The Annals of Statistics
2002, Vol. 30, No. 5, 1441–1459

KERNEL DENSITY ESTIMATION FOR LINEAR PROCESSES

BY WEI BIAO WU[1] AND JAN MIELNICZUK

University of Chicago and Polish Academy of Sciences

In this paper we provide a detailed characterization of the asymptotic behavior of kernel density estimators for one-sided linear processes. The conjecture that asymptotic normality for the kernel density estimator holds under short-range dependence is proved under minimal assumptions on bandwidths. We also depict the dichotomous and trichotomous phenomena for various choices of bandwidths when the process is long-range dependent.

1. Introduction. The kernel method introduced by Rosenblatt (1956) has received considerable attention in nonparametric estimation of probability densities. Let $\{X_t\}_{t=1}^{\infty}$ be a stationary sequence with a marginal density f. Then the kernel density estimator of f is defined as

$$f_n(x_0) = \frac{1}{nb_n} \sum_{j=1}^{n} K\left(\frac{x_0 - X_j}{b_n}\right), \qquad x_0 \in \mathbb{R}, \tag{1}$$

where the kernel K is some not necessarily positive function such that $\int_{\mathbb{R}} K(s)\,ds = 1$, and the bandwidths $\{b_n\}$ satisfy *natural conditions*, $b_n \to 0$ and $nb_n \to \infty$. Many of the previous results concerning asymptotic properties of f_n are established under the assumption that the X_t are independent. See Silverman (1986) and references therein. A classical result of Parzen (1962) states that when X_i are i.i.d. and the bandwidths satisfy the natural conditions, $(nb_n)^{1/2}[f_n(x_0) - \mathbb{E} f_n(x_0)]$ converges in distribution to the normal law with zero mean and variance $f(x_0)\int_{\mathbb{R}} K^2(s)\,ds$. A natural problem is: What is the limiting behavior of (1) when dependence between observations is allowed?

For weakly dependent observations Parzen's result was generalized under various mixing conditions; see, for example, Robinson (1983), Castellana and Leadbetter (1986) and Bosq [(1996), Theorem 2.3]. It is proved that in such cases asymptotic law of the centered density estimator standardized by $(nb_n)^{1/2}$ is the same as in the i.i.d. case. However, it turns out that this result does not hold for strongly dependent sequences: the asymptotic law as well as standardization might be completely different [see, e.g., Csörgő and Mielniczuk (1995)]. Here, we say that a sequence is *short-range dependent* (SRD) if its covariances are absolutely summable and *long-range dependent* (LRD) if they are not; see Beran

Received December 2000; revised November 2001.

[1]Supported in part by a U.S. Army Research Office grant.

AMS 2000 subject classifications. Primary 60F05, 60F17; secondary 60G35.

Key words and phrases. Long- and short-range dependence, kernel density estimators, linear process, martingale central limit theorem.

(1994) for an overview of the rapidly growing literature on statistical properties of LRD sequences. Ho (1996) considered an instantaneous transformation of long-range dependent Gaussian sequences and exhibited dichotomous behavior of $f_n(x_0)$ depending on the size of the bandwidths. For small bandwidths, LRD notwithstanding, the asymptotic law of $f_n(x_0)$ is the same as in the i.i.d. case. For large bandwidths, LRD prevails, affecting standardization and the asymptotic law of $f_n(x_0)$.

We assume throughout the paper that X_n, $n \in \mathbb{Z}$, is an infinite order moving-average process given by $\sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$, where $\{\varepsilon_i,\ i \in \mathbb{Z}\}$ are i.i.d. random variables with zero mean and finite variance, and $\{a_i\}_{i=0}^{\infty}$ is a sequence of real numbers such that $\sum_{i=0}^{\infty} a_i^2 < \infty$. This model was proposed by Woodroofe (1970) in the context of studying the asymptotic normality of (1). The setting is very general since many important time series models, such as ARMA and fractional ARIMA, admit this form. Withers (1981) showed that one needs restrictive conditions on the decay rate of a_n in order to get the strong mixing property for the linear process; see Pham and Tran (1985), Gorodetskii (1977) and Doukhan (1994) for more discussion about mixing properties of linear processes. For SRD processes X_t, there has been a long-standing conjecture stating that Parzen's result holds under the natural conditions on $\{b_n\}$ specified above. An important early contribution to the solution of this problem is Chanda (1983). Using some modification of Chanda's approach, Hallin and Tran (1996) showed the asymptotic normality of (1) under conditions $|a_n| = \mathcal{O}[n^{-(4+\delta)}]$ and $nb_n^{(13+2\delta)/(2+2\delta)}/\log\log n \to \infty$ for some $\delta > 0$. Under the same condition on a_n, Coulon-Prieur and Doukhan (2000) improved Hallin and Tran's results by imposing only the natural conditions on b_n. Honda (2000) obtained similar results using developments for empirical processes based on moving averages due to Giraitis, Koul and Surgailis (1996).

In this paper, we shall give a detailed characterization of the limiting behavior of (1) in accordance with different rates of $b_n \to 0$ and the coefficients $\{a_n\}$ of the linear process. In particular, we provide an affirmative answer to the stated conjecture proving that, for SRD sequences, asymptotic normality of the kernel density estimate is a rule rather than an exception. Namely, we show that Parzen's result holds under natural conditions on $\{b_n\}$ and $\sum_{i=1}^{\infty} |a_i| < \infty$. The last condition roughly corresponds to the SRD case. In particular, it is satisfied if $a_i = \mathcal{O}(i^{-\beta})$ for some $\beta > 1$. A frequently assumed form $a_n = n^{-\beta} L(n)$ for some slowly varying function L is not even required in those conditions. For the LRD situation $a_n = n^{-\beta} L(n)$, $1/2 < \beta < 1$, we observe dichotomous and trichotomous behavior under different conditions depending on the strength of the dependence and the decay rate of the bandwidth sequence. Specifically, let the bandwidth $b_n = n^{\alpha} L_1(n)$, where $-1 < \alpha < 0$ and $L_1(\cdot)$ is a slowly varying function. Let r_0 be the smallest $j \geq 1$ such that the jth derivative $f^{(j)}(x_0) \neq 0$ provided it exists. If $r_0 = 1$, then $f_n(x_0) - \mathbb{E} f_n(x_0)$ is asymptotically normal with a form of norming sequence depending on whether $\alpha < 2\beta - 2$ or $\alpha > 2\beta - 2$. In the case $r_0 \geq 2$, a richer structure exists. On the quadrangle $\{(\beta, \alpha) \in (1/2, 1) \times (-1, 0) : \alpha(r_0 - 1/2) + 1 - \beta < 0,\ \alpha + 1 - r_0(2\beta - 1) < 0\}$ as well as on the triangle $\{(\beta, \alpha) \in (1/2, 1) \times (-1, 0) : \alpha(r_0 - 1/2) + 1 - \beta > 0,\ \alpha + \beta - 1/2 > 0\}$ (cf. Figure 1), $f_n(x_0) - \mathbb{E} f_n(x_0)$ is asymptotically normal but with different norming sequences. However, on the triangle $\{(\beta, \alpha) \in (1/2, 1) \times (-1, 0) : \alpha + \beta - 1/2 < 0,\ \alpha + 1 - r_0(2\beta - 1) > 0\}$, the suitably normalized sequence $f_n(x_0) - \mathbb{E} f_n(x_0)$ converges to a multiple Wiener–Itô integral. A condition describing the borderline of the dichotomy amounts to comparison of a variance of a sample mean of the long-range dependent sequence and that of a kernel density estimate for an i.i.d. sample with the same marginal density f. This coincides with Hall and Hart's (1990) dichotomy condition for the mean integrated squared error of (1) in the LRD case.

In this paper we demonstrate that the martingale central limit theorem may be successfully applied to answer the question about the asymptotic law of f_n in the case of SRD sequences and is vital for analogous analysis in the LRD case. In the context of nonlinear functionals of linear processes, the martingale approach was first cleverly employed by Ho and Hsing (1996, 1997). We believe that this is also an effective method for other nonparametric functional estimation problems involving dependent variables such as regression analysis. From the practical point of view, the main message for statisticians is that, for a SRD sequence, the normal approximation may be used to construct pointwise confidence intervals for $f(x_0)$ which asymptotically attain a prescribed level of confidence.

