The Annals of Statistics
2002, Vol. 30, No. 2, 397–430

WEAK DEPENDENCE BEYOND MIXING AND ASYMPTOTICS FOR NONPARAMETRIC REGRESSION

BY PATRICK ANGO NZE, PETER BÜHLMANN AND PAUL DOUKHAN

Université Lille 3, ETH Zürich and Université Cergy Pontoise

We consider a new concept of weak dependence, introduced by Doukhan and Louhichi [*Stochastic Process. Appl.* **84** (1999) 313–342], which is more general than the classical frameworks of mixing or associated sequences. The new notion is broad enough to include many interesting examples such as very general Bernoulli shifts, Markovian models or time series bootstrap processes with discrete innovations.

Under such a weak dependence assumption, we investigate nonparametric regression which represents one (among many) important statistical estimation problems. We justify in this more general setting the "whitening by windowing principle" for nonparametric regression, saying that asymptotic properties remain in first order the same as for independent samples. The proofs borrow previously used strategies, but precise arguments are developed under the new aspect of general weak dependence.

1. Introduction. The analysis of statistical procedures for dependent data usually relies on some decay of dependence as the distance, say in time or space, between observations increases. The most popular notions describing such a decay of dependence are from the framework of mixing sequences [10]. However, mixing conditions can be very hard to verify for particular models or are even too strong to be true. Not much is known in asymptotic theory about the behavior of a statistical procedure when the data generating stationary process exhibits dependencies which are beyond classical mixing or, as another example for a framework describing dependence, beyond association (see Section 3.4). Examples where mixing or association fails to hold include the following: (i) Bernoulli shifts driven by discrete innovations, (ii) Markov processes driven by discrete innovations and (iii) processes arising from model- or sieve-based time series bootstraps. More precise definitions are given in Section 3.

Doukhan and Louhichi [12] have introduced a new concept of weak dependence for stationary processes which generalizes the notions of mixing and association. Relaxation of mixing or association conditions and assuming only the new notion of weak dependence yields a fairly tractable framework for the analysis of

---

Received July 1999; revised September 2001.

AMS 2000 subject classifications. Primary 60F05, 62M99; secondary 60E15, 60G10, 60G99, 62G07, 62G09.

Key words and phrases. Bernoulli shift, bootstrap, central limit theorem, Lindeberg method, Markov process, mixing, nonparametric estimation, positive dependence, stationary sequence, time series.

statistical procedures with very general data generating processes, for example, examples (i)–(iii) mentioned above.

For a stationary time series, weak dependence as defined in [12] is measured in terms of covariances of functions. For convenient functions h and k, we assume that

$$\mathrm{Cov}\big(h(\text{"past"}), k(\text{"future"})\big) \tag{1.1}$$

is small when the distance between the "past" and the "future" is sufficiently large. Regarding the functions h and k, we focus on the class of bounded Lipschitz functions and modifications thereof. This class is small enough to actually prove for various processes that the quantities in (1.1) can be bounded uniformly over the function class as a function of the distance between the "past" and the "future." On the other hand, the function class is rich enough to obtain high order moment bounds and central limit theorems (CLTs) for sums whenever suitable uniform bounds for (1.1) hold. Alternative function classes are also possible and another similar proposal for weak dependence has been given in [4]; see Section 2.2.

Under the general notion of weak dependence in terms of requiring only a suitable (uniform) decay of the covariances in (1.1), it is still possible to get fairly good bounds for moment and exponential inequalities and the CLT still holds (see Section 2.3). However, these bounds are usually less tight than for mixing or associated sequences. Another complication arises when dealing with transformed values $g(Z_t, \dots, Z_{t-v})$ for fixed v, where $g(\cdot)$ is nonsmooth: the covariance bound in (1.1) controls only for transforms $g(\cdot)$ which are bounded Lipschitz (or modifications thereof), although nonsmooth $g(\cdot)$'s can be handled under additional concentration assumptions for the process $(Z_t)_{t\in\mathbb{Z}}$. Some of the strategies for various proofs can be borrowed from previous work in mixing or association, but the difficulties mentioned above indicate that precise arguments have to be developed under the new aspect of weak dependence.

As *one* important example of an estimation problem, we focus on point and interval estimation for conditional expectations. Consider a strictly stationary process $(Z_t)_{t\in\mathbb{Z}}$ taking values in $\mathbb{R}^D$ for some $D \in \mathbb{N}$: $(Z_t)_{t\in\mathbb{Z}}$ does not need to satisfy a mixing or association, but only a weak dependence assumption as mentioned above. For $2 \le D = 1 + d$, writing the components as $Z_t = (X_t, Y_t)$ with $X_t \in \mathbb{R}^d$ and $Y_t \in \mathbb{R}$, the problem of interest here is nonparametric point and interval estimation of the function

$$r(x) = \mathbb{E}(Y_0 | X_0 = x).$$

For simplicity, we often restrict ourselves to $D = 2$ with $X_t \in \mathbb{R}$, but extensions to $D = 1 + d > 2$ with $X_t \in \mathbb{R}^d$ are straightforward. Known estimation techniques for $r(x)$, namely the kernel estimator and a local bootstrap thereof are shown to have essentially the same first order asymptotic properties as in the independent case. This phenomenon is known as the *whitening by windowing* principle and was first proved for mixing processes in [29]. There seem to be few difficulties extending our results to other smoothing techniques such as local polynomial estimators. It indicates validity of this principle for many windowing and smoothing methods applied to very general, stationary weakly dependent observations, including nonmixing or nonassociated sequences.

The paper is organized as follows. The definition and tools for weak dependence are given in Section 2; examples of processes where weak dependence holds are given in Section 3; Section 4 describes point and interval estimators for conditional expectations; asymptotic properties are given in Section 5; Section 6 contains the proofs.

2. Weak dependence.

2.1. *Definition.* We define here the new notion of dependence, thereby closely following [12]. Generally, let E be some normed measurable space with norm $\|\cdot\|$, although we restrict later attention to the case where $E = \mathbb{R}^D$. Denote by $\mathbb{L}^\infty(E^u)$ $(u \in \mathbb{N})$ the set of measurable bounded functions on E^u and set $\mathbb{L}^\infty = \bigcup_{u=1}^\infty \mathbb{L}^\infty(E^u)$. We then equip E^u with the l^1-norm $\|(x_1, \dots, x_u)\|_1 = \|x_1\| + \cdots + \|x_u\|$, where $x \in E^u$. Moreover, denote by

$$\mathrm{Lip}(h) = \sup_{x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|_1}$$

the Lipschitz modulus of a function h: $E^u \to \mathbb{R}$ with respect to the l^1-norm in E^u. Define

$$\mathcal{L} = \bigcup_{u=1}^\infty \{h \in \mathbb{L}^\infty(E^u);\ \mathrm{Lip}(h) < \infty,\ \|h\|_\infty \le 1\}.$$

DEFINITION 1 [12]. The E-valued sequence $(Z_t)_{t\in\mathbb{Z}}$ is called $(\theta, \mathcal{L}, \psi)$-weak dependent if for some monotone sequence $\theta = (\theta_r)_{r\in\mathbb{N}}$ decreasing to zero at infinity and some real-valued function ψ with arguments $(h, k, u, v) \in \mathcal{L}^2 \times \mathbb{N}^2$,

$$\left|\mathrm{Cov}\left(h(Z_{i_1}, \dots, Z_{i_u}), k(Z_{j_1}, \dots, Z_{j_v})\right)\right| \le \psi(h, k, u, v)\theta_r$$

for any u-tuple $(i_1, \dots, i_u)$, any v-tuple $(j_1, \dots, j_v)$ with $i_1 \le \cdots \le i_u < i_u + r \le j_1 \le \cdots \le j_v$ and all $u, v \in \mathbb{N}$.

Various choices for the function ψ turn out to be convenient. Particularly, consider

$$\psi_1(h, k, u, v) = u\,\mathrm{Lip}(h) + v\,\mathrm{Lip}(k), \qquad \psi_1'(h, k, u, v) = v\,\mathrm{Lip}(k),$$
$$\psi_2(h, k, u, v) = uv\,\mathrm{Lip}(h)\,\mathrm{Lip}(k), \qquad \psi_2'(h, k, u, v) = v\,\mathrm{Lip}(h)\,\mathrm{Lip}(k),$$

where the functions h and k are defined on E^u and E^v, respectively ($u, v \in \mathbb{N}$). Clearly $(\theta, \mathcal{L}, \psi'_j)$-weak dependence implies $(\theta, \mathcal{L}, \psi_j)$-weak dependence ($j = 1, 2$). The functions ψ_j ($j = 1, 2$) are adapted to provide moment bounds by using techniques in [12], whereas the functions ψ'_j ($j = 1, 2$) provide nice CLTs via the Lindeberg–Rio method used, for example, in [9]. The distinction with the subscripts 1 and 2 corresponds to our examples, where we always consider ψ_1, ψ'_1 for Bernoulli shifts and Markov processes, and ψ_2, ψ'_2 for associated sequences (see Section 3).

If the class $\mathcal{L}$ is replaced by $\mathbb{L}^\infty$ and $\psi(h, k, u, v) = 4\|h\|_\infty \|k\|_\infty$, one obtains strong mixing processes with $\theta_r = \alpha_r$ as defined by Rosenblatt (cf. [10]). However, such strong mixing conditions refer to the total variation norm of two distributions rather than an appropriate distance between random variables. This is often an unnecessarily strong requirement. In Section 3, we will discuss some examples where the mixing condition is too restrictive.

The notion of $\mathcal{L}$-weak dependence can also be modified to deal with indicator functions, which are not Lipschitzian, and empirical processes. In the latter case with indicators of half-lines in $\mathbb{R}$, we consider instead of $\mathcal{L}$ the class

$$\mathcal{I} = \bigcup_{u=1}^{\infty} \left\{ \bigotimes_{i=1}^{u} g_{x_i};\ x_i \in \mathbb{R}^+ \text{ for } i = 1, \ldots, u \right\},$$

where $g_x(y) = \mathbb{1}_{\{x \le y\}} - \mathbb{1}_{\{y \le -x\}}$, $x \in \mathbb{R}^+$. Under $\mathcal{L}$-weak dependence and additional regularity assumptions for the underlying process, $\mathcal{I}$-weak dependence can be established. For instance, under smoothness conditions for the distribution of the process, the following uniform covariance bounds (as in Definition 1) over $\mathcal{I}$ can be established: $v\sqrt{u+v}\theta_r$, $(u+v)^{4/3}\theta_r^{1/3}$ and $(u+v)\theta_r^{1/3}$ in the $(\theta, \mathcal{L}, \psi'_1)$-, $(\theta, \mathcal{L}, \psi_2)$- and $(\theta, \mathcal{L}, \psi'_2)$-weak dependent cases, respectively.

2.2. *Relation to ν-mixing.* As discussed above, it is sometimes desirable to bound covariances of non-Lipschitz functions. In [4], another type of weak dependence, called ν-mixing, was introduced. Similar to Definition 1, uniform covariance bounds over classes of functions with smooth averaged modulus of continuity are required. This framework is closely related to the theory of weak convergence as in Bhattacharya and Ranga Rao [3]. Example 3.3 in [4] explains that weak dependence as in Definition 1 implies ν-mixing; the reverse implication is generally not true. For a given process, it is therefore harder to prove weak dependence than ν-mixing. On the other hand, within the framework of weak dependence, covariance and exponential inequalities and CLTs for sums have been established [12], whereas for the more general notion of ν-mixing, only covariance bounds have been derived [4].

