta - 1/2$ if $(J+1)(2\beta-1) \leq 1$, and $\tau < -J(\beta-1/2)+\min\{(J+1)(\beta-1/2), 1/2+\alpha(\beta-1/2)/2\}$ if $(J+1)(2\beta-1) > 1$. Applying Taylor's expansion to $K_\infty(0,\theta_n)-K_\infty(0,\theta)$ in (A.4) up to the $(J+1)$-th term, we see from (3.1) that $R_{n,J} = G_n + K_\infty^{(0,J+1)}(0,\theta^*)(\theta_n-\theta)^{J+1}/(J+1)!$ Since $|\theta_n - \theta|^{J+1}$ is of the same order as the bound on the right side of (A.11), the representation (3.1) in part (i) holds. To prove part (ii) we follow the same steps as in part (i) but adopt a different upper bound for $\sup_{y\in I_n} |G_n(y,\theta)|$.

First, $\sup_{y\in I_n}|B_n(y,\theta)| = o_p(n^{-1/2-(\beta-1/2)+\delta})$ for any $\delta > 0$, since $J(2\beta-1) \geq 1$ implies $\bar{Y}_{n,J} = o_p(n^{-1/2+\delta})$. Next, we aim to bound $\sup_{y\in I_n}|C_n(y,\theta)|$ and $\sup_{y\in I_n}|A_n(y,\theta)|$. As noted before, (A.6), (A.7), (A.8), and (A.9) hold whether $J(2\beta-1)$ is greater than one or not. Now that the current case assumes $J(2\beta-1) \geq 1$, the upper bound on the righthand side of (A.10) is $C \cdot D_n^{\alpha} n^{-1+\delta}$. Hence

$$\sup_{y\in I_n}|C_n(y,\theta)| = o_p(n^{-(1/2+\alpha(\beta-1/2)/2)+\delta}). \tag{A.12}$$

To bound $\sup_{y\in I_n}|A_n(y,\theta)|$, we first note that for any j, $2 \leq j \leq J-1$, $|\bar{Y}_{n,j}|D_n^{J-j+1} = o_p(n^{-(J+1)(\beta-1/2)+\delta})$ if $j(2\beta-1) < 1$, and $|\bar{Y}_{n,j}|D_n^{J-j+1} = o_p(n^{-1/2-(J-j+1)(\beta-1/2)+\delta})$ if $j(2\beta-1) \geq 1$. Since $J(2\beta-1) \geq 1$, it is clear that for $1 \leq \alpha \leq 2$,

$$\max\{-(J+1)(\beta-1/2), -1/2-(J-j+1)(\beta-1/2)\} < -1/2-(\beta-1/2) < -1/2-\alpha(\beta-1/2)/2.$$

Hence $\sup_{y\in I_n}|A_n(y,\theta)| = o_p(n^{-(1/2+(\beta-1/2))+\delta})$, which is dominated by the bound in 2pt (A.12) for $\sup_{y\in I_n}|C_n(y,\theta)|$. Therefore, $\sup_{y\in I_n}|G_n(y,\theta)| = o_p(n^{-(1/2+\alpha(\beta-1/2)/2)+\delta})$, and, from the identity $R_{n,J} = G_n + K_\infty^{(0,J+1)}(0,\theta^*)(\theta_n - \theta)^{J+1}/(J+1)!$, part (ii) follows. For the case of the indicator function, the proof is almost the same, but simpler (see Remark 2-(2)) and omitted (part (i) is essentially done in Ho and Hsing (1996)).

Proof of Theorem 3.2. The argument is almost the same as in the part (ii) of Theorem 3.1. Because of the assumption $\sum_i |a_i| < \infty$, only a few changes are made: J is identically one (hence $A_n(y,\theta)$ in (A.5) vanishes), $D_n = n^{-1/2}$, and $|\bar{Y}_{n,1}| = o_p(n^{-1/2+\delta})$.

References

Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930-934.

Avram, F. and Taqqu, M. S. (1987) Noncentral limit theorems and Appell polynomials. *Ann. Probab.* **15**, 767-775.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics* (Edited by E. Eberlein and M. S. Taqqu), 162-192. Birkhäuser, Boston.

Fox, R. and Taqqu, M. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* **74**, 213-240.

Gastwirth, J. L. (1971). On the sign test for symmetry. *J. Amer. Staist. Assoc.* **66**, 821-823.

Giraitis, L. (1985). Central limit theorem for functionals of a linear process. *Lithuanian Math. J.* **25**, 25-35.

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

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

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25**, 1636-1669.

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

Major, P. (1981). Multiple Wiener-Ito integrals: with applications to limit theorems. *Lecture Notes in Math.* **849**, Springer, New York.

Ralescu, S. and Puri, M. L. (1984). On the Berry-Esseen rates, a law of the iterated logarithm and an invariance principle for the proportion of the sample below the sample mean. *J. Multivariate Anal.* **14**, 231-247.

Rosenblatt, M. (1961). Independence and dependence. *Proc. Fourth Berkeley Symp. Math. Statis. Probab.*, 431-443. Univ. California Press.

Sen, P. K. (1972) On the Bahadur representation of sample quantiles for sequences of phi-mixing random variables. *J. Multivariate Anal.* **2**, 77-95.

Surgalis, D. (1983). Zones of attraction of self-similar multiple integrals. *Lithuanian Math. J.* **22**, 327-340.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

E-mail: hcho@stat.sinica.edu.tw

(Received April 2001; accepted May 2002)