The remainder of the paper is structured as follows. Preliminaries and heuristics of the presented approach are given in Section 2. In Sections 3 and 4 we discuss cases of SRD and LRD processes, respectively. Proofs are deferred to Section 5. In Section 6 some unsolved problems are discussed.

2. Notation and preliminaries. Denote by $f_1(\cdot)$ and $f(\cdot)$ the density functions of ε_t and X_t, respectively. Before stating the main results, it might be helpful to explain intuitively the discussed approach. Let

$$\mathbb{E} f_n(x_0) = b_n^{-1} \int_{\mathbb{R}} K[b_n^{-1}(x_0 - u)] f(u)\, du = \int_{\mathbb{R}} K(v) f(x_0 - b_n v)\, dv$$

be the mean of $\mathbb{E} f_n(x_0)$ and $\widetilde{X}_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$ be the shift process. Then we have the following decomposition:

$$nb_n[f_n(x_0) - \mathbb{E} f_n(x_0)] = M_n + N_n, \tag{2}$$

where

$$M_n = \sum_{j=1}^{n} \left\{ K\left(\frac{x_0 - X_j}{b_n}\right) - \mathbb{E}\left[K\left(\frac{x_0 - X_j}{b_n}\right) \Big| \widetilde{X}_{j-1}\right]\right\}, \tag{3}$$

$$N_n = \sum_{j=1}^{n} \left\{ \mathbb{E}\left[K\left(\frac{x_0 - X_j}{b_n}\right) \Big| \widetilde{X}_{j-1}\right] - b_n \mathbb{E} f_n(x_0)\right\}. \tag{4}$$

Put $\sigma^2(x_0) = f(x_0)\int_{\mathbb{R}} K^2(s)\,ds$. The major thrust of the decomposition (2) is due to the fact that the summands of the term M_n form a martingale difference sequence. Thus $M_n/\sqrt{nb_n}$ is always asymptotically normal $N[0, \sigma^2(x_0)]$, regardless whether X_t is short- or long-range dependent (see Lemma 2 in Section 3). We assume throughout that $\int_{\mathbb{R}} K^2(s)\,ds < \infty$ and $a_0 = 1$. For the second part, N_n, we have

$$N_n = b_n \int_{\mathbb{R}} K(v) H_n(-b_n v)\,dv,$$

where

$$H_n(z) = \sum_{t=1}^{n} \{f_1(x_0 - R_t + z) - f(x_0 + z)\}. \tag{5}$$

Observing that $R_t = X_t - \varepsilon_t$ is $\sigma(\widetilde{X}_{t-1})$-measurable and $a_0 = 1$, we have

$$\begin{aligned}\mathbb{E}\left[K\left(\frac{x_0 - X_t}{b_n}\right)\Big|\widetilde{X}_{t-1}\right] &= \int_{\mathbb{R}} K\left(\frac{x_0 - R_t - u}{b_n}\right) f_1(u)\,du \\ &= b_n \int_{\mathbb{R}} K(v) f_1(x_0 - R_t - b_n v)\,dv\end{aligned}$$

almost surely.

Intuitively speaking, if $b_n \to 0$ sufficiently fast, then the first term M_n in (2) dominates. From this observation, convergence to a Gaussian limit with a norming sequence $\sqrt{nb_n}$ follows. On the other hand, if $b_n \to 0$ at an appropriately slow rate, then the second term in (2) may also count, which possibly results in a non-Gaussian limit. This situation occurs for LRD sequences.

3. SRD sequences.

In the SRD case, we will show that N_n in (2) has a negligible contribution under fairly weak conditions and hence the central limit theorem holds. Therefore, the conjecture stated in the Introduction is true. More precisely, we have the following Theorem 1 in which $\Rightarrow$ denotes convergence in distribution. We say that a function g is Lipschitz continuous if there exists a constant $L > 0$ such that $|g(x) - g(y)| \le L|x - y|$ holds for all $x, y \in \mathbb{R}$.

THEOREM 1. *Assume that f_1 is Lipschitz continuous, $f(x_0) \neq 0$, $\int_{\mathbb{R}} K^2(s)\,ds < \infty$ and $\sum_{i=1}^{\infty} |a_i| < \infty$. Let $b_n \to 0$ and $nb_n \to \infty$. Then we have*

$$(nb_n)^{1/2}[f_n(x_0) - \mathbb{E} f_n(x_0)] \Rightarrow N\left[0, f(x_0)\int_{\mathbb{R}} K^2(s)\,ds\right]. \tag{6}$$

Concerning methodology, let us note that there is a fundamental difference between Theorem 1 and previous results such as Hallin and Tran (1996) and Rosenblatt (1970). The traditionally employed approach consists in using Bernstein's small-block–large-block method in conjunction with various mixing conditions.

See also Robinson (1983) and Tran (1992). However, mixing conditions are usually unverifiable and might be too restrictive [see, e.g., Carbon and Tran (1996) and Withers (1981)]. This drawback is partly due to an i.i.d. approximation which is the essence of Bernstein's method. That is why the martingale approximation, which entails only very mild conditions on the sequence $\{a_i\}$, seems to be superior to blocking techniques. Another method for proving asymptotic normality is Rio's (1995) central limit theorem. Rio's method is applied by Coulon-Prieur and Doukhan (2000).

A replacement of $\mathbb{E} f_n(x_0)$ by $f(x_0)$ in (6) is a routine problem in density estimation theory. For example, if f has $p-1$ absolutely continuous derivatives, $f^{(p)}$ is bounded and K is of order p, then $\mathbb{E} f_n(x) - f(x) = \mathcal{O}(b_n^p)$ for any x and thus the centering $\mathbb{E} f_n(x_0)$ may be changed to $f(x_0)$ without affecting the above result provided that $nb_n^{2p+1} \to 0$.

Using standard methods together with the proof of Theorem 1, it may be shown that under the stated conditions the vector $V_n = (nb_n)^{1/2}[f_n(x_0) - \mathbb{E} f_n(x_0), \dots, f_n(x_k) - \mathbb{E} f_n(x_k)]$ converges in distribution to a product of normal laws $N[0, f(x_i)\int K^2(s)\,ds]$ for $i = 0, \dots, k$ at different points $x_0, x_1, \dots, x_k \in \mathbb{R}$. This follows from the Cramér–Wold lemma, after noting that components of V_n become asymptotically uncorrelated. Moreover, using similar reasoning, an analogous result for derivatives of $f_n(\cdot)$ may be proved.

4. LRD sequences. For LRD processes, we need to assume that a_n has the form $n^{-\beta}L(n)$, where $1/2 < \beta < 1$ and $L(\cdot)$ is a slowly varying function at ∞. In this case, a closer investigation of (4) is necessary as N_n may also have a significant contribution. Because of the local character of N_n its behavior is determined by the behavior of $H_n(\cdot)$ and its derivatives at 0. It follows from the breakthrough work by Ho and Hsing (1997) that the distributional limit of $H_n(0)$ is related to the power rank of $G(s) := f_1(x_0 - s) - f(x_0)$ with respect to the distribution of R_1. See Ho and Hsing (1997) for a definition of the power rank.

Assume that f_1 is Lipschitz continuous. By Lemma 1, $f(x)$ exists and is equal to $\mathbb{E} f_1(x - R_1)$. Let $G_\infty(s) := \mathbb{E}[G(s + R_1)] = f(x_0 - s) - f(x_0)$. If there exists an $r_0 = r(x_0) \in \mathbb{N}$ such that the rth derivative of $f(\cdot)$ at x_0, $f^{(r)}(x_0)$, is zero for all $r < r_0$ and is nonzero for $r = r_0$, then it follows that the power rank of $G(\cdot)$ is r_0. If f_1 is r_0 times differentiable, then by Taylor's expansion,

$$H_n(-b_n v) = \sum_{j=0}^{r_0-1} H_n^{(j)}(0)\frac{(-b_n v)^j}{j!} + \frac{(-b_n v)^{r_0}}{(r_0)!} H_n^{(r_0)}(\xi) \tag{7}$$

for some $\xi \in \mathbb{R}$. Let $G^{(r)}(s) = f_1^{(r)}(x_0 - s) - f^{(r)}(x_0)$. Then $G_\infty^{(r)}(s) = \mathbb{E}G^{(r)}(s + R_1) = f^{(r)}(x_0 - s) - f^{(r)}(x_0)$. Hence the power rank for $G^{(r)}$ with respect to R_1 is $r_0 - r$ when $0 \le r \le r_0 - 1$. Then the result follows from Ho and Hsing (1996, 1997), where the limiting behavior of $H_n^{(r)}(0)$ is studied. Let

$$\sigma_{n,r}^2 = C(\beta, r)n^{2-r(2\beta-1)}L^{2r}(n)[\mathbb{E}(\varepsilon^2)]^r,$$

where

$$C(\beta, r) = \{r![1 - r(\beta - \tfrac{1}{2})][1 - r(2\beta - 1)]\}^{-1} \left[\int_0^\infty (x + x^2)^{-\beta}\, dx\right]^r.$$

Roughly speaking, $H_n^{(r)}(0)$ is of order σ_{n,r_0-r} if $(r_0 - r)(2\beta - 1) < 1$ and $\sqrt{n}$ otherwise. From (2) and (7), we see that among terms M_n, $H_n^{(r)}(0)b_n^{r+1}$, $0 \le r \le r_0 - 1$, only the ones with their order reaching $\max[\sqrt{nb_n}, b_n^{r+1}\max(\sigma_{r_0-r}, \sqrt{n})]$ have a major contribution.