2.3. *Available tools under weak dependence.* We briefly review the most important tools when dealing with a stationary process $(Z_t)_{t\in\mathbb{Z}}$ satisfying a suitable weak dependence condition with coefficient θ_r.

Rosenthal's inequality, bounding higher moments of sums, becomes

$$\left|\mathbb{E}\left(\sum_{t=1}^{n}(Z_t - \mathbb{E}[Z_0])\right)^q\right| \\ \leq A_q \max\left\{\left(Cn\sum_{r=0}^{n-1}\theta_r\right)^{q/2}, BM^{q-2}n\sum_{r=0}^{n-1}(r+1)^{q-2}\theta_r\right\}, \\ q \text{ an integer } \geq 2,$$

where A_q, B and C are positive constants, and the centered random variables $|Z_t - \mathbb{E}[Z_0]| \leq M$ are bounded; see [12]. An exponential inequality, assuming $|Z_t - \mathbb{E}[Z_0]| \leq M$ and $\mathbb{E}|Z_0 - \mathbb{E}[Z_0]|^2 \leq \sigma^2 < \infty$, looks as follows:

$$\mathbb{P}\left[\left|\sum_{t=1}^{n}(Z_t - \mathbb{E}[Z_0])\right| \geq x\sigma\sqrt{n}\right] \leq B\exp(-A\sqrt{\beta x})$$

for universal positive constants A, B, and $\beta > 0$ is a constant depending on the decay of θ_r; see [12]. The difficulty with both inequalities is the complicated dependence of their constants on the decay of the weak dependence coefficients θ_r, yielding less tight bounds than in the mixing framework.

Central limit theorems

$$n^{-1/2}\sum_{t=1}^{n}(Z_t - \mathbb{E}[Z_0]) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \lim_n\left(n^{-1}\operatorname{Var}\left(\sum_{t=1}^{n}Z_t\right)\right)\right)$$

can be established using either Bernstein's blocking technique or Lindeberg's method. We usually prefer the latter, as described in [28], since it often works under slightly weaker conditions.

From an asymptotic view as $n \to \infty$, we have qualitatively the same behavior as in the mixing framework, although with different constants for inequalities. Therefore, many strategies for proofs assuming mixing conditions carry over to the more general setting of weak dependence. However, as already indicated, the different constants require careful arguments which, unfortunately, have to be given in a case-by-case manner.

3. Examples. We present here examples where weak dependence, as defined in Section 2.1, holds. First, we focus on some general classes of processes and will then specialize to specific models.

3.1. *Bernoulli shifts.*

DEFINITION 2. Let $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of real-valued random variables and let $F\colon \mathbb{R}^{\mathbb{Z}} \to E$ be a measurable function. The sequence $(Z_t)_{t\in\mathbb{Z}}$ defined by

$$Z_t = F(\xi_{t-j} : j \in \mathbb{Z}) \tag{3.1}$$

is called a Bernoulli shift.

The class of Bernoulli shifts is very general. It provides examples of processes that are weakly dependent but not mixing (see [30]). A bound for the decay of weak dependence can be obtained as follows.

For any $k \in \mathbb{N}$, let $\delta_k \in \mathbb{R}$ be such that

$$\sup_{t\in\mathbb{Z}} \mathbb{E}\| F(\xi_{t-j} : j \in \mathbb{Z}) - F\big(\xi_{t-j}\mathbb{1}_{\{|j|<k\}} : j \in \mathbb{Z}\big)\| \le \delta_k. \tag{3.2}$$

Such sequences $(\delta_k)_{k\in\mathbb{N}}$ are related to the modulus of uniform continuity of F. The sequence $(\delta_k)_{k\in\mathbb{N}}$ can be evaluated under regularity conditions on the function F. For example, if

$$\|F(u_j : j \in \mathbb{Z}) - F(v_j : j \in \mathbb{Z})\| \le \sum_{j\in\mathbb{Z}} a_j |u_j - v_j|^b,$$

for positive constants $(a_j)_{j\in\mathbb{Z}}$ and some $0 < b \le 1$, and if the sequence ξ_t has finite bth-order moment for all t, we can choose $\delta_k = \sum_{|j|\ge k} a_j \mathbb{E}|\xi_j|^b$. Bernoulli shifts then satisfy the condition of

$$(\theta, \mathcal{L}, \psi_1)\text{-weak dependence with } \theta_r = \delta_{r/2},$$

(see [12], Corollary 2 and Lemma 8). In the case of a causal Bernoulli shift where $Z_t = F(\xi_{t-j}, j \in \mathbb{N}_0)$, then also $(\theta, \mathcal{L}, \psi_1')$-weak dependence holds with $\theta_r = \delta_r$.

3.2. *Markov processes.* Markov processes can be represented as Bernoulli shifts. Consider an $\mathbb{R}^D$-valued Markov process, driven by the recurrence equation

$$Z_t = f(Z_{t-1}, \xi_t) \qquad (t \in \mathbb{Z}) \tag{3.3}$$

for some i.i.d. sequence $(\xi_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}(\xi_0) = 0$, ξ_t independent of $\{Z_s;\ s < t\}$ and $f\colon \mathbb{R}^D \times \mathbb{R} \to \mathbb{R}^D$. Then the function F in (3.1) is defined implicitly (if it exists) by the relation

$$F(\xi) = f\big(F(\xi'), \xi_t\big) \qquad \text{where } \xi = (\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots),$$
$$\xi' = B\xi = (\xi_{t-1}, \xi_{t-2}, \xi_{t-3}, \ldots),$$

with B denoting the backshift operator.

Assume now in representation (3.3) that Z_0 is independent of the sequence $(\xi_t)_{t\in\mathbb{N}}$. Suppose that, for some $0 \le c_i < 1$,

$$\begin{aligned}&\mathbb{E}|f(0,\xi_1)| < \infty \quad \text{and} \quad \mathbb{E}|f(u,\xi_1) - f(v,\xi_1)| \le \sum_{i=1}^{d} c_i |u_i - v_i|,\\ &c = \sum_{i=1}^{d} c_i < 1 \qquad \text{for all } u, v \in \mathbb{R}^D.\end{aligned} \tag{3.4}$$

Duflo [14] shows under the condition in (3.4) that the Markov process $(Z_t)_{t\in\mathbb{N}}$ has a stationary distribution μ with finite first moment. Assume now in addition that Z_0 is distributed with μ, that is, the Markov chain is stationary. Then, if (3.4) holds, such a Markov chain is

$$(\theta, \mathcal{L}, \psi_1')\text{-weak dependent with } \theta_r = c^r \mathbb{E}|Z_0|.$$

See [12].

3.3. *Chaotic representations.* We now specialize the Bernoulli-shift representation in (3.1) to expansions associated with the discrete chaos generated by the sequence $(\xi_t)_{t\in\mathbb{Z}}$. In a condensed formulation we write

$$\begin{aligned}F(u) &= \sum_{k=0}^{\infty} F_k(u) \qquad (u \in \mathbb{R}^{\mathbb{Z}}),\\ F_k(u) &= \sum_{j_1\in\mathbb{Z}} \sum_{j_2\in\mathbb{Z}} \cdots \sum_{j_k\in\mathbb{Z}} a^{(k)}_{j_1,\dots,j_k} u_{j_1} u_{j_2} \cdots u_{j_k} \qquad (k \ge 1),\end{aligned} \tag{3.5}$$

where $F_k(u)$ denotes the kth-order chaos contribution and $F_0(u) = a_0^{(0)}$ is only a centering constant. In short we write, in vector notation, $F_k(u) = \sum_{j\in\mathbb{Z}^k} a_j^{(k)} u_j$. Processes associated with a finite number of chaos, that is, $F_k(u) \equiv 0$ if $k > k_0$ for some $k_0 \in \mathbb{N}$, are also called Volterra processes. A first example of Volterra processes is the class of linear processes, including the popular ARMA models. It corresponds to the expansion in (3.5) with $F_k(u) \equiv 0$ for all $k > 1$.

A simple and general condition for $\mathbb{L}^1$-convergence of the expansion in (3.5), still written in a condensed notation, is $\sum_{k=0}^{\infty} \sum_{j\in\mathbb{Z}^k} |a_j^{(k)}|\ \mathbb{E}|\xi_0|^k < \infty$, which allows us to define the distribution of such shift processes. Then

$$\begin{aligned}&(\theta, \mathcal{L}, \psi_1)\text{-weak dependence holds with } \theta_r = \delta_{r/2},\\ &\delta_r = \sum_{k=0}^{\infty} \sum_{j\in\mathbb{Z}^k;\, \|j\|_\infty > r} |a_j^{(k)}|\ \mathbb{E}|\xi_0|^k < \infty.\end{aligned}$$

Under causality (see the end of Section 3.1), $(\theta, \mathcal{L}, \psi_1')$-weak dependence also holds with $\theta_r = \delta_r$.

3.4. *Associated sequences.*

DEFINITION 3 [15]. The sequence $(Z_t)_{t\in\mathbb{Z}}$ is associated, if for all coordinatewise increasing real-valued functions h and k,

$$\operatorname{Cov}\bigl(h(Z_t : t \in A), k(Z_t : t \in B)\bigr) \geq 0$$

for all finite subsets A and B of $\mathbb{Z}$.

Associated sequences are

$$(\theta, \psi_2, \mathcal{L})\text{-weak dependent with } \theta_r = \sup_{k\geq r} \operatorname{Cov}(X_0, X_r);$$

see [12]. Note that broad classes of examples of associated processes result from the fact that any independent sequence is associated and that monotonicity preserves association (cf. [23]).

The case of Gaussian sequences is analogous by setting $\theta_r = \sup_{k\geq r} |\operatorname{Cov}(X_0, X_k)|$; see [12]. For associated or Gaussian sequences, ψ_2'-weak dependence also holds with $\theta_r = \sum_{k\geq r} |\operatorname{Cov}(X_0, X_k)|$.

3.5. *More specific examples.*

Nonparametric AR model. Consider the real-valued functional (nonparametric) autoregressive model

$$Z_t = r(Z_{t-1}) + \xi_t,$$

where $r\colon \mathbb{R} \to \mathbb{R}$ and $(\xi_t)_{t\in\mathbb{Z}}$ as in (3.3). This a special example of a Markov process in (3.3). Assume that $|r(u) - r(u')| \leq c|u - u'|$ for all $u,\ u' \in \mathbb{R}$ and for some $0 \leq c < 1$, and $\mathbb{E}|\xi_0| < \infty$. Then (3.4) with $D = 1$ holds and implies

$$(\theta, \mathcal{L}, \psi_1')\text{-weak dependence with } \theta_r = \delta_r = c^r \mathbb{E}|Z_0|.$$

We emphasize here that the marginal distribution of the innovations ξ_t can be discrete. In such a case, classical mixing properties can fail to hold.

As an example, consider the simple linear AR(1) model,

$$Z_t = \phi Z_{t-1} + \xi_t = \sum_{j\geq 0} \phi^j \xi_{t-j}, \qquad |\phi| < 1.$$

Let $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. Bernoulli variables with parameter $s = \mathbb{P}[\xi_t = 1] = 1 - \mathbb{P}[\xi_t = 0]$. The AR(1) process $(Z_t)_{t\in\mathbb{Z}}$ with innovations $(\xi_t)_{t\in\mathbb{Z}}$ and AR parameter $\phi \in]0, \frac{1}{2}]$ is $(\theta, \mathcal{L}, \psi_1')$-weak dependent with $\theta_r = \delta_r = \phi^r \mathbb{E}|Z_0|$, but it is known to be nonmixing (cf. [30]). Note that concentration holds; for example, Z_t is uniform if $s = \frac{1}{2}$ and it has a Cantor marginal distribution if $s = \frac{1}{3}$. Hence, without a regularity condition on the marginal distribution of ξ_0, Bernoulli shifts or Markov processes may not be mixing.

Nonparametric ARCH model. Consider the real-valued functional (nonparametric) ARCH model

$$Z_t = s(Z_{t-1})\xi_t,$$

where $s\colon \mathbb{R} \to \mathbb{R}^+$ and $(\xi_t)_{t\in\mathbb{Z}}$ as in (3.3) with $\mathbb{E}|\xi_0|^2 = 1$. This is a special example of a Markov process in (3.3) with $f(u, v) = s(u)v$. Assume that $|s(u) - s(u')| \le c|u - u'|$ for all $u,\ u' \in \mathbb{R}$ and for some $0 \le c < 1$. Then (3.4) with $D = 1$ holds and implies

$$(\theta, \mathcal{L}, \psi_1')\text{-weak dependence with } \theta_r = c^r \mathbb{E}|Z_0|.$$

Again, the innovation distribution is allowed to be discrete.

Nonparametric AR–ARCH model. An often used combination of the two models above is a process having nonparametric conditional mean and variance structure,

$$Z_t = r(Z_{t-1}) + s(Z_{t-1})\xi_t,$$

with $r(\cdot)$, $s(\cdot)$ and $(\xi_t)_{t\in\mathbb{Z}}$ as in the examples above. Assume the Lipschitz conditions as above for $r(\cdot)$ and $s(\cdot)$ with constants c_r and c_s, respectively. If $c_r + c_s = c < 1$, the process satisfies

$$(\theta, \mathcal{L}, \psi_1')\text{-weak dependence with } \theta_r = c^r \mathbb{E}|Z_0|.$$

Bilinear model. Consider the simple bilinear process with recurrence equation

$$Z_t = aZ_{t-1} + bZ_{t-1}\xi_{t-1} + \xi_t,$$

where $(\xi_t)_{t\in\mathbb{Z}}$ is as in (3.3). Such causal processes are associated with the chaotic representation in (3.5) with

$$F(u) = \sum_{j=0}^{\infty} u_j \prod_{s=1}^{j} (a + bu_s), \qquad u = (u_0, u_1, u_2, \ldots).$$

Under Tong's [34] stationarity condition and if $c = \mathbb{E}|a + b\xi_0| < 1$, the process satisfies

$$(\theta, \mathcal{L}, \psi_1')\text{-weak dependence with } \theta_r = \frac{c^r(r+1)}{1-c}.$$

AR-sieve bootstrap for time series. The AR-sieve bootstrap for $\mathbb{R}$-valued time series $(Z_t)_{t=1}^n$ resamples from an estimated autoregressive model of order $p = p_n \to \infty,\ p_n/n \to 0\ (n \to \infty)$, defined recursively by

$$Z_t^* = \sum_{j=1}^{p_n} \hat{\phi}_j Z_{t-j}^* + \xi_t^* \qquad (t \in \mathbb{Z}),$$

with $(\xi_t^*)_{t\in\mathbb{Z}}$ an i.i.d. sequence, $\xi_t^* \sim \widehat{P}_\xi$ (see below), independent of $\{Z_s^*;\ s < t\}$; see [5, 17]. Asymptotically, since $p_n \to \infty$, this is not a finite order Markov process anymore. However, under the conditions (A1) and (A2) below, it is still a Bernoulli shift. The parameter estimates $\hat{\phi}_1, \ldots, \hat{\phi}_{p_n}$ are from the Yule–Walker method in an AR(p_n) model, yielding residuals $\hat{\xi}_{p+1}, \ldots, \hat{\xi}_n$. The innovation distribution P_ξ of ξ_t can then be estimated by $\widehat{P}_\xi = (n-p)^{-1} \sum_{t=p+1}^{n} \Delta_{\tilde{\xi}_t}$, where $\tilde{\xi}_t = \hat{\xi}_t - (n-p)^{-1} \sum_{s=p+1}^{n} \hat{\xi}_s$ are the centered residuals and Δ_x is the point mass at $x \in \mathbb{R}$.

The following assumptions imply a weak-dependence property:

(A1) The data-generating process is AR(∞),

$$Z_t = \sum_{j=0}^{\infty} \phi_j Z_{t-j} + \xi_t \qquad (t \in \mathbb{Z}),$$

where $(\xi_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}(\xi_0) = 0$, $\mathbb{E}|\xi_t|^4 < \infty$, ξ_t independent of $\{Z_s;\ s < t\}$.

(A2) $\Phi(z) = 1 - \sum_{j=1}^{\infty} \phi_j z^j$ is bounded away from zero for $|z| \le 1$ ($z \in \mathbb{C}$) and $\sum_{j=0}^{\infty} j^m |\phi_j| < \infty$ for some $m \in \mathbb{N}$. Moreover, the approximating autoregressive order satisfies $p_n \to \infty,\ p_n = o((n/\log(n))^{1/(2m+2)})$.

Note that assumption (A2) requires $\sum_{j=0}^{\infty} j^m |\phi_j| < \infty$, which becomes stronger for large m, and $p_n = o((n/\log(n))^{1/(2m+2)})$, which is also more restrictive with large m. However, the assertion in Proposition 1 below is stronger with large m as well. The condition should be interpreted to mean that the underlying process satisfies (A2) with a maximal m and the sieve bootstrap is then required to work with a correspondingly small enough p_n. If the approximating order is chosen from the data via minimizing the Akaike information criterion (AIC), then Shibata [33] has shown that $\hat{p}_{\mathrm{AIC}} \sim const{\cdot}\, n^{1/(2\beta)}$, if $|\phi_j| \sim const{\cdot}\, j^{-\beta}$ as $j \to \infty$: thus, (A2) holds in conjunction with AIC for the maximal $m \in \mathbb{N}$ which is strictly smaller than $\beta - 1$.

PROPOSITION 1. *Assume* (A1) *and* (A2) *with* $m \in \mathbb{N}$. *Then the AR-sieve bootstrapped process* $(Z_t^*)_{t\in\mathbb{Z}}$ *is* $(\theta, \mathcal{L}, \psi_1)$*-weakly dependent with* $\theta_r = Cr^{-m}$, $C > 0$ *a constant*, *on a set* A_n *with* $\mathbb{P}[A_n] \to 1$.

PROOF. Due to Wiener's theorem, our assumption (A2) is equivalent to assumption A2 in [5]. Section 5.1 and Lemma 5.1 in [5] also describe that

$$Z_t^* = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \xi_{t-j}^*, \qquad \hat{\psi}_{0,n} \equiv 1,$$

and there exists a random variable N such that

$$\sup_{n \ge N} \sum_{j=0}^{\infty} j^m |\hat{\psi}_{j,n}| < \infty \qquad \text{a.s.}$$

Hence, $(Z_t^*)_{t\in\mathbb{Z}}$ is a causal Bernoulli shift and the bound in the discussion following formula (3.2) implies the assertion. $\square$

It is *not* required to resample ξ_t^* from a smoothed version of $\widehat{P}_\xi$. The result here is an elegant extension of the work in [4], where smooth resampling was needed to prove a version of weak dependence (namely ν-mixing, see Section 2.2) for the sieve bootstrapped process.

4. Nonparametric estimation of conditional expectations. As an important *example* of a statistical problem, we investigate asymptotic properties of nonparametric estimation under weak dependence. This problem allows us to compare sharpness of our results with known properties from the frameworks of mixing or associated sequences, discussed in greater detail in Section 5.6.

4.1. *Point estimation.* We restrict ourselves in the sequel to the case of stationary processes $(Z_t)_{t\in\mathbb{Z}}$ with $Z_t = (X_t, Y_t)$, where $X_t,\ Y_t \in \mathbb{R}$. The quantity of interest is $r(x) = \mathbb{E}(Y_0|X_0 = x)$. The extension to the case where $X_t \in \mathbb{R}^d$ for some $d > 1$ is straightforward. Let K be some kernel function integrating to 1, Lipschitzian and rapidly convergent to 0 at infinity (faster than any polynomial decay). For simplicity, we assume throughout the paper that it is compactly supported. The kernel estimator (cf. [31]) is defined by

$$\hat{r}(x) = \hat{r}_{n,h}(x) = \frac{\hat{g}_{n,h}(x)}{\hat{f}_{n,h}(x)} \quad \text{if } \hat{f}_{n,h}(x) \neq 0;\ \hat{r}(x) = 0 \quad \text{otherwise;}$$

$$\hat{f}(x) = \hat{f}_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right),$$

$$\hat{g}(x) = \hat{g}_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right)Y_t.$$

Here $(h_n)_{n\in\mathbb{N}}$ is a sequence of bandwidths (positive real numbers). We always assume that $h_n \to 0,\ nh_n \to \infty\ (n \to \infty)$. The corresponding population versions are the marginal density $f(\cdot)$ of X_t and $g(x) = f(x)r(x)$.

4.1.1. *Bias.* We briefly recall the classical analysis for the deterministic part.

DEFINITION 4. Let $\rho = a + b$ with $a \in \mathbb{N}$ and $0 < b \le 1$. Set

$$\begin{aligned}\mathcal{C}_\rho = \Big\{u\colon \mathbb{R} \to \mathbb{R};\ u \in \mathcal{C}_a \text{ and there exists } A \ge 0 \\ \text{such that } |u^{(a)}(x) - u^{(a)}(y)| \le A|x - y|^b \\ \text{for all } x, y \text{ in any compact subset of } \mathbb{R}\Big\}.\end{aligned}$$

Here, $\mathcal{C}_a$ is the set of a-times continuously differentiable functions and $\mathcal{C}_\rho$ is known as the set of ρ-regular functions.

Using only the stationarity assumption, we have $\mathbb{E}(\hat{g}(x)) = g_h(x)$ with $g_h(x) = \int_{-\infty}^{\infty} K(s)g(x - sh)\,ds$. The study of the bias $b_h(x) = g_h(x) - g(x)$ is purely analytical and does not depend on dependence properties of the sequence $(Z_t)_{t\in\mathbb{Z}}$.

Assuming $g \in \mathcal{C}_\rho$, one can choose a kernel function K of order ρ (not necessarily nonnegative; see [26] or [2]) such that the bias b_h satisfies

$$b_h(x) = g_h(x) - g(x) = O(h^\rho),$$

where the O-term is uniform on any compact subset of $\mathbb{R}$ (cf. [31]). If ρ is an integer with $b = 1$, $\rho = a - 1$, then with an appropriately chosen kernel K of order ρ,

$$b_h(x) \sim \frac{\int s^\rho K(s)\,ds}{\rho!}\, g^{(\rho)}(x)h^\rho,$$

uniformly on any compact interval.