4.1. *Dichotomy.* Denote by $\mathbf{b}$ the bandwidth sequence $\{b_n\}$. For $r_0 = 1$ and $c \in [0, \infty]$, we define

$$\mathbf{D}_c = \left\{(\mathbf{b}, \beta, L(\cdot)) : \lim_{n\to\infty} \frac{b_n \sigma_{n,1}}{\sqrt{nb_n}} = c\right\}.$$

Hence $\mathbf{D}_0$ and $\mathbf{D}_\infty$ correspond to $(\mathbf{b}, \beta, L(\cdot))$ such that $b_n\sigma_{n,1} = o(\sqrt{nb_n})$ and $\sqrt{nb_n} = o(b_n\sigma_{n,1})$, respectively. Recall that $r_0 = 1$ is equivalent to $f^{(1)}(x_0) \neq 0$.

THEOREM 2. *Assume that* (a) $\mathbb{E}(\varepsilon_1^4) < \infty$ *and* f_1 *is three times differentiable with bounded, continuous and integrable derivatives*, (b) $f(x_0) \neq 0$, $f^{(1)}(x_0) \neq 0$ *and* (c) $\int_{\mathbb{R}} |vK(v)|\, dv < \infty$.

(i) *If* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{D}_0$, *then*

$$\sqrt{nb_n}[f_n(x_0) - \mathbb{E}f_n(x_0)] \Rightarrow N\left[0, f(x_0)\int_{\mathbb{R}} K^2(s)\, ds\right]. \tag{8}$$

(ii) *If* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{D}_\infty$, *then*

$$\frac{n}{\sigma_{n,1}}[f_n(x_0) - \mathbb{E}f_n(x_0)] \Rightarrow N[0, |f^{(1)}(x_0)|^2]. \tag{9}$$

It follows from the proof of Theorem 2 (see Section 5) and Theorem 1(i) in Giraitis and Surgailis (1999) that the assumption $\mathbb{E}\varepsilon_1^4 < \infty$ may be reduced to $\mathbb{E}|\varepsilon_1|^{2+\eta} < \infty$ for some $\eta > 0$ provided that $|\mathbb{E}e^{iu\varepsilon_1}|$ is bounded by $C/(1 + |u|)^\delta$, $u \in \mathbb{R}$, for some $\delta > 0$ and a finite constant C.

The following Corollary 1 is an immediate consequence of Theorem 2, where b_n is assumed to have the special form $b_n = n^\alpha L_1(n)$ where $L_1(\cdot)$ is some slowly varying function and $-1 < \alpha < 0$. The boundary line with equation $\alpha - 2\beta + 2 = 0$ divides the rectangle $(1/2, 1) \times (-1, 0)$ into two triangles pertaining to different norming sequences. On both triangles the limit of (1) is normal. Triangle $\mathbf{D}_0$ corresponds to $\sqrt{nb_n}$ norming, while triangle $\mathbf{D}_\infty$ corresponds to $n/\sigma_{n,1}$ norming. Thus, for any $1/2 < \beta < 1$ corresponding to the LRD case, by choosing sufficiently small bandwidths $b_n = n^\alpha L_1(n)$ with $-1 < \alpha < 2\beta - 2$

we obtain white noise as a limit of the process $(nb_n)^{1/2}(f_n(\cdot) - \mathbb{E}f_n(\cdot))$. In contrast, for large bandwidths corresponding to $0 > \alpha > 2\beta - 2$ the effect of long-range dependence prevails. We conjecture that in this case under assumptions of Theorem 2 the process $(n/\sigma_{n,1})[f_n(\cdot) - \mathbb{E}f_n(\cdot)]$ converges in $\mathbb{C}[-\infty, +\infty]$ to a degenerate process $f^{(1)}(\cdot)Z$, where Z is a standard normal random variable [see, e.g., Proposition 1 in Csörgő and Mielniczuk (1995) and Ho and Hsing (1996)]. The behavior on the boundary $\mathbf{D}_c$, which roughly corresponds to $b_n \approx n^{2\beta-2}$, is handled in Theorem 4(a). The limiting distribution is the convolution of distributions in (8) and (9).

COROLLARY 1. *Assume that* $b_n = n^\alpha L_1(n)$, *where* $L_1(\cdot)$ *is a slowly varying function and* $-1 < \alpha < 0$. *Then under the conditions of Theorem* 2 *convergence in* (8) *holds for* $\alpha < 2\beta - 2$, *whereas for* $\alpha > 2\beta - 2$ *convergence in* (9) *holds*.

4.2. *Trichotomy.* For $r_0 \geq 2$ we define

$$\mathbf{T}_{1c} = \left\{(\mathbf{b}, \beta, L(\cdot)) : \lim_{n\to\infty} \frac{nb_n}{\sigma_{n,1}} = c\right\}, \tag{10}$$

$$\mathbf{T}_{2c} = \left\{(\mathbf{b}, \beta, L(\cdot)) : \lim_{n\to\infty} \frac{b_n^{r_0}\sigma_{n,1}}{\sqrt{nb_n}} = c\right\} \tag{11}$$

and

$$\mathbf{T}_{3c} = \left\{(\mathbf{b}, \beta, L(\cdot)) : \lim_{n\to\infty} \frac{b_n\sigma_{n,r_0}}{\sqrt{nb_n}} = c\right\}. \tag{12}$$

It can be easily verified that

$$\frac{b_n\sigma_{n,r_0}}{\sqrt{nb_n}} \times \left[\frac{nb_n}{\sigma_{n,1}}\right]^{r_0-1} \times \left[\frac{b_n^{r_0}\sigma_{n,1}}{\sqrt{nb_n}}\right]^{-1} = \frac{C(\beta, r_0)}{[C(\beta,1)]^{r_0}}, \tag{13}$$

which explains the interplay between conditions appearing in (10)–(12).

Recall that when $r(2\beta - 1) < 1$, the multiple Wiener–Itô integral (MWI) $Z_{r,\beta}$ is defined as

$$Z_{r,\beta} = C(\beta, r)^{-1/2} \int_{\mathcal{S}} \left\{\int_0^1 \prod_{i=1}^{r} [\max(v - u_j, 0)]^{-\beta}\, dv\right\} d\mathbb{B}(u_1)\cdots d\mathbb{B}(u_r), \tag{14}$$

where $\{\mathbb{B}(u), u \in \mathbb{R}\}$ is a standard two-sided Brownian motion,

$$\mathcal{S} = \{(u_1, \ldots, u_r) \in \mathbb{R}^r : -\infty < u_1 < \cdots < u_r < 1\}$$

and the norming constant $C(\beta, r)^{-1/2}$ [defined below (7)] ensures that $\mathbb{E}(Z_{r,\beta}^2) = 1$. The MWI $Z_{r,\beta}$ is Gaussian for $r = 1$ and non-Gaussian for $r > 1$ [see, e.g., Taqqu (1979)]. Let $\kappa_j = \int_{\mathbb{R}} K(v)v^j\, dv/j!$ provided it exists.

THEOREM 3. *Assume that* (a) $\mathbb{E}[\varepsilon_1^{\max(4,2r_0)}] < \infty$, f_1 *is* $r_0 + 2$ *times differentiable with bounded, continuous and integrable derivatives*; (b) $f(x_0) \neq 0$, $f^{(r_0)}(x_0) \neq 0$ *while* $f^{(r)}(x_0) = 0$ *for* $1 \le r < r_0$; *and* (c) $\int_{\mathbb{R}} |v^{r_0} K(v)|\, dv < \infty$.

(i) *If* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{20} \cap \mathbf{T}_{30}$ *then*

$$\sqrt{nb_n}[f_n(x_0) - \mathbb{E} f_n(x_0)] \Rightarrow N\left[0, f(x_0) \int_{\mathbb{R}} K^2(s)\, ds\right]. \tag{15}$$

(ii) *If* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{2\infty} \cap \mathbf{T}_{1\infty}$ *and* $\kappa_{r_0-1} \neq 0$, *then*

$$\frac{nb_n}{b_n^{r_0}\sigma_{n,1}}[f_n(x_0) - \mathbb{E} f_n(x_0)] \Rightarrow N\big[0, \big|f^{(r_0)}(x_0)\kappa_{r_0-1}\big|^2\big]. \tag{16}$$

(iii) *If* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{10} \cap \mathbf{T}_{3\infty}$, *then*

$$\frac{n}{\sigma_{n,r_0}}[f_n(x_0) - \mathbb{E} f_n(x_0)] \Rightarrow (-1)^{r_0} f^{(r_0)}(x_0) Z_{r_0,\beta}. \tag{17}$$

In order to better understand the conditions assumed in the last result it is helpful to note that $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{10}$ is equivalent to $b_n^{r_0}\sigma_{n,1}/(b_n\sigma_{n,r_0}) \to 0$, whereas the condition $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{1\infty}$ coincides with $b_n^{r_0}\sigma_{n,1}/(b_n\sigma_{n,r_0}) \to \infty$. The former condition implies that for $1 \le j \le r_0 - 1$, $b_n^{j+1}\sigma_{n,r_0-j} = o(b_n\sigma_{n,r_0})$, whereas the latter implies that for $0 \le j \le r_0 - 2$, $b_n^{j+1}\sigma_{n,r_0-j} = o(b_n^{r_0}\sigma_{n,1})$. The convergence (15) occurs in the situation when M_n dominates N_n. The situation when N_n dominates M_n splits into two cases depending on whether the first or the last term in the expansion (7) dominates (we disregard for the time being the case when they are of the same order). In the former case we still have the normal limit (16) but with different standardization, whereas in the latter both the standardization and the limit change. Theorem 3 readily entails the following corollary.

COROLLARY 2. *Assume that* $r_0 \ge 2$ *and conditions* (a) *and* (b) *in Theorem* 3 *hold. Let* $b_n = n^\alpha L_1(n)$, *where* $L_1(\cdot)$ *is a slowly varying function and* $-1 < \alpha < 0$. *Then* (15), (16) *or* (17) *holds*, *respectively*, *if*

$$\alpha(r_0 - 1/2) + 1 - \beta < 0, \qquad \alpha + 1 - r_0(2\beta - 1) < 0;$$
$$\alpha(r_0 - 1/2) + 1 - \beta > 0, \qquad \alpha + \beta - 1/2 > 0;$$

or

$$\alpha + \beta - 1/2 < 0, \qquad \alpha + 1 - r_0(2\beta - 1) > 0.$$

Figure 1 contains a graphic representation of the corollary. Three lines, T_1, T_2, T_3, in the graph are described by equations $\alpha + \beta - 1/2 = 0$, $\alpha(r_0 - 1/2) +$

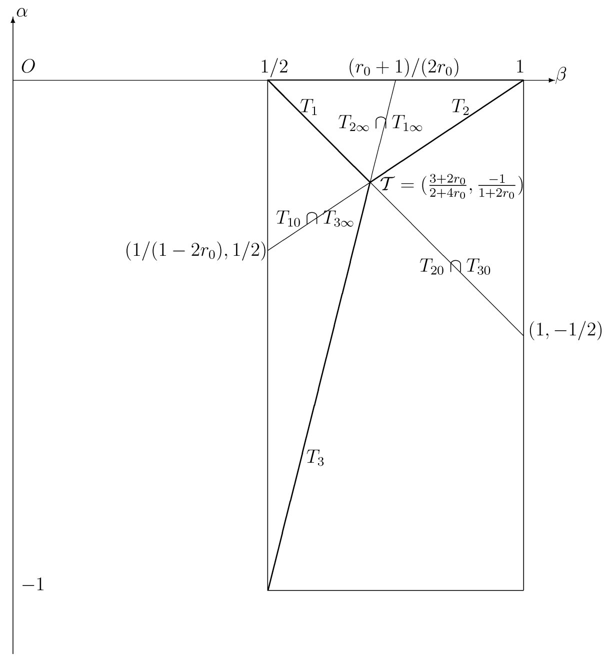

FIG. 1. *Graphic representation of Corollary* 2.

$1-\beta=0$ and $\alpha+1-r_0(2\beta-1)=0$, respectively. Interestingly enough, they have a common joint point $\mathcal{T}=(\frac{3+2r_0}{2+4r_0}, \frac{-1}{1+2r_0})$. This observation is a consequence of (13). Lines T_1, T_2, T_3 emanating from $\mathcal{T}$ divide the rectangle $(1/2, 1)\times(-1, 0)$ into three regions corresponding to a different limiting distribution for each region. On the quadrangle $\mathbf{T}_{20}\cap\mathbf{T}_{30}$, the use of the norming sequence $\sqrt{nb_n}$ yields a Gaussian limit. On the triangle $\mathbf{T}_{2\infty}\cap\mathbf{T}_{1\infty}$, we have also a Gaussian limit but with different norming sequence $nb_n/(b_n^{r_0}\sigma_{n,1})$. A non-Gaussian limit is obtained on the triangle $\mathbf{T}_{10}\cap\mathbf{T}_{3\infty}$ with norming sequence $n/\sigma_{n,r_0}$. This is the reason for the term *trichotomy*.

4.3. *Boundary cases.* Theorem 4 discusses boundary cases. It is shown that the asymptotic variances are different from those obtained in the previous limit theorems. The result is not complete. It covers the case when the martingale term M_n is of the same order as the last term in the expansion (7) which in turn dominates all other summands of the expansion. Moreover, in part (c) we study the case when all summands in (7) for $j=0,1,\dots,r_0-1$ are of the same order and they dominate M_n. The asymptotic distribution of $f_n(x_0)-\mathbb{E}f_n(x_0)$ can be interpreted as the convolution of independent components appearing as limits in Theorem 3.

THEOREM 4. *Assume that conditions* (a)–(c) *in Theorem* 3 *hold. Let* $r_0 = r(x_0) \geq 1$ *be the power rank of* $f(\cdot)$ *at* x_0 *and assume* $\kappa_{r_0-1} \neq 0$. *If* (a) $r_0 = 1$ *and* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{D}_c$ *or* (b) $r_0 \geq 2$ *and* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{2c} \cap \mathbf{T}_{1\infty}$ *for some* $0 < c < \infty$, *then*

$$\sqrt{nb_n}[f_n(x_0) - \mathbb{E}f_n(x_0)] \Rightarrow N[0, \sigma^2(x_0, c)], \tag{18}$$

where

$$\sigma^2(x_0, c) = f(x_0)\int_{\mathbb{R}} K^2(s)\,ds + c^2\left|f^{(r_0)}(x_0)\kappa_{r_0-1}\right|^2.$$

(c) *If* $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{1c} \cap \mathbf{T}_{2\infty}$, *then*

$$\begin{aligned}&\frac{nb_n}{b_n^{r_0}\sigma_{n,1}}[f_n(x_0) - \mathbb{E}f_n(x_0)]\\ &\quad\Rightarrow f^{(r_0)}(x_0)\sum_{j=0}^{r_0-1}\left[\frac{C(\beta, r_0 - j)}{C^{r_0-j-1}(\beta, 1)}\right]^{1/2}\frac{\kappa_j}{c^{r_0-j-1}}Z_{r_0-j,\beta}.\end{aligned} \tag{19}$$

5. Proofs. In Lemma 1 the existence and continuity of marginal density functions is studied. Its variant was proved in Lemma 1(i) of Giraitis, Koul and Surgailis (1996) under an assumption on the tail behavior of the characteristic function of ε_1.