In the following, a ρ-regularity assumption for g (or f) will always be associated with using a kernel K of order ρ for the corresponding estimate.

4.2. *Interval estimation with local bootstrap.* Interval estimation of $r(x) = \mathbb{E}(Y_0|X_0 = x)$ has been proposed with local bootstrap schemes without using normal approximation (see [22], [24]; see also [21]). All these cited references assume an α- or β-mixing condition for the stationary underlying process.

The local bootstrap for nonparametric regression is defined as follows. Consider the empirical distribution function for Y_t given $X_t = x$,

$$\hat{F}_b(\cdot|x) = \sum_{t=1}^{n} K\left(\frac{x - X_t}{b}\right)\frac{\mathbb{1}_{\{Y_t \le \cdot\}}}{\hat{f}_{n,b}(x)},$$

with kernel $K(\cdot)$ and estimator $\hat{f}_{n,b}(\cdot)$ as in Section 4.1, but with bandwidth b generally different from h used in the estimator $\hat{r}_{n,h}$. Construct the bootstrap sample as

$$(X_1, Y_1^*), \dots, (X_n, Y_n^*),\ Y_t^* \sim \hat{F}_b(\cdot|X_t) \qquad (t = 1, \dots, n), \tag{4.1}$$

such that Y_t^* is conditionally independent of Y_s^* $(s \neq t)$, given the data. Thus it involves only some independent resampling. For the particular problem of bootstrapping nonparametric estimators of conditional expectations, this turns out to be sufficient, although the underlying process can be very general. The reason for this is the whitening-by-windowing principle mentioned already in Section 1: as will be shown in Theorem 2 and Proposition 6, the asymptotic distribution of $\hat{r}(x)$ is the same as in the independent case, thus indicating that a bootstrap mechanism does not need to mimic dependence properties of $(Z_t)_{t\in\mathbb{Z}}$. In the sequel, bootstrap moments and distributions induced by the resampling random mechanism in (4.1) are always equipped with an asterisk *.

The bootstrapped estimator is defined with the plug-in rule,

$$\hat{r}^*_{n,h}(x) = \frac{\hat{g}^*_{n,h}(x)}{\hat{f}_{n,h}(x)}, \qquad \hat{g}^*_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right) Y^*_t.$$

Bootstrap percentile confidence intervals for $\mathbb{E}[\hat{r}(x)]$ or $r(x)$ are then constructed as usual. Finite-sample numerical results of this local bootstrap are given in [24] and [6].

5. Asymptotic properties. For asymptotic analysis we assume the following:

$$\begin{array}{ll} & \text{the marginal density } f(\cdot) \text{ of } X_t \text{ exists and is continuous;} \\ & f(x) > 0 \text{ for the point } x \text{ of interest;} \\ (5.1) & \text{the function } r(\cdot) = \mathbb{E}(Y_0 | X_0 = \cdot) \text{ exists and is continuous;} \\ & \text{for some } p \geq 1 \text{ and all } m \leq p\ (m \in \mathbb{N}),\ g_m(\cdot) = f(\cdot)\mathbb{E}(|Y_0|^m | X_0 = \cdot) \text{ exist} \\ & \text{and are continuous.} \end{array}$$

We set $g = fr$ with obvious short notation. Moreover, assume one of the following moment assumptions. Either

$$\mathbb{E}(|Y_0|^S) < \infty \qquad \text{for some } S \tag{5.2}$$

or

$$\mathbb{E}\big(\exp(|Y_0|)\big) < \infty. \tag{5.3}$$

We also consider a conditionally centered version of g_2 appearing in the asymptotic variance of the estimator $\hat{r}$,

$$G_2(x) = f(x)\operatorname{Var}(Y_0 | X_0 = x) = g_2(x) - f(x) r^2(x).$$

5.1. *Variance and asymptotic normality of $\hat{g}$.* Denote by $f_{(k)}$ the density of the pair (X_0, X_k) and assume that there exists some constant $C > 0$ such that

$$\begin{array}{l} \sup_{k\in\mathbb{N}} \|f_{(k)}\|_\infty \leq C, \text{ and} \\ r_{(k)}(x, x') = \mathbb{E}\big(|Y_0 Y_k| \big| X_0 = x, X_k = x'\big) \\ \text{are continuous, uniformly over all } k \in \mathbb{N}. \end{array} \tag{5.4}$$

Under this assumption, the functions $g_{(k)} = f_{(k)} r_{(k)}$ are locally bounded. The following result extends Lemma 1 in [11] for density estimation to the estimate $\hat{g}$ under weak dependence with either ψ_1 or ψ_2.

PROPOSITION 2. *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying conditions* (5.1) *with* $p = 2$, (5.3) *and* (5.4). *Suppose $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with*

$\theta_r = O(r^{-a})$ and $a > 2 + j$ ($j = 1$ or 2). Then, uniformly in x belonging to any compact subset of $\mathbb{R}$,

$$\mathrm{Var}(\hat{g}(x)) = \frac{1}{nh}\, g_2(x) \int K^2(u)\,du + o\left(\frac{1}{nh}\right)$$

and

$$\mathrm{Var}(\hat{g}(x) - r(x)\hat{f}(x)) = \frac{1}{nh} G_2(x) \int K^2(u)\,du + o\left(\frac{1}{nh}\right).$$

REMARK 1. The moment condition in (5.3) can be relaxed. Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies conditions (5.1) with $p = 2$, (5.2) with $S > 2$ and (5.4). Let $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > \frac{(2+j)S-4}{S-2} + \frac{2}{\delta(S-2)}$ ($j = 1$ or 2). Then the assertions of Proposition 2 still hold.

We now investigate central limit theorems.

PROPOSITION 3. *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies conditions* (5.1) *with* $p = 2$, (5.3) *and* (5.4). *Suppose that $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi'_j, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > 2 + j$ ($j = 1$ or 2). Then*

$$\sqrt{nh}(\hat{g}(x) - \mathbb{E}\hat{g}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, g_2(x)\int K^2(u)\,du\right)$$

and

$$\sqrt{nh}([\hat{g}(x) - r(x)\hat{f}(x)] - \mathbb{E}[\hat{g}(x) - r(x)\hat{f}(x)]) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, G_2(x)\int K^2(u)\,du\right).$$

REMARK 2. The results stated in Proposition 3 also hold for finite dimensional convergence at different x's: the components are asymptotically mutually independent, as for i.i.d. samples. Proposition 3 will be used for proving Theorem 2 in Section 5.3.

REMARK 3. The results stated in Proposition 3 also hold for $(\theta, \psi_j, \mathcal{L})$-weakly dependent sequences. See Remark 6.1 and Proposition 6.1 in [1].

5.2. *Higher order moments of $\hat{g}$.* We give here rate-optimal bounds for higher order moments of $\hat{g}$ which turn out to be useful for asymptotics of $\hat{r}$.

THEOREM 1. *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying conditions* (5.1) *with* $p = 2$, (5.3) *and* (5.4). *Suppose that $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. Let $q \geq 3$ be some integer. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is*

$(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > \max(q-1, \frac{4+2j\delta(q-1)}{(q-2)+\delta(4-q)})$ $(j = 1$ or $2)$. Then, for all x belonging to some compact set,

$$\limsup_{n\to\infty}(nh)^{q/2}\left|\mathbb{E}\big(\hat{g}(x) - \mathbb{E}\hat{g}(x)\big)^q\right| < \infty.$$

REMARK 4. Theorem 1 also holds under a weaker moment assumption than (5.3). See Remark 4 in [1].

5.3. *Asymptotic normality of the regression estimator.*

PROPOSITION 4. *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies conditions (5.1) with $p = 2$, (5.3) and (5.4). Consider a positive kernel K. Let $f,\ g \in \mathcal{C}_\rho$ for some $\rho \in]0, 2]$, and let $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume one of the following*:

(i) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_1, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > \max(3, 9\delta)$.*

(ii) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_2, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > \max(3, 12\delta)$.*

Then, uniformly in x belonging to any compact subset of $\mathbb{R}$,

$$\mathbb{E}\big(\hat{r}(x)\big) = r(x) + O\left(h^\rho + \frac{1}{nh}\right).$$

PROPOSITION 5. *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies conditions (5.1) with $p = 2$, (5.3) and (5.4). Consider a positive kernel K. Let $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0, 2]$, and let $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume one of the following*:

(i) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_1, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > \max(5, \frac{30\delta}{7-5\delta})$.*

(ii) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_2, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $a > \max(5, \frac{40\delta}{7-5\delta})$.*

Then, uniformly in x belonging to any compact subset of $\mathbb{R}$,

$$\operatorname{Var}\big(\hat{r}(x)\big) = \frac{G_2(x)}{nhf^2(x)}\int K^2(u)\,du + o\left(\frac{1}{nh}\right).$$

THEOREM 2. *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying conditions (5.1) with $p = 2$, (5.3) and (5.4). Consider a positive kernel K. Let $f,\ g \in \mathcal{C}_\rho$ for some $\rho \in]0, 2]$, and let $nh^{1+2\rho} \to 0,\ n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and*

$$a > \min\left(\max\big(2+j, 3(2+j)\delta\big), \max\left(2+j+\frac{1}{\delta}, \frac{2+2(2+j)\delta}{1+\delta}\right)\right) \quad (j = 1 \text{ or } 2).$$

Then, for all x belonging to any compact subset of $\mathbb{R}$,

$$\sqrt{nh}\big(\hat{r}(x) - r(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \frac{G_2(x)}{f^2(x)} \int K^2(u)\,du\right).$$

In the case where the bandwidth only satisfies the condition $n^\delta h \to \infty$ for some $\delta \in]0, 1[$, asymptotic normality when centered at $\mathbb{E}[\hat{r}(x)]$ still holds (see [1], Proposition 6.2). From this, asymptotic normality of $\sqrt{nh}(\hat{r}(x) - r(x))$ with mean squared error rate-optimal bandwidth $h \sim n^{-1/(1+2\rho)}$ is expected to hold with an asymptotic, nonvanishing bias term.

5.4. *Almost-sure convergence properties.*

We assume for the next Sections 5.4 and 5.5 that the kernel K is also differentiable on its support.

THEOREM 3. *Let* $(Z_t)_{t\in\mathbb{Z}}$ *be a stationary sequence satisfying conditions* (5.1) *with* $p = 2$, (5.3), (5.4) *and assume that it is either* $(\theta, \psi_1, \mathcal{L})$*- or* $(\theta, \psi_2, \mathcal{L})$*-weakly dependent with* $\theta_r \le a^r$ *for some* $0 < a < 1$:

(i) *If* $nh/\log^4(n) \to \infty$, *then for any* $M > 0$, *almost surely*,

$$\sup_{|x|\le M} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| = O\left(\frac{\log^2(n)}{\sqrt{nh}}\right).$$

(ii) *For any* $M > 0$, *if* $f,\ g \in \mathcal{C}_\rho$ *for some* $\rho \in]0, \infty[$, $h \sim (\frac{\log^4(n)}{n})^{1/(1+2\rho)}$ *and* $\inf_{|x|\le M} f(x) > 0$, *then*, *almost surely*,

$$\sup_{|x|\le M} |\hat{r}(x) - r(x)| = O\left\{\left(\frac{\log^4(n)}{n}\right)^{\rho/(1+2\rho)}\right\}.$$

The bound in assertion (ii) is almost optimal: in the i.i.d. setting or also in the framework of mixing processes, the logarithmic factor is $\log(n)$ instead of $\log^4(n)$ here; see also Section 5.6.