LEMMA 1. (a) *If the density function* f_1 *of* ε_t *is Lipschitz continuous*, *then the density function* f *of* X_t *exists and is also Lipschitz continuous*.

(b) *If* f_1 *is bounded and* $a_0 \neq 0, \#\{i : a_i \neq 0\} > 1$, *then* $\tilde{f}_1$, *the density function of* $R_t = X_t - a_0\varepsilon_t$, *is also bounded*.

(c) *If* f_1 *is* p *times differentiable with bounded*, *continuous derivatives*, *then* f *also satisfies the same properties*.

PROOF. (a) Let $C > 0$ be the Lipschitz constant of f_1. Then $\mathbb{E}|f_1(x - R_t) - f_1(x)| \leq \mathbb{E}(C|R_t|) \leq C[\mathbb{E}(R_t^2)]^{1/2} < \infty$. Hence $\mathbb{E}[f_1(x - R_t)]$ is finite. Let F, F_1 and $\tilde{F}_1$ be the distribution functions of X_t, ε_t and R_t, respectively. Recall that $a_0 = 1$. Since $F(x) = \int_{\mathbb{R}} F_1(x - y)\tilde{F}_1(dy)$, by the mean value theorem

$$\begin{aligned}&|F(x + \delta) - F(x) - \delta\mathbb{E}[f_1(x - R_t)]|\\ &\quad\leq \int_{\mathbb{R}} |F_1(x - y + \delta) - F_1(x - y) - \delta f_1(x - y)|\tilde{F}_1(dy)\\ &\quad= \int_{\mathbb{R}} |\delta f_1(x - y + \delta\tau) - \delta f_1(x - y)|\tilde{F}_1(dy) \leq C\delta^2,\end{aligned}$$

where $0 \leq \tau = \tau(\delta) \leq 1$. Hence by letting $\delta \to 0$, we conclude that the derivative $f(x) = dF/dx = \mathbb{E}[f_1(x - R_t)]$ exists. In addition, $|f(x) - f(y)| \leq \mathbb{E}|f_1(x - R_t) - f_1(y - R_t)| \leq C|x - y|$, which proves the Lipschitz continuity of f. Parts (b) and (c) can be similarly obtained by elementary methods. $\square$

LEMMA 2. *Let $b_n \to 0$, $nb_n \to \infty$ and K be such that $\int_{\mathbb{R}} K^2(v)\,dv < \infty$. For $X_n = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$ such that the density function f_1 of ε_1 is Lipschitz continuous and bounded, we always have*

$$(nb_n)^{-1/2} M_n \Rightarrow N[0, \sigma^2(x_0)], \qquad \sigma^2(x_0) = f(x_0) \int_{\mathbb{R}} K^2(s)\,ds.$$

PROOF. Let

$$\zeta_{n,t} = (nb_n)^{-1/2} K\big(b_n^{-1}(x_0 - X_t)\big) \quad \text{and} \quad \xi_{n,t} = \zeta_{n,t} - \mathbb{E}[\zeta_{n,t} \mid \widetilde{X}_{t-1}].$$

By the martingale central limit theorem, it suffices to show that $\sum_{t=1}^{n} \mathbb{E}[\xi_{n,t}^2 \mid \widetilde{X}_{t-1}] \xrightarrow{\mathcal{P}} \sigma^2(x_0)$ and the Lindeberg condition $n\mathbb{E}[\xi_{n,t}^2 \mathbf{1}_{[|\xi_{n,t}| > \varepsilon]}] = o(1)$ holds for any $\varepsilon > 0$. In order to prove the first statement, observe that since f_1 is bounded $|\mathbb{E}[\zeta_{n,t} | \widetilde{X}_{t-1}]| \le (nb_n)^{-1/2} b_n C$, where $C = \sup f_1(u) \int_{\mathbb{R}} |K(u)|\,du < \infty$. Thus we have

$$\left| \sum_{t=1}^{n} \mathbb{E}[\zeta_{n,t}^2 \mid \widetilde{X}_{t-1}] - \sum_{t=1}^{n} \mathbb{E}[\xi_{n,t}^2 \mid \widetilde{X}_{t-1}] \right| \le \sum_{t=1}^{n} (\mathbb{E}[\zeta_{n,t} \mid \widetilde{X}_{t-1}])^2 \le C^2 \sum_{i=1}^{n} \frac{b_n^2}{nb_n} = \mathcal{O}(b_n).$$

Hence we only need to show that $\sum_{t=1}^{n} \mathbb{E}[\zeta_{n,t}^2 \mid \widetilde{X}_{t-1}] \xrightarrow{\mathcal{P}} \sigma^2(x_0)$. Note that $\mathbb{E}[\zeta_{n,t}^2 \mid \widetilde{X}_{t-1}] = n^{-1} \int_{\mathbb{R}} K^2(z) f_1(x_0 - R_t - b_n z)\,dz$ and $\int_{\mathbb{R}} K^2(z) f(x_0 - b_n z)\,dz \to \sigma^2(x_0)$ in view of the continuity of f. Thus the required convergence follows from

$$n^{-1} \int_{\mathbb{R}} K^2(v) H_n(-b_n v)\,dv = o_P(1).$$

In order to prove it, observe that the linear process R_t is ergodic as a "moving function" of the ergodic stationary sequence ε_i and whence $f_1(x_0 - R_t)$ is ergodic. Thus from the ergodic theorem we obtain $\mathbb{E}|H_n(0)| = o(n)$ and the last displayed formula follows from $\int_{\mathbb{R}} K^2(v) |H_n(-b_n v) - H_n(0)|\,dv = o(n)$. However, $n^{-1}|H_n(-b_n v) - H_n(0)| = \mathcal{O}[\min(1, |b_n v|)]$ by the Lipschitz continuity of f_1 and f and we obtain $\int_{\mathbb{R}} K^2(v) |H_n(-b_n v) - H_n(0)| = o(n)$ by the Lebesgue dominated convergence theorem, ensuring that $\int_{\mathbb{R}} K^2(v) \min(1, |b_n v|)\,dv \to 0$.

The Lindeberg condition results from Corollary 9.5.2 in Chow and Teicher (1988) which implies that

$$n\mathbb{E}\big[\xi_{n,t}^2 \mathbf{1}_{[|\xi_{n,t}| > \varepsilon]}\big] \le 4n\mathbb{E}\big[\zeta_{n,t}^2 \mathbf{1}_{[|\zeta_{n,t}| > \varepsilon/2]}\big] = \mathcal{O}(1) \int_{[|K(v)| > \sqrt{nb_n}\varepsilon/2]} K^2(v)\,dv.$$

The right-hand side is $o(1)$ since $nb_n \to \infty$. $\square$

Before proving Theorem 1, we establish Lemma 3, which implies that N_n in (2) is negligible. Then the major term becomes M_n, which is asymptotically normal. For a random variable Y, we write $\|Y\| = [\mathbb{E}(Y^2)]^{1/2}$.

LEMMA 3. *Under the conditions of Theorem* 1, *we have*

$$\sup_{x\in\mathbb{R}} \|H_n(x)\| = \mathcal{O}(\sqrt{n}). \tag{20}$$

PROOF. Let ε_t', $t \in \mathbb{Z}$, be i.i.d. copies of ε_t, $t \in \mathbb{Z}$. Then $R_{t+1}^* := (R_{t+1} - a_t\varepsilon_1) + a_t\varepsilon_1'$ is a coupled version of R_{t+1}. Observe that for $1 \le t \le n$,

$$\begin{aligned}&\mathbb{E}[f_1(x - R_{t+1}) \mid \widetilde{X}_1] - \mathbb{E}[f_1(x - R_{t+1}) \mid \widetilde{X}_0]\\ &\quad = \mathbb{E}[f_1(x - R_{t+1}) - f_1(x - R_{t+1}^*) \mid \widetilde{X}_1].\end{aligned}$$