REMARK 5. Under the conditions of Theorem 3(ii), but assuming only the weaker condition $n^\delta h \to \infty$ for some $\delta \in]0, 1[$, we obtain

$$\sup_{|x|\le M} |\hat{r}(x) - r(x)| = o(1) \qquad \text{almost surely.}$$

REMARK 6. The assertions of Theorem 3 also hold under weaker moment assumptions than (5.3). (See [1], Proposition 6.3.)

5.5. *Validity of the local bootstrap.* We denote

$$Z_t^{(s,1)} = (Y_t^s, X_t), \qquad s \in \mathbb{N}.$$

THEOREM 4. *Suppose that the stationary sequences $(Z_t^{(s,1)})_{t\in\mathbb{Z}}$ satisfy the weak dependence and moment assumptions of Theorem* 3, *with $p = 4$ in condition* (5.1), *for $s = 1, 2, 3, 4$. Assume that $n^\delta h \to \infty$ for some $\delta \in]0, 1[$, and the pilot bandwidth is $b = O(n^{-\beta})$ for some $\beta \in]0, 1/3[$. Then*:

(i)

$$\sup_{v\in\mathbb{R}} \big|\mathbb{P}^*\big(\sqrt{nh}\big(\hat{r}^*_{n,h}(x) - \mathbb{E}^*\big(\hat{r}^*_{n,h}(x)\big)\big) \le v\big) \\ - \mathbb{P}\big(\sqrt{nh}\big(\hat{r}_{n,h}(x) - \mathbb{E}\big(\hat{r}_{n,h}(x)\big)\big) \le v\big)\big| = o_P(1).$$

(ii) *In the case of no asymptotic bias where* $\lim \sqrt{nh}(\mathbb{E}(\hat{r}_{n,h}(x)) - r(x)) = 0$, *that is*, $nh^{1+2\rho} \to 0$ *if* $f,\ g \in \mathcal{C}_\rho$ *for some* $\rho \in]0, 2]$, *the term* $\mathbb{E}(\hat{r}_{n,h}(x))$ *in* (i) *can be replaced by* $r(x)$.

REMARK 7. In the case of asymptotic bias in (ii), the local bootstrap also is expected to work if additional conditions on the kernel K and on the regularity of the function r hold, and if the pilot bandwidth b satisfies $b/h \to \infty$ ($n \to \infty$); see, for example, [24]. Then

$$\sup_{v\in\mathbb{R}} \Big|\mathbb{P}^*\big(\sqrt{nh}\big(\hat{r}^*_{n,h}(x) - \hat{r}_{n,b}(x)\big) \le v\big) - \mathbb{P}\big(\sqrt{nh}\big(\hat{r}_{n,h}(x) - r(x)\big) \le v\big)\Big| = o_P(1).$$

REMARK 8. Weak dependence of $(Z_t^{(s,1)})$ ($s = 1, 2, 3, 4$) holds if $(Y_t)_{t\in\mathbb{Z}}$ is a suitably regular $\mathbb{R}$-valued Bernoulli shift and $X_t = Y_{t-\ell}$ for some $\ell \in \mathbb{N}$. This is so because, for $Y_t = F(\xi_{t-j},\ j \in \mathbb{Z})$,

$$\begin{aligned} &\big|F^2(\xi_{t-j},\ j \in \mathbb{Z}) - F^2\big(\xi_{t-j}\mathbb{1}_{\{|j|<k\}}, j \in \mathbb{Z}\big)\big| \\ &\quad = \big|F(\xi_{t-j},\ j \in \mathbb{Z}) - F\big(\xi_{t-j}\mathbb{1}_{\{|j|<k\}},\ j \in \mathbb{Z}\big)\big| \\ &\qquad \times \big|F(\xi_{t-j}, j \in \mathbb{Z}) + F\big(\xi_{t-j}\mathbb{1}_{\{|j|<k\}}, j \in \mathbb{Z}\big)\big|. \end{aligned}$$

Assuming some regularity similar to that in the discussion following (3.2) and $\mathbb{E}|Y_t|^{1+\kappa} < \infty$ for some $0 < \kappa < \infty$ yields by Hölder's inequality weak dependence for (Y_t^2), typically with a slower decay of θ_r. By finite iteration, weak dependence of $(Y_t^s)_{t\in\mathbb{Z}}$ follows for $s \in \mathbb{N}$.

5.6. *Comparison with other frameworks.* Similar asymptotic results as in Sections 5.1–5.5 have been derived assuming suitable mixing conditions (which are much stronger than weak dependence). Robinson [29] was first to prove a CLT for $\hat{r}$ under an α-mixing condition. Tran obtains optimal uniform rates of convergence $O((\log(n)/(nh))^{1/2})$ for density estimators which should be compared with our Theorem 3(i): [35] deals with a weak form of β-mixing, and linear processes are considered in [36]. Masry and Tjostheim [19] study nonparametric estimation in ARCH models and provide uniform rates of convergence assuming an α-mixing condition: under an additional smoothness assumption, they obtain optimal uniform rates of convergence; see the right-hand side in (5.5) below. Asymptotic normality of a local polynomial estimator in ARCH models was established in the β-mixing framework [16]. Alternatively, the data generating process may be an associated sequence. Roussas [32] proves uniform rates of convergence under association: his bounds are suboptimal.

Our results are almost as sharp as in the classical framework of mixing sequences. The best comparative aspect is the uniform rates of convergence for the estimator $\hat{r}(\cdot)$, rather than the conditions for establishing a CLT. Zhao and Fang [39] prove the optimal bound for almost-sure convergence, uniformly on compact sets, of the kernel regression estimator for strongly mixing stationary processes,

$$\sup_{|x|\le M} |\hat{r}(x) - r(x)| = O\Big(\big(n^{-1}\log(n)\big)^{\rho/(1+2\rho)}\Big) \quad \text{a.s. } (M > 0). \tag{5.5}$$

For more details about underlying assumptions, see [1], Section 5.6. The difference from Theorem 3 is a slightly better rate by the factor $\log(n)^{-3\rho/(1+2\rho)}$; besides that this result has been shown under a polynomial decay for the mixing coefficients, whereas Theorem 3 requires an exponential decay of weak dependence. Truong and Stone [38] establish optimal pointwise, L^2 and L^∞ bounds on compacts (the latter as above) under some α-mixing conditions.

McKeague and Zhang [20] show asymptotic properties for the nonparametric AR–ARCH model

$$Z_t = r(Z_{t-1}) + s(Z_{t-1})\xi_t,$$

described in Section 3.5 by using a rather different martingale approach. Their results are about integrated conditional mean and variance functions rather than the functions themselves. Moreover, they assume a variance property of the estimator [their assumption (A3)] which was justified by assuming a mixing condition on the data generating process. Our result here justifies their technique: the condition (A3) in [20] can be shown via weak dependence which is implied by a Lipschitz condition on $r(\cdot)$ and $s(\cdot)$; see Section 3.5.

Tran, Roussas, Yakowitz and Truong Van [37] stress the difficulties with time series having discrete innovations: to cope with such problems, they focus on *linear* processes. However, our framework of weak dependence also captures discrete innovations in general *nonlinear* models; see Section 3.

6. Proofs.

6.1. *Variance and asymptotic normality of $\hat{g}$.*

PROOF OF PROPOSITION 2. We give the proof for the more general case described in Remark 1. Denote by C (different) constants whose values are allowed to change. Let

$$T_t(x) = Y_t \mathbb{1}_{\{|Y_t| \le M(n)\}} K\left(\frac{x - X_t}{h}\right) \qquad (t = 1, \dots, n). \tag{6.1}$$

Then the truncated kernel estimator of $g(x)$,

$$\tilde{g}(x) = \frac{1}{nh} \sum_{t=1}^{n} T_t(x), \tag{6.2}$$

satisfies

$$\begin{aligned}(nh)\mathbb{E}\big(\hat{g} - \tilde{g} - \mathbb{E}(\hat{g} - \tilde{g})\big)^2(x) &\le \frac{2n^2}{nh}\mathbb{E}\left[Y_0^2 \mathbb{1}_{\{|Y_0| > M(n)\}} K^2\left(\frac{x - X_0}{h}\right)\right] \\ &\le \mathrm{C}M^{(2-S)}(n) n h^{-1} \to 0 \qquad (n \to \infty)\end{aligned}$$

for $M(n) = M_0 n^\gamma,\ M_0 > 0,\ \gamma \ge (1+\delta)/(S-2)$. It remains to estimate

$$\mathrm{Var}\big(\tilde{g}(x)\big) = \frac{1}{nh^2}\mathrm{Var}\big(T_0(x)\big) + \frac{2}{n^2h^2}\sum_{r=1}^{n-1}(n-r)\,\mathrm{Cov}\big(T_0(x), T_r(x)\big).$$

A classical result (see, e.g., [26], page 37) shows that

$$\mathrm{Var}\big(T_0(x)\big) = h g_2(x) \int K^2(u)\,du + o(h).$$

It follows from the boundedness assumptions on densities that

$$\big|\mathrm{Cov}\big(T_0(x), T_r(x)\big)\big| \le \mathrm{C}h^2.$$

Moreover, the $(\theta, \psi_1, \mathcal{L})$-weak dependence assumption yields

$$\big|\mathrm{Cov}\big(T_0(x), T_r(x)\big)\big| \le \mathrm{C}\theta_r h^{-1} M^2(n). \tag{6.3}$$

Next, we use a truncation device due to Tran [35]: if $a > \frac{3S-4}{S-2} + \frac{2}{\delta(S-2)}$, there exists $\varsigma \in]0,1[$ such that $(2\delta + 2\gamma)/(a-1) < \varsigma < \delta$, so that

$$\frac{2(nh)}{n^2h^2}\sum_{r=1}^{[n^\varsigma]}(n-r)\big|\mathrm{Cov}\big(T_0(x), T_r(x)\big)\big| \le \mathrm{C}n^{\varsigma - \delta} \to 0 \qquad (n \to \infty)$$

and

$$\frac{2(nh)}{n^2h^2}\sum_{r=1+[n^\varsigma]}^{n-1}(n-r)\big|\mathrm{Cov}\big(T_0(x),T_r(x)\big)\big| \le \mathrm{C}\frac{M^2(n)}{h^2}\sum_{r\ge[n^\varsigma]}\theta_r$$
$$\le \mathrm{C}n^{\varsigma(1-a)+2\delta-2\gamma}\to 0 \qquad (n\to\infty).$$

Using the bounds given above, the first assertion of Proposition 2 follows. For a $(\theta,\psi_2,\mathcal{L})$-weakly dependent process, the result follows from

$$\big|\mathrm{Cov}\big(T_0(x),T_r(x)\big)\big| \le \mathrm{C}\theta_r h^{-2}M^2(n)$$

in place of (6.3).