Let $H_n^\diamond(x) = \sum_{t=1}^n f_1(x - R_{t+1}) - nf(x)$. Then $H_n^\diamond(x)$ and $H_n(x - x_0)$ are identically distributed. Let $i^+ = \max(0, i)$. Since f_1 is Lipschitz continuous, we have

$$\begin{aligned}&\big\|\mathbb{E}[H_n^\diamond(x) \mid \widetilde{X}_{i+1}] - \mathbb{E}[H_n^\diamond(x) \mid \widetilde{X}_i]\big\|\\
&\quad \le \sum_{j=i^++1}^{n} \big\|\mathbb{E}[f_1(x - R_{j+1}) \mid \widetilde{X}_{i+1}] - \mathbb{E}[f_1(x - R_{j+1}) \mid \widetilde{X}_i]\big\|\\
&\quad = \sum_{j=i^++1}^{n} \big\|\mathbb{E}[f_1(x - R_{j-i}) \mid \widetilde{X}_1] - \mathbb{E}[f_1(x - R_{j-i}) \mid \widetilde{X}_0]\big\|\\
&\quad = \sum_{j=i^++1}^{n} \big\|\mathbb{E}[f_1(x - R_{j-i}) - f_1(x - R_{j-i}^*) \mid \widetilde{X}_1]\big\|\\
&\quad \le \sum_{j=i^++1}^{n} \|f_1(x - R_{j-i}) - f_1(x - R_{j-i}^*)\|\\
&\quad = \sum_{j=i^++1}^{n} \mathcal{O}(|a_{j-i-1}|)\end{aligned}$$

for all $i \le n - 1$. Note that $\mathbb{E}[H_n^\diamond(x) \mid \widetilde{X}_{i+1}] - \mathbb{E}[H_n^\diamond(x) \mid \widetilde{X}_i] = 0$ almost surely for $i \ge n$. Then by the conditions of Theorem 1,

$$\begin{aligned}\|H_n(x - x_0)\| &= \|H_n^\diamond(x)\|^2\\
&= \sum_{i=-\infty}^{n-1} \big\|\mathbb{E}[H_n^\diamond(x) \mid \widetilde{X}_{i+1}] - \mathbb{E}[H_n^\diamond(x) \mid \widetilde{X}_i]\big\|^2\\
&\le \left[\sum_{i=-\infty}^{-1} + \sum_{i=0}^{n-1}\right]\left[\sum_{j=1+i^+}^{n} |a_{j-i-1}|\right]^2\\
&= \left[\sum_{i=-(n-1)}^{0} + \sum_{i=1}^{\infty}\right]\left[\sum_{j=1+(-i)^+}^{n} |a_{j+i-1}|\right]^2\end{aligned}$$

$$
\begin{aligned}
&= \mathcal{O}(n) + \sum_{i=0}^{\infty} \sum_{j_1, j_2 = 1}^{n} |a_{j_1+i}||a_{j_2+i}| \\
&= \mathcal{O}(n) + \sum_{j_1=1}^{n} \sum_{i=0}^{\infty} \sum_{j_2=1}^{n} |a_{j_1+i}||a_{j_2+i}| \\
&\le \mathcal{O}(n) + \mathcal{O}(n) \left[\sum_{i=0}^{\infty} |a_i| \right]^2 = \mathcal{O}(n).
\end{aligned}
$$
□

PROOF OF THEOREM 1. By Lemma 3, we have

$$
\begin{aligned}
\mathbb{E}\left| \int_{\mathbb{R}} K(v) H_n(-b_n v)\, dv \right| &\le \int_{\mathbb{R}} |K(v)| \times \mathbb{E}|H_n(-b_n v)|\, dv \\
&\le \int_{\mathbb{R}} |K(v)| \times \|H_n(-b_n v)\|\, dv \\
&= \mathcal{O}(\sqrt{n}),
\end{aligned}
$$

which yields (6) via Lemma 2 in conjunction with (2) and $b_n \to 0$. □

For a LRD process, we need the following Lemma 4 to describe the asymptotic expansions of H_n in (5) and (7). Ho and Hsing (1996) introduced a powerful asymptotic expansion for empirical processes of linear sequences. Let

$$
Y_{n,r} = \sum_{t=1}^{n} \sum_{0 \le j_1 < \cdots < j_r} \prod_{s=1}^{r} a_{j_s} \varepsilon_{t-j_s}, \qquad r \ge 1,\ Y_{n,0} = n. \tag{21}
$$

If $r(2\beta - 1) < 1$ and $\mathbb{E}(\varepsilon^{2r}) < \infty$, then $\mathbb{E}(Y_{n,r}^2) \sim \sigma_{n,r}^2$, where $a_n \sim b_n$ means $a_n / b_n \to 0$ when $n \to \infty$, and

$$
\frac{Y_{n,r}}{\sigma_{n,r}} \Rightarrow Z_{r,\beta}, \tag{22}
$$

where $Z_{r,\beta}$ is the multiple Wiener–Itô integral defined in (14). See Avram and Taqqu (1987) and Surgailis (1982) for a proof of (22).

LEMMA 4. *Assume conditions* (a) *and* (b) *of Theorem* 3 *with the moment condition for* ε_1 *weakened to* $\mathbb{E}(\varepsilon_1^4) < \infty$. *Then*

$$
H_n^{(j)}(0) = (-1)^{r_0 - j} f^{(r_0)}(x_0) Y_{n, r_0 - j} + o_P(\sigma_{n, r_0 - j})
$$

for any nonnegative integer j *such that* $1 \le r_0 - j < (2\beta - 1)^{-1}$ *and*

$$
\sup_{x \in \mathbb{R}} |H_n^{(r_0)}(x)| = \mathcal{O}_P(\sigma_{n,1}).
$$

PROOF. Let $\tilde{F}_{n,1}(\cdot)$ be the empirical distribution function of R_t, $1 \le t \le n$ and $\tilde{F}_1$ be the distribution function of R_t. We first consider the case $j \le r_0 - 1$. Since f_1 is $r_0 + 1$ times differentiable and $f^{(j)}(x_0) = \int_{\mathbb{R}} f_1^{(j)}(x_0 - u)\, d\tilde{F}_1(u) = \int_{\mathbb{R}} \tilde{F}_1(u) f_1^{(j+1)}(x_0 - u)\, du$, we have

$$
\begin{aligned}
H_n^{(j)}(0) &= \sum_{t=1}^{n} f_1^{(j)}(x_0 - R_t) - n f^{(j)}(x_0) \\
&= n \int_{\mathbb{R}} f_1^{(j)}(x_0 - u)\, d\tilde{F}_{n,1}(u) - n f^{(j)}(x_0) \\
&= n \int_{\mathbb{R}} [\tilde{F}_{n,1}(u) - \tilde{F}_1(u)] f_1^{(j+1)}(x_0 - u)\, du.
\end{aligned}
$$

By Theorem 2.1 in Ho and Hsing (1996) and the assumed moment condition on ε_1,

$$
n\big[\tilde{F}_{n,1}(u) - \tilde{F}_1(u)\big] = \sum_{r=1}^{r_0 - j} (-1)^r \tilde{F}_1^{(r)}(u) Y_{n,r} + S_{n, r_0 - j}(u),
$$

where $\sup_{u \in \mathbb{R}} |S_{n, r_0 - j}(u)| = o_P(\sigma_{n, r_0 - j})$. Then the first statement of the lemma follows readily since $\int_{\mathbb{R}} \tilde{F}_1^{(r)}(u) f_1^{(j+1)}(x_0 - u)\, du = f^{(r+j)}(x_0)$ for $r \le r_0 - j$ which is equal $f^{(r_0)}(x_0)$ when $r = r_0 - j$, and 0 when $r < r_0 - j$. As of the case $j = r_0$, we have similarly that $n[\tilde{F}_{n,1}(u) - \tilde{F}_1(u)] = S_{n,1}(u) - \tilde{f}_1(u) Y_{n,1}$ and $\sup_{u \in \mathbb{R}} |S_{n,1}(u)| = o_P(\sigma_{n,1})$, which ensures the second half of the lemma since $\int_{\mathbb{R}} |f_1^{(r_0+1)}(u)|\, du < \infty$ and $Y_{n,1} = \mathcal{O}_P(\sigma_{n,1})$. □

PROOF OF THEOREM 2. Applying Taylor's expansion and Lemma 4 with $r_0 = 1$, we have

$$
\begin{aligned}
H_n(-b_n v) &= H_n(0) - b_n v H_n'(\xi) \\
&= H_n(0) - b_n v \sigma_{n,1} \mathcal{O}_P(1).
\end{aligned}
$$

Then by the decomposition (2),

$$
\begin{aligned}
n b_n [f_n(x_0) - \mathbb{E} f_n(x_0)] &= M_n + b_n \int_{\mathbb{R}} K(v) H_n(-b_n v)\, dv \\
&= M_n + b_n H_n(0) + b_n^2 \sigma_{n,1} \mathcal{O}_P(1).
\end{aligned}
$$

Hence by Lemma 2, (8) follows from the assumption in (i) via

$$
\frac{n b_n}{\sqrt{n b_n}} [f_n(x_0) - \mathbb{E} f_n(x_0)] = \frac{M_n}{\sqrt{n b_n}} + o_P(1).
$$

On the other hand, (ii) implies (9) via $H_n(0) = -f^{(1)}(x_0) Y_{n,1} + \sigma_{n,1} \mathcal{O}_P(1)$, an application of Lemma 4 with $r_0 = 1$, $j = 0$ and (22). □

PROOF OF THEOREM 3. (i) We consider the cases $r_0 < (2\beta-1)^{-1}$ and $r_0 \geq (2\beta-1)^{-1}$ separately and show that in both cases $N_n = o_P(\sqrt{nb_n})$, which yields (15) by Lemma 2 and (2). For the first case, by the conditions in (i) we have $\max(b_n^{j+1}\sigma_{n,r_0-j}, 0 \leq j \leq r_0-1) = o(\sqrt{nb_n})$, which implies by (7) and Lemma 4 that

$$N_n = b_n \int_{\mathbb{R}} K(v) H_n(-b_n v)\, dv = o_P(\sqrt{nb_n}).$$

For the second case $r_0 \geq (2\beta-1)^{-1}$, let j_0 be the largest integer that is less than $r_0 + 1 - (2\beta-1)^{-1}$. Again by Taylor's expansion (7) and Lemma 4, we can write

$$H_n(-b_n v) = \sum_{j=0}^{j_0} + \sum_{j=j_0+1}^{r_0-1} + b_n^{r_0} \mathcal{O}_P(\sigma_{n,1}) v^{r_0}.$$

For the first summand, we know that $j \leq j_0$ implies $r_0 - j + 1 > (2\beta-1)^{-1}$. Then by the proof of Theorem 3.2 in Ho and Hsing (1997) and the proof of Lemma 4 under assumed conditions we have $H_n^{(j)}(0) = \mathcal{O}_P(\sqrt{n})$. As to the second summand, clearly, $j \geq j_0 + 1 \geq r_0 + 1 - (2\beta-1)^{-1}$ entails $r_0 - j < (2\beta-1)^{-1}$. Hence by Lemma 4,

$$N_n = b_n \sum_{j=0}^{j_0} b_n^j \mathcal{O}_P(\sqrt{n}) + b_n \sum_{j=j_0+1}^{r_0-1} b_n^j \mathcal{O}_P(\sigma_{n,r_0-j}) + b_n^{r_0} \mathcal{O}_P(\sigma_{n,1}) = o_P(\sqrt{nb_n}),$$

which is actually similar to the former case $r_0 < (2\beta-1)^{-1}$.

(ii) We proceed similarly by considering $r_0 < (2\beta-1)^{-1}$ and $r_0 \geq (2\beta-1)^{-1}$ separately. For the former case by the conditions imposed in (ii), we have

$$\max_{0 \leq j \leq r_0-2} b_n^{j+1}\sigma_{n,r_0-j} = o(b_n^{r_0}\sigma_{n,1}) \quad \text{and} \quad \sqrt{nb_n} = o(b_n^{r_0}\sigma_{n,1}).$$

By (2) and (7), conclusion (16) follows from

$$\frac{nb_n}{b_n^{r_0}\sigma_{n,1}}[f_n(x_0) - \mathbb{E} f_n(x_0)] = \frac{Y_{n,1}}{\sigma_{n,1}} c_{r_0} + o_P(1),$$

where $c_r = (-1)^r f^{(r)}(x_0) \int_{\mathbb{R}} K(v) v^{r-1}/(r-1)!\, dv$.

The latter case $r_0 \geq (2\beta-1)^{-1}$ can be dealt with similarly as in the proof of (i) via Lemma 4.

(iii) Under the imposed conditions, we have $\beta \leq (3+2r_0)/(2+4r_0)$ (see Figure 1) which implies $(2\beta-1)^{-1} \geq r_0 + 1/2 > r_0$ and thus the conditions of Lemma 4 are satisfied for $j = 0$. Moreover,

$$\max_{1 \leq j \leq r_0-1} b_n^{j+1}\sigma_{n,r_0-j} = o(b_n\sigma_{n,r_0}) \quad \text{and} \quad \sqrt{nb_n} = o(b_n\sigma_{n,r_0}).$$

Then by Lemma 4, (17) results from (22), (2) and (7). $\square$

PROOF OF THEOREM 4. By Avram and Taqqu (1987) and the decomposition (2), part (c) follows from (22) and (7) and Lemma 4 after elementary manipulations. The form of coefficient pertaining to $Z_{r_0-j,\beta}$ in (19) follows from the asymptotic equality

$$\frac{1}{b_n^{r_0-j-1}\sigma_{n,1}} \sim \frac{1}{\sigma_{n,r_0-j}}\left\{\frac{C(\beta,r_0-j)}{C^{r_0-j-1}(\beta,1)}\right\}\frac{1}{c^{r_0-j-1}}$$

for $j=0,1,\dots,r_0-1$. Now we prove only (a) since (b) can be obtained in a similar manner. So let $r_0=1$. Let $s_n=\sum_{i=0}^n a_i$ for $n\ge 0$ and $s_n=0$ for $n<0$. Recall the proof of Lemma 2 for the definitions of $\xi_{n,t}$ and $\zeta_{n,t}$. For any fixed nonzero $\lambda_1,\lambda_2\in\mathbb{R}$ let

$$\eta_n(t)=\lambda_1\xi_{n,t}+\frac{\lambda_2}{\sigma_{n,1}}\varepsilon_t s_{n-t}, \qquad t=1,2,\dots,n$$

and

$$\eta_n(t)=\frac{\lambda_2}{\sigma_{n,1}}\varepsilon_t(s_{n-t}-s_{-t}), \qquad t=0,-1,-2,\dots.$$

By the Cramér–Wold device in view of the proof of Theorem 2 it suffices to show the asymptotic normality $N(0,\lambda_1^2\sigma^2(x_0)+\lambda_2^2)$ of $\sum_{t=-\infty}^n\eta_n(t)$. Observe that the sequence $\{\eta_n(t)\}_{t=-\infty}^n$ is a (triangular) martingale difference w.r.t. $\widetilde{X}_t$. Thus by the martingale central limit theorem, we only need to verify the convergence of the conditional variances

$$\sum_{t=-\infty}^{n}\mathbb{E}[\eta_n^2(t)\mid\widetilde{X}_{t-1}]\xrightarrow{\mathcal{P}}\lambda_1^2\sigma^2(x_0)+\lambda_2^2 \tag{23}$$

and the Lindeberg condition

$$\sum_{t=-\infty}^{0}+\sum_{t=1}^{n}\mathbb{E}\big[\eta_n^2(t)\mathbf{1}_{|\eta_n(t)|>\varepsilon}\big]=o(1) \qquad \text{for any } \varepsilon>0. \tag{24}$$

Clearly, in view of the proof of Lemma 2, the convergence of the conditional variances (23) follows from the following relation for the cross-product terms:

$$\sum_{t=1}^{n}\mathbb{E}\left|\xi_{n,t}\frac{\varepsilon_t s_{n-t}}{\sigma_{n,1}}\right|=\mathcal{O}(\sqrt{b_n})\to 0.$$

The last claim is an immediate consequence of $s_n=\mathcal{O}[n^{1-\beta}L(n)]$ which follows from Karamata's theorem [see, e.g., Theorem 0.6 in Resnick (1987)] and

$$\begin{aligned}\mathbb{E}\left|\varepsilon_1 K\left(\frac{x_0-X_1}{b_n}\right)\right| &= \iint_{\mathbb{R}^2}\left|uK[b_n^{-1}(x_0-u-v)]\right|f_1(u)\tilde{f}_1(v)\,du\,dv\\ &= b_n\iint_{\mathbb{R}^2}|uK(z)|f_1(u)\tilde{f}_1(x_0-u-b_nz)\,du\,dz\\ &\le b_nC\int_{\mathbb{R}^2}|uK(z)|f_1(u)\,du\,dz=\mathcal{O}(b_n),\end{aligned}$$

where $\tilde{f}_1 = \tilde{F}_1^{(1)}$ is the density function of R_t with an upper bound C in view of Lemma 1 and $\sigma_{n,1}^{-1} \sum_{t=1}^{n} s_{n-t+1} = \mathcal{O}(n^{1/2})$.