The second assertion follows by replacing Y_t with $Y_t - r(x)$. $\square$

PROOF OF PROPOSITION 3. We proceed as in [27] and more specifically as in [9] for density estimation. Consider a sequence $(W_n)_{n\in\mathbb{N}}$ of i.i.d. $\mathcal{N}(0,1)$ r.v.'s, independent of $(X_t,Y_t)_{t\in\mathbb{Z}}$. Set $M(n)=\log(n)$, $nh\sigma_n^2=\mathrm{Var}(\sum_{t=1}^n T_t(x))$ with $T_t(x)$ given in (6.1), and define the following:

$$\xi_t = \frac{1}{\sigma_n\sqrt{nh}}\big(T_t(x)-\mathbb{E}T_t(x)\big);$$
$$S_k=\sum_{t=1}^k \xi_t,\quad 1\le k\le n \quad\text{and}\quad S_0=0;$$
$$\tau_k=\sum_{t=k}^n V_t,\quad 1\le k\le n \quad\text{and}\quad \tau_{n+1}=0;$$

where $V_k=\frac{\sqrt{v_k}}{\sigma_n}W_k$ and $\frac{v_k}{\sigma_n^2}=|\mathrm{Var}(S_k)-\mathrm{Var}(S_{k-1})|$. For applying the Lindeberg method, denote by φ some three times differentiable function with bounded derivatives, and consider

$$U_t = S_{t-1}+\tau_{t+1},$$
$$R_t(x)=\varphi(U_t+x)-\varphi(U_t)-\frac{v_t}{2\sigma_n^2}\varphi''(U_t)\qquad (t=1,\dots,n).$$

Now, we want to show that

$$\sigma_n \to \left(g_2(x)\int K^2(u)\,du\right)^{1/2}, \tag{6.4}$$

$$\sqrt{nh}\big(\hat g(x)-\mathbb{E}\hat g(x)\big)-\sigma_n S_n \xrightarrow{\mathbb{L}^2} 0, \tag{6.5}$$

$$S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{6.6}$$

We consider either a $(\theta, \psi_1, \mathcal{L})$- or $(\theta, \psi_1', \mathcal{L})$-weakly dependent sequence $(Z_t)_{t\in\mathbb{Z}}$. Analogously as in the proof of Proposition 2, we show that (6.4) holds true if the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_1', \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ and $M^2(n)h_n^{(a-3)} \to 0$. The same result is valid in the $(\theta, \psi_1, \mathcal{L})$-weakly dependent case. Formula (6.5) is easily proved using the exponential moment assumption (5.3).

To prove (6.6), we apply the so-called Lindeberg–Rio method [28]. Note that

$$\begin{aligned}\varphi(S_n) - \varphi(\tau_n) &= \sum_{t=1}^{n} \varphi\Biggl(\sum_{s=1}^{t} \xi_s + \sum_{s=t+1}^{n} V_s\Biggr) - \varphi\Biggl(\sum_{s=1}^{t-1} \xi_s + \sum_{s=t}^{n} V_s\Biggr)\\ &= \sum_{t=1}^{n} R_t(\xi_t) - R_t(V_t).\end{aligned}$$

Then

$$|\mathbb{E}\varphi(S_n) - \mathbb{E}\varphi(W_0)| \le \sum_{t=1}^{n} |\mathbb{E}R_t(\xi_t)| + \sum_{t=1}^{n} |\mathbb{E}R_t(V_t)|. \tag{6.7}$$

Since

$$\begin{aligned}|\mathbb{E}R_t(V_t)| &= \bigl|\mathbb{E}\bigl(\varphi(U_t + V_t) - \varphi(U_t) - V_t\varphi'(U_t) - V_t^2\varphi''(U_t)/2\bigr)\bigr|\\ &\le \mathbb{E}\bigl|V_t^3\varphi^{(3)}(U_t + \vartheta V_t)\bigr|/6 \qquad (\text{with } 0 < \vartheta < 1)\\ &\le (\|\varphi^{(3)}\|_\infty/6)(v_t/\sigma_n^2)^{3/2}\mathbb{E}|W_0|^3\end{aligned}$$

and

$$\begin{aligned}\frac{v_t}{\sigma_n^2} &= \bigl|\operatorname{Var}(S_t) - \operatorname{Var}(S_{t-1})\bigr|\\ &\le C\Biggl(\frac{1}{n} + \sum_{j=1}^{t-1} \frac{1}{nh} \min\biggl(\frac{M^2(n)\theta_{t-j}}{h}, h^2\biggr)\Biggr),\end{aligned}$$

we obtain

$$\begin{aligned}\sum_{t=1}^{n} |\mathbb{E}R_t(V_t)| &\le \frac{C}{\sigma_n^3} \sum_{t=1}^{n} v_t^{3/2}\\ &\le \frac{C}{n^{3/2}\sigma_n^3} \sum_{t=1}^{n} \Biggl(1 + \sum_{j=1}^{t-1} \min\biggl(\frac{M^2(n)\theta_j}{h^2}, h\biggr)\Biggr)^{3/2}.\end{aligned} \tag{6.8}$$

Moreover,

$$\begin{aligned}R_t(\xi_t) &= \varphi(U_t + \xi_t) - \varphi(U_t) - \frac{v_t}{2\sigma_n^2}\varphi''(U_t)\\ &= \xi_t\varphi'(U_t) + \frac{1}{2}\biggl(\xi_t^2 - \frac{v_t}{2\sigma_n^2}\biggr)\varphi''(U_t)\\ &\quad + \frac{1}{6}\xi_t^3\varphi^{(3)}(U_t + \vartheta_t\xi_t) \qquad \text{with } 0 < \vartheta_t < 1.\end{aligned}$$

It then follows that

$$
\begin{aligned}
\sum_{t=1}^{n} |\mathbb{E}R_t(\xi_t)| \le {}& \left|\mathbb{E}\left(\sum_{t=1}^{n} \mathrm{Cov}(\xi_0,\xi_t) \sum_{k=t+1}^{n} \varphi''(U_k)\right)\right| \\
&+ \left|\sum_{t=1}^{n} \mathbb{E}\big(\varphi''(U_t)(\xi_t^2 - \mathbb{E}\xi_t^2)\big)\right| \Big/ 2 \\
&+ \left|\sum_{t=1}^{n}\sum_{j=1}^{t-1} \mathrm{Cov}\big(\varphi''(S_{t-1-j}+\tau_{j+1})\xi_{t-j}, \xi_t\big)\right| \\
&+ \left|\sum_{t=1}^{n}\sum_{j=1}^{t-1} \mathrm{Cov}\big(\varphi^{(3)}(S_{t-1-j}+\tau_{j+1}+\vartheta_{t-j}\xi_{t-j})\xi_{t-j}^2, \xi_t\big)\right| \Big/ 2 \\
&+ \left|\sum_{t=1}^{n} \mathbb{E}\big(\varphi^{(3)}(U_t+\theta_t\xi_t)\xi_t^3\big)\right| \Big/ 6 = E_1+E_2+E_3+E_4+E_5.
\end{aligned}
\tag{6.9}
$$

We now follow [9] to bound the five terms above:

$$
\begin{aligned}
E_1 &\le \sum_{t=1}^{n} \big|\mathrm{Cov}(\xi_0,\xi_t)\big| \left|\sum_{k=t+1}^{n} \mathbb{E}\varphi''(U_k)\right| \\
&\le \frac{C}{nh\sigma_n^2} \sum_{t=1}^{n} (n-t)\min\left(\frac{M^2(n)\theta_t}{h}, h^2\right).
\end{aligned}
\tag{6.10}
$$

Denoting by ι_j some numbers in $]0,1[$,

$$
\begin{aligned}
E_2 &\le \frac{1}{2}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\left|\mathrm{Cov}\left(\varphi^{(3)}(S_{j-1}+\iota_j\xi_j+\tau_{j+1})\xi_j, \frac{\xi_t^2}{2}\right)\right| \\
&\quad + \sum_{t=1}^{n}\left|\mathrm{Cov}\left(\varphi''(\tau_{t+1}), \frac{\xi_t^2}{2}\right)\right| \\
&\le \begin{cases}
\dfrac{CM(n)}{(\sigma_n\sqrt{nh})^3}\displaystyle\sum_{t=2}^{n}\sum_{j=1}^{t-1}\min\left(\frac{M^2(n)(t-j)\theta_j}{h}, h^2\right) + \frac{Ch}{\sigma_n^2}, \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{in the } (\theta, \psi_1, \mathcal{L}) \text{ case}, \\
\dfrac{CM(n)}{(\sigma_n\sqrt{nh})^3}\displaystyle\sum_{t=2}^{n}\sum_{j=1}^{t-1}\min\left(\frac{M^2(n)\theta_j}{h}, h^2\right) + \frac{Ch}{\sigma_n^2}, \quad \text{in the } (\theta, \psi_1', \mathcal{L}) \text{ case},
\end{cases}
\end{aligned}
\tag{6.11}
$$

$$(6.12)\qquad E_3 \le \begin{cases} \dfrac{C}{nh\sigma_n^2}\displaystyle\sum_{t=1}^{n}\sum_{j=1}^{t-1}\min\biggl(\frac{M^2(n)(t-j)\theta_j}{h}, h^2\biggr), & \text{in the } (\theta, \psi_1, \mathcal{L}) \text{ case},\\ \dfrac{C}{nh\sigma_n^2}\displaystyle\sum_{t=1}^{n}\sum_{j=1}^{t-1}\min\biggl(\frac{M^2(n)\theta_j}{h}, h^2\biggr), & \text{in the } (\theta, \psi_1', \mathcal{L}) \text{ case},\end{cases}$$

$$(6.13)\qquad E_4 \le \begin{cases} \dfrac{C}{(\sigma_n\sqrt{nh})^3}\displaystyle\sum_{t=1}^{n}\sum_{j=1}^{t-1}\min\biggl(\frac{M^3(n)(t-j)\theta_j}{h}, h^2M(n)\biggr),\\ \quad\text{in the } (\theta, \psi_1, \mathcal{L}) \text{ case},\\ \dfrac{C}{(\sigma_n\sqrt{nh})^3}\displaystyle\sum_{t=1}^{n}\sum_{j=1}^{t-1}\min\biggl(\frac{M^3(n)\theta_j}{h}, h^2M(n)\biggr),\\ \quad\text{in the } (\theta, \psi_1', \mathcal{L}) \text{ case},\end{cases}$$

$$(6.14)\qquad E_5 \le \frac{C}{(\sigma_n\sqrt{nh})^3}\sum_{t=1}^{n}M(n)h \le \frac{CM(n)}{\sigma_n^3\sqrt{nh}}.$$

If the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_1', \mathcal{L})$-weakly dependent with $\theta_r = O(r^{-a})$ for some $a > 3$, then by using (6.8)–(6.14), the right-hand side of (6.7) tends to zero, which implies the first assertion.

For a $(\theta, \psi_1, \mathcal{L})$-weakly dependent sequence, again using (6.8)–(6.14), we need $\theta_r = O(r^{-a})$ with

$$a > \max\biggl(3 + \frac{1}{\delta}, \frac{6\delta + 2}{\delta + 1}\biggr).$$

The second assertion is an application of the first one when replacing Y_t with $Y_t - r(x)$. $\square$

6.2. *Higher order moments of $\hat{g}$.*

PROOF OF THEOREM 1. We keep the notation from (6.1) and (6.2) and denote again by C a universal constant (whose value might change). Since $\mathbb{E}(\exp(|Y_0|)) < \infty$,

$$\begin{aligned}&(nh)^{q/2}\mathbb{E}|\hat{g} - \tilde{g} - \mathbb{E}(\hat{g} - \tilde{g})|^q(x)\\ &\quad\le \frac{(2n)^q}{(nh)^{q/2}}\mathbb{E}\biggl[|Y_0|^q\mathbb{1}_{\{|Y_0|>M(n)\}}\biggl|K\biggl(\frac{x - X_0}{h}\biggr)\biggr|^q\biggr]\\ &\quad\le \mathrm{C}\frac{(2n)^q}{(nh)^{q/2}}\|K\|_\infty^q(\mathbb{E}|Y_0|^{2q})^{1/q}n^{-M_0/2}\bigl(\mathbb{E}\exp(|Y_0|)\bigr)^{1/2} \to 0 \qquad (n\to\infty)\end{aligned}$$

for $M(n) = M_0 \log(n)$, $M_0 > 0$ big enough. It remains to check that $S_n = (\tilde{g} - \mathbb{E}\tilde{g})(x) = \frac{1}{nh}\sum_{t=1}^{n} W_t$, $W_t = T_t - \mathbb{E}T_t$, satisfies

$$\limsup_{n\to\infty} (nh)^{q/2} |\mathbb{E}S_n|^q < \infty.$$

We set, for any integer $l \in [1, n]$,

$$A_l(\tilde{g}) := \sum_{1\le t_1\le\cdots\le t_l\le n} |\mathbb{E}(W_{t_1}\cdots W_{t_l})|.$$

As in [12], the method relies on the inductive relationship

$$A_q(\tilde{g}) \le n|\mathbb{E}W_0^q| + C_q n \sum_{r=1}^{n-1}(r+1)^{q-2}C_{r,q}(\tilde{g}) + \sum_{m=2}^{q-2} A_m(\tilde{g})A_{q-m}(\tilde{g}).$$

Here $C_{r,q} = C_{r,q}(\tilde{g})$ is defined as

$$\begin{aligned} C_{r,q} = \sup\{&|\operatorname{Cov}(W_{t_1}\cdots W_{t_m}, W_{t_{m+1}}\cdots W_{t_q})|;\\ &1 \le t_1 \le \cdots \le t_q \le n,\ t_{m+1} - t_m = r,\ 1 \le m \le q-1\}, \end{aligned}$$

which is a bound for the expressions

$$|\operatorname{Cov}(W_{t_1}\cdots W_{t_m}, W_{t_{m+1}}\cdots W_{t_q})|,$$

for all increasing sequences $(t_i)_{i=1}^q$ with the same length $r = t_{m+1} - t_m$ of the mth spacing. Since $|W_0| \le C\log(n)$ almost surely and the kernel K is compactly supported, we obtain

$$n|\mathbb{E}W_0^q| \le \mathrm{C}nh\log^q(n). \tag{6.15}$$

The boundedness assumptions on the densities in (5.4) (implying boundedness uniformly in k by uniform continuity) yield a first bound for $C_{r,q}$,

$$C_{r,q} \le \mathrm{C}\log(n)^{q-2}h^2. \tag{6.16}$$

The $(\theta, \psi_1, \mathcal{L})$-weak dependence of the process yields a second bound for $C_{r,q}$,

$$C_{r,q} \le \mathrm{C}\log(n)^q h^{-1}\theta_r. \tag{6.17}$$

We again use Tran's truncation technique. Assume that $q \ge 4$ is an even integer. If $a > \max(q-1, \frac{6\delta(q-1)}{q-2+\delta(4-q)})$, then there exists $\varsigma \in]0,1[$ such that $((2+q)\delta + 2 - q)/2(a-q+1) < \varsigma < ((4-q)\delta + q - 2)/2(q-1)$, so that

$$\begin{aligned} &(nh)^{-q/2}\sum_{r=1}^{[n^\varsigma]} n(r+1)^{q-2}C_{r,q}\\ &\quad\le \mathrm{C}(nh)^{-q/2}nh^2\big(\log(n)\big)^{q-2}[n^\varsigma]^{q-1} \qquad (6.18)\\ &\quad\le n^{(2\varsigma(q-1)-((4-q)\delta+q-2))/2}\big(\log(n)\big)^{q-2} \to 0 \qquad (n\to\infty) \end{aligned}$$

and

$$\begin{aligned}&(nh)^{-q/2}\sum_{r=1+[n^\varsigma]}^{n-1} n(r+1)^{q-2}C_{r,q}\\ &\quad\le \mathrm{C}(nh)^{-q/2}nh^{-1}\bigl(\log(n)\bigr)^q[n^\varsigma]^{q-1-a}\\ &\quad\le n^{(-2\varsigma(a-q+1)+(q+2)\delta+2-q)/2}\bigl(\log(n)\bigr)^q \to 0 \qquad (n\to\infty).\end{aligned}\tag{6.19}$$

Formulae (6.15), (6.18) and (6.19) show that the numbers $B_l = (nh)^{-l/2}A_l(\tilde g)$ satisfy the relation $B_q \le \sum_{m=2}^{q-2} B_m B_{q-m} + c_0$. Therefore (see [8] for more details on Catalan numbers B_l), $B_q \le c_0\frac{(2q-2)!}{q!(q-1)!}$, that is,

$$\limsup_{n\to\infty}(nh)^{q/2}|\mathbb{E}S_n|^q \le \limsup_{n\to\infty} q!(nh)^{-q/2}A_q(\tilde g) \le c_0,$$

which completes the proof for $q \ge 4$ an even integer and for a $(\theta, \psi_1, \mathcal{L})$-weakly dependent process. The case when q is an odd integer greater than or equal to 3 is analogous.

If the process is $(\theta, \psi_2, \mathcal{L})$-weakly dependent, the bound in (6.17) for $C_{r,q}$ becomes

$$C_{r,q} \le \mathrm{C}\log(n)^q h^{-2}\theta_r. \qquad \square$$

6.3. *Asymptotic normality of the regression estimator.*

PROOF OF PROPOSITION 4. We closely follow [7] and [25]. From the expansion

$$u^{-1} = \sum_{i=0}^{p}(-1)^i\frac{(u-u_0)^i}{u_0^{i+1}} + (-1)^{p+1}\frac{(u-u_0)^{p+1}}{uu_0^{p+1}},\tag{6.20}$$

we deduce with $p=2$, $u=b_n$, $u_0=\mathbb{E}b_n=1$,

$$\begin{aligned}\mathbb{E}(\hat r(x)) &= \mathbb{E}a_n - \mathbb{E}\bigl((a_n-\mathbb{E}a_n)(b_n-\mathbb{E}b_n)\bigr) + (\mathbb{E}a_n)\mathbb{E}(b_n-\mathbb{E}b_n)^2\\ &\quad + \mathbb{E}\bigl((a_n-\mathbb{E}a_n)(b_n-\mathbb{E}b_n)^2\bigr) - \mathbb{E}\bigl(\hat r(x)(b_n-\mathbb{E}b_n)^3\bigr),\end{aligned}\tag{6.21}$$

where $\hat r(x) = a_n/b_n$ (if $b_n \ne 0$) with

$$a_n = \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)\Big/\left(n\mathbb{E}K\left(\frac{x-X_0}{h}\right)\right),$$

$$b_n = \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\Big/\left(n\mathbb{E}K\left(\frac{x-X_0}{h}\right)\right).$$

Clearly, since f and g are ρ-regular,

$$\mathbb{E}a_n = r(x) + O(h^\rho).$$

Assume that the sequence is $(\theta, \psi_1, \mathcal{L})$-weakly dependent. By the Cauchy–Schwarz inequality, we bound the second right-hand side term in (6.21) with the variance bounds given in Proposition 2 yielding the order $O((nh)^{-1})$ if $a > 3$. The third term has the same order.

Rosenthal type results, stated in Theorem 1, then yield

$$\begin{aligned}
&\bigl|\mathbb{E}(a_n - \mathbb{E}a_n)(b_n - \mathbb{E}b_n)^2\bigr| \\
&\quad \le \left(\frac{h}{\mathbb{E}K((x - X_0)/h_n)}\right)^3 \bigl(\mathbb{E}\bigl(\hat{g}(x) - \mathbb{E}\hat{g}(x)\bigr)^4\bigr)^{1/2} \bigl(\mathbb{E}\bigl(\hat{f}(x) - \mathbb{E}\hat{f}(x)\bigr)^2\bigr)^{1/2} \\
&\quad = O\bigl((nh)^{-3/2}\bigr)
\end{aligned}$$

if $a > \max(3, 9\delta)$.

For the last term, we use a truncation device

$$\begin{aligned}
\bigl(\mathbb{E}\bigl(\hat{r}(x)\bigr)^4\bigr)^{1/4} &\le M(n) + \left(\mathbb{E}\left(\frac{\sum_{i=1}^n Y_i^4 \mathbb{1}_{\{|Y_i| > M(n)\}} K((x - X_i)/h_n)}{\sum_{j=1}^n K((x - X_j)/h_n)}\right)\right)^{1/4} \\
&= O\bigl(\log(n) + n^{(2 - M_0)/8}\bigr)
\end{aligned}$$

if $M(n) = M_0 \log(n)$. Thus, by Hölder's inequality,

$$\mathbb{E}\hat{r}(x)(b_n - \mathbb{E}b_n)^3 = O\bigl(\log(n)(nh)^{-3/2}\bigr)$$

if $a > \max(3, 9\delta)$. This completes the proof in case of $(\theta, \psi_1, \mathcal{L})$-weak dependence. The $(\theta, \psi_2, \mathcal{L})$-weakly dependent case is similar and details are omitted. $\square$

PROOF OF PROPOSITION 5. We use Collomb's expansion (6.20) with $p = 1$, $u = \hat{f}(x)$, $u_0 = \mathbb{E}\hat{f}(x)$. Thus,

$$\begin{aligned}
&\hat{r}(x) - \mathbb{E}\hat{r}(x) \\
&\quad = \frac{\hat{g}(x)\mathbb{E}\hat{f}(x) - \hat{f}(x)\mathbb{E}\hat{g}(x)}{(\mathbb{E}\hat{f}(x))^2} \\
&\qquad + \frac{(\mathbb{E}\hat{g}(x) - \hat{g}(x))(\hat{f}(x) - \mathbb{E}\hat{f}(x)) + \mathbb{E}((\hat{g}(x) - \mathbb{E}\hat{g}(x))(\hat{f}(x) - \mathbb{E}\hat{f}(x)))}{(\mathbb{E}\hat{f}(x))^2} \\
&\qquad + \frac{\hat{r}(x)(\hat{f}(x) - \mathbb{E}\hat{f}(x))^2 - \mathbb{E}(\hat{r}(x)(\hat{f}(x) - \mathbb{E}\hat{f}(x))^2)}{(\mathbb{E}\hat{f}(x))^2}.
\end{aligned}$$