We check now Lindeberg condition (24). Elementary computations show that the first summand in (24) is $o(1)$. For the second one, let $\phi_n(t) = \lambda_1 \zeta_{n,t} + \lambda_2 \varepsilon_t s_{n-t}/\sigma_{n,1}$, $1 \le t \le n$. Then $\eta_n(t) = \phi_n(t) - \mathbb{E}[\phi_n(t) \mid \widetilde{X}_{t-1}]$. In view of Corollary 9.5.2 in Chow and Teicher (1988) it remains to show that

$$\sum_{t=1}^{n} \mathbb{E}\big[\phi_n^2(t)\mathbf{1}_{[|\phi_n(t)|>\varepsilon]}\big] = o(1) \qquad \text{for any } \varepsilon > 0.$$

To this end, let

$$G(n) = \min\left\{\left|\frac{\varepsilon\sqrt{nb_n}}{2\lambda_1}\right|, \left|\frac{\varepsilon\sigma_{n,1}}{2\lambda_2 s_{n-t}}\right|, \ t = 1, 2, \dots, n\right\}.$$

Then $\lim_{n\to\infty} G(n) = \infty$ and for all $1 \le t \le n$,

$$\mathbf{1}_{[|\phi_n(t)|>\varepsilon]} \le \mathbf{1}_{[|K[b_n^{-1}(x_0 - X_t)]| \ge G(n)]} + \mathbf{1}_{[|\varepsilon_t| \ge G(n)]}.$$

Applying the inequality $(A+B)^2 \le 2(A^2 + B^2)$, we have

$$\phi_n^2(t) \le 2\lambda_1^2 \zeta_{n,t}^2 + 2\lambda_2^2 \sigma_{n,1}^{-2} [\varepsilon_t s_{n-t}]^2.$$

In view of Karamata's theorem, $\sum_{i=1}^{n} s_{n-t}^2/\sigma_{n,1}^2 = \mathcal{O}(1)$. Thus

$$\sum_{t=1}^{n} E\big\{\sigma_{n,1}^{-2}[\varepsilon_t s_{n-t}]^2 \mathbf{1}_{[|\varepsilon_t| \ge G(n)]}\big\} \le E\big\{\varepsilon_t^2 \mathbf{1}_{[|\varepsilon_t| \ge G(n)]}\big\} = o(1)$$

and since f is bounded

$$\sum_{t=1}^{n} E\big\{\zeta_{n,t}^2 \mathbf{1}_{|K[b_n^{-1}(x_0 - X_1)]| \ge G(n)}\big\} = \mathcal{O}(1) \int_{|K(z)| \ge G(n)} K^2(z)\, dz = o(1).$$

Thus we can now conclude the proof of the Lindeberg condition by first noting that

$$\begin{aligned}
&\mathbb{E}\left|K^2\left(\frac{x_0 - X_t}{b_n}\right)\mathbf{1}_{[|\varepsilon_t| \ge G(n)]}\right| \\
&\qquad = \int_{|u| \ge G(n)} \int_{\mathbb{R}} K^2[b_n^{-1}(x_0 - u - v)] f_1(u) \tilde{f}_1(v)\, du\, dv \\
&\qquad = b_n \int_{|u| \ge G(n)} \int_{\mathbb{R}} K^2(z) f_1(u) \tilde{f}_1(x_0 - u - b_n z)\, dz\, du \\
&\qquad = \mathcal{O}(b_n) \int_{|u| \ge G(n)} f_1(u)\, du = o(b_n)
\end{aligned}$$

and then, reasoning as before,

$$
\begin{aligned}
&\sum_{t=1}^{n} \mathbb{E}\{[\sigma_{n,1}^{-1}\varepsilon_t s_{n-t}]^2 \mathbf{1}_{|K[b_n^{-1}(x_0-X_t)]|\ge G(n)}\} \\
&\qquad \le \mathbb{E}\{\varepsilon_1^2 \mathbf{1}_{|K[b_n^{-1}(x_0-X_1)]|\ge G(n)}\} \\
&\qquad = b_n \iint_{|K(z)|\ge G(n)} u^2 f_1(u)\tilde{f}_1(x_0-u-b_n z)\,dz\,du \\
&\qquad = \mathcal{O}(b_n)\int_{|K(z)|\ge G(n)} dz \\
&\qquad \le \mathcal{O}(b_n)G^{-2}(n)\int_{\mathbb{R}} K^2(s)\,ds = o(b_n). \qquad \square
\end{aligned}
$$

6. Conclusion. Clearly Theorem 4 does not cover the cases $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{3c} \cap \mathbf{T}_{10}$ and $(\mathbf{b}, \beta, L(\cdot)) \in \mathbf{T}_{3c} \cap \mathbf{T}_{1c'}$ for some $c, c' > 0$. The latter is more interesting and it corresponds to the case when bandwidths satisfy the hyperbolic decay condition of Corollary 1 to the situation that (β, α) is exactly the common joint point $\mathcal{T}$. We conjecture that in this situation the limiting distribution is a convolution of random normal variable and multiple Wiener–Itô integrals.

Acknowledgments. The comments of two referees and an Associate Editor, which helped improve former versions of this paper, are gratefully acknowledged.

REFERENCES

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

BERAN, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

BOSQ, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. *Lecture Notes in Statist.* **110**. Springer, New York.

CARBON, M. and TRAN, L. T. (1996). On histograms of linear processes. *J. Statist. Plann. Inference* **53** 403–419.

CASTELLANA, J. V. and LEADBETTER, M. R. (1986). On smoothed probability density estimation. *Stochast. Process. Appl.* **21** 179–193.

CHANDA, K. C. (1983). Density estimation for linear processes. *Ann. Inst. Statist. Math.* **35** 439–446.

CHANDA, K. C. (1995). Corrigendum to density estimation for linear processes. Unpublished note.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory*, 2nd ed. Springer, New York.

COULON-PRIEUR, C. and DOUKHAN, P. (2000). A triangular central limit theorem under a new weak dependence condition. *Statist. Probab. Lett.* **47** 61–68.

CSÖRGŐ, S. and MIELNICZUK, J. (1995). Density estimation under long-range dependence. *Ann. Statist.* **23** 990–999.

DOUKHAN, P. (1994). *Mixing*: *Properties and Examples*. Springer, 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. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *J. Statist. Plann. Inference* **80** 81–93.

GORODETSKII, V. V. (1977). On the strong mixing property for linear sequences. *Theory Probab. Appl.* **22** 411–412.

HALL, P. and HART, J. (1990). Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory Related Fields* **87** 253–274.

HALLIN, M. and TRAN, L. T. (1996). Kernel density estimation for linear processes: Asymptotic normality and optimal bandwidth derivation. *Ann. Inst. Statist. Math.* **48** 430–448.

HO, H. C. (1996). On central and non-central limit theorems in density estimation for sequences of long-range dependence. *Stochastic Process. Appl.* **63** 153–174.

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.

HONDA, T. (2000). Nonparametric density estimation for a long-range dependent linear process. *Ann. Inst. Statist. Math.* **52** 599–611.

PARZEN, E. (1962). On the estimation of a probability density and the mode. *Ann. Math. Statist.* **33** 1965–1976.

PHAM, T. D. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303.

RESNICK, S. I. (1987). *Extreme Values*, *Regular Variation and Point Processes*. Springer, New York.

RIO, E. (1995). About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Statist.* **1** 35–61.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

ROSENBLATT, M. (1970). Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 199–213. Cambridge Univ. Press.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

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

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

TRAN, L. T. (1992). Kernel density estimation for linear processes. *Stochastic Process. Appl.* **26** 281–296.

WITHERS, C. S. (1981). Conditions for linear process to be strongly mixing. *Z. Wahrsch. Verw. Gebiete* **57** 477–480.

WOODROOFE, M. (1970). Discussion on Rosenblatt's paper. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 199–213. Cambridge Univ. Press, London.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 S. UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637
E-MAIL: wbwu@galton.uchicago.edu

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
ORDONA 21
01–237 WARSAW
POLAND
E-MAIL: miel@ipipan.waw.pl