Now,

$$\operatorname{Var}\bigl(\hat{r}(x)\bigr) = \frac{\mathbb{E}(\hat{g}(x)\mathbb{E}\hat{f}(x) - \hat{f}(x)\mathbb{E}\hat{g}(x))^2}{(\mathbb{E}\hat{f}(x))^4} + o\left(\frac{1}{nh}\right) \tag{6.22}$$

under the conditions

$$nh\mathbb{E}|\hat{g}(x) - \mathbb{E}\hat{g}(x)|^3 \le \big((nh)^{4/3}\mathbb{E}|\hat{g}(x) - \mathbb{E}\hat{g}(x)|^4\big)^{3/4} \to 0,$$
$$nh\mathbb{E}\big|\hat{r}(x)\big(\hat{g}(x) - \mathbb{E}\hat{g}(x)\big)\big|^3 \le nh\big(\mathbb{E}\hat{r}^4(x)\big)^{1/4}\big(\mathbb{E}\big(\hat{f}(x) - \mathbb{E}\hat{f}(x)\big)^4\big)^{3/4} \to 0,$$
$$nh\big(\mathrm{Var}\big(\hat{g}(x)\big)\big)^2 \to 0$$

and

$$nh\big(\mathbb{E}\big(\hat{f}(x) - \mathbb{E}\hat{f}(x)\big)^6\big)^{1/2}\big(\mathbb{E}\big(\hat{g}(x) - \mathbb{E}\hat{g}(x)\big)^6\big)^{1/6}\big(\mathbb{E}|\hat{r}(x)|^3\big)^{1/3} \to 0.$$

The first and second bounds are obtained if the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent ($j = 1$ or 2) with $\theta_r = O(r^{-a})$, where

$$a > \max\left(3, \frac{9(j+2)\delta}{5-2\delta}, \frac{3(j+2)\delta}{1+2\delta}\right).$$

The third bound is satisfied if

$$a > \max\left(1, \frac{2(j+2)\delta}{1+\delta}\right).$$

The last bound holds true if

$$a > \max\left(5, \frac{10(j+2)\delta}{7-5\delta}\right).$$

The first right-hand term in (6.22) can then be handled along the same lines as in the proof of Proposition 2 (note that $a > 2 + j$). $\square$

PROOF OF THEOREM 2. Consider the following identity:

$$\begin{aligned}\hat{r}(x) - r(x) = {} & \frac{(\hat{g} - r\hat{f})(x) - \mathbb{E}((\hat{g} - r\hat{f})(x))}{\mathbb{E}\hat{f}(x)} \\ & - \left(r(x) - \frac{\mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}(x)}\right) + \frac{(r(x) - \mathbb{E}\hat{g}(x)/\mathbb{E}\hat{f}(x))(\hat{f}(x) - \mathbb{E}\hat{f}(x))}{\mathbb{E}\hat{f}(x)} \\ & - \frac{(\hat{g}(x) - \mathbb{E}\hat{g}(x))(\hat{f}(x) - \mathbb{E}\hat{f}(x))}{(\mathbb{E}\hat{f}(x))^2} + \frac{\hat{r}(x)(\hat{f}(x) - \mathbb{E}\hat{f}(x))^2}{(\mathbb{E}\hat{f}(x))^2}.\end{aligned}$$

The term

$$\sqrt{nh}\left(r(x) - \frac{\mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}(x)}\right) = O(nh^{1+2\rho})^{1/2},$$

as a contribution from bias. Assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent ($j = 1$ or 2). Then,

$$\mathbb{E}\left|\sqrt{nh}\frac{(r(x) - \mathbb{E}\hat{g}(x)/\mathbb{E}\hat{f}(x))(\hat{f}(x) - \mathbb{E}\hat{f}(x))}{\mathbb{E}\hat{f}(x)}\right| = O(nh^{(1+2\rho)})^{1/2},$$

$$\begin{aligned}
&\mathbb{E}\left|\sqrt{nh}\frac{(\hat{g}(x) - \mathbb{E}\hat{g}(x))(\hat{f}(x) - \mathbb{E}\hat{f}(x))}{(\mathbb{E}\hat{f}(x))^2}\right| \\
&\qquad \le C\sqrt{nh}\big(\mathbb{E}\big(\hat{g}(x) - \mathbb{E}\hat{g}(x)\big)^2\big)^{1/2}\big(\mathbb{E}\big(\hat{f}(x) - \mathbb{E}\hat{f}(x)\big)^2\big)^{1/2} \\
&\qquad = O(nh)^{-1/2} \quad \text{if } a > 2 + j,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left|\sqrt{nh}\frac{\hat{r}(x)(\hat{f}(x) - \mathbb{E}\hat{f}(x))^2}{(\mathbb{E}\hat{f}(x))^2}\right|^{1/2} \\
&\qquad \le (nh)^{1/4}\big(\mathbb{E}|\hat{r}(x)|\big)^{1/2}\big(\mathbb{E}\big(\hat{f}(x) - \mathbb{E}\hat{f}(x)\big)^2\big)^{1/2} \\
&\qquad = O\big((nh)^{-1/4}\log^{1/2}(n)\big) \quad \text{if } a > 2 + j.
\end{aligned}$$

The result then follows from Proposition 3. $\square$

The CLT for the centered regression function can be proved by the same device. Consider the numbers

$$\alpha_1(\delta) = \min\left(\max(3, 9\delta), \max\left(3 + \frac{1}{\delta}, \frac{6\delta + 2}{\delta + 1}\right)\right),$$

$$\alpha_2(\delta) = \min\left(\max(4, 12\delta), \max\left(4 + \frac{1}{\delta}, \frac{8\delta + 2}{\delta + 1}\right)\right),$$

$$\beta_1(\delta) = \max\left(\alpha_1(\delta), \max\left(3, \frac{27\delta}{7 - 4\delta}\right)\right),$$

$$\beta_2(\delta) = \max\left(\alpha_2(\delta), \max\left(3, \frac{36\delta}{7 - 4\delta}\right)\right).$$

Note that $3 \le \alpha_1(\delta) \le \frac{3(1+\sqrt{3})}{2}$ and $4 \le \alpha_2(\delta) \le 6$.

PROPOSITION 6. *Suppose that the stationary sequence* $(Z_t)_{t\in\mathbb{Z}}$ *satisfies conditions* (5.1) *with* $p = 2$, (5.3) *and* (5.4). *Let* $n^\delta h \to \infty$ *for some* $\delta \in]0, 1[$. *In addition, assume that the sequence* $(Z_t)_{t\in\mathbb{Z}}$ *is* $(\theta, \psi_j, \mathcal{L})$-*weakly dependent* ($j = 1$ *or* 2), *and* $\theta_r = O(r^{-a})$, *where* $a > \beta_j(\delta)$. *Then*

$$\sqrt{nh}\big(\hat{r}(x) - \mathbb{E}\hat{r}(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{G_2(x)}{f^2(x)}\int K^2(u)\,du\right).$$

The proof of Proposition 6 is based upon Collomb's expansion. Details are given in [1] (proof of Proposition 6.2).

6.4. *Almost-sure convergence.*

PROOF OF THEOREM 3. We keep the notation as in (6.1) and (6.2) and again denote by C a universal constant (whose value might change). Since $\mathbb{E}(\exp(|Y_0|)) < \infty$,

$$\mathbb{P}\Big(\sup_{|x|\le M} |\hat{g} - \tilde{g}|(x) > 0\Big) \le n\mathbb{P}\big(|Y_0| > M_0 \log(n)\big) \le \mathrm{C}n^{1-M_0},$$

and, by the Cauchy–Schwarz inequality,

$$\sup_{|x|\le M} \mathbb{E}\big(|\hat{g} - \tilde{g}|(x)\big) \le \frac{1}{h}\,\mathbb{E}\bigg[|Y_0|\mathbb{1}_{\{|Y_0|>M_0\log(n)\}}\bigg|K\bigg(\frac{x-X_0}{h}\bigg)\bigg|\bigg] \le \frac{1}{h}h^{1/3}n^{-M_0}.$$

We can now reduce the computations to those of a density estimator, as given in [12]. Assume that the interval $[-M, M]$ is covered by L_ν intervals with diameter $1/\nu$ [$\nu = \nu(n)$ depends here on n]; we denote by I_j the jth interval and x_j its center. Assume that the relation $h\nu \to \infty$ holds ($n \to \infty$). We then follow a strategy described in detail in [18]. We can bound suprema over an interval I_j as follows:

$$\begin{aligned}\sup_{x\in I_j} |\tilde{g} - \mathbb{E}\tilde{g}|(x) &\le |\tilde{g}(x_j) - \mathbb{E}\tilde{g}(x_j)| \\ &\quad + \frac{\mathrm{C}}{h\nu}\big(|\tilde{g}' - \mathbb{E}\tilde{g}'|(x_j) + 2|\mathbb{E}\tilde{g}'|(x_j)\big),\end{aligned} \tag{6.23}$$

where $\tilde{g}'$ is another suitable kernel density estimator as defined below. To prove (6.23), set

$$w(z) = \|K'\|_\infty \mathbb{1}_{\{|z|\le 2R_0\}},$$

where $[-R_0, R_0]$ is the support of the kernel K. This $w(\cdot)$ is an even kernel, decreasing on $[0, \infty[$, constant on $[0, 2R_0]$, taking the value 0 at $z = 3R_0$, whereas the kernel K vanishes for $z > R_0$. Now take $x \in I_j$. Then

$$\bigg|K\bigg(\frac{x-X_t}{h}\bigg) - K\bigg(\frac{x_j-X_t}{h}\bigg)\bigg| \le \frac{1}{h\nu}w\bigg(\frac{x_j-X_t}{h}\bigg).$$

If $|(x_j - X_t)/h| > 2R_0$, we have

$$\bigg|\frac{x-X_t}{h}\bigg| \ge \bigg|\frac{x_j-X_t}{h}\bigg| - \bigg|\frac{x_j-x}{h}\bigg| \ge 2R_0 - \frac{1}{h\nu} > R_0$$

for n big enough since $h\nu \to \infty$. Define now another kernel-type estimate as

$$\tilde{g}'(x) = \frac{1}{nh}\sum_{t=1}^{n} |Y_t|\mathbb{1}_{\{|Y_t|\le M_0\log(n)\}} w\bigg(\frac{x-X_t}{h}\bigg).$$

Recalling that $x \in I_j$,

$$|\tilde{g}(x) - \mathbb{E}\tilde{g}(x)| \le |\tilde{g}(x_j) - \mathbb{E}\tilde{g}(x_j)| + |\tilde{g}(x) - \tilde{g}(x_j)| + |\mathbb{E}\tilde{g}(x) - \mathbb{E}\tilde{g}(x_j)|$$
$$=: \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

By the Lipschitz property of K,

$$\mathrm{II} \le \frac{1}{h\nu}\frac{1}{nh}\sum_{t=1}^{n} |Y_t| \mathbb{1}_{\{|Y_t| \le M_0 \log(n)\}} w\left(\frac{x_j - X_t}{h}\right) = \frac{1}{h\nu}\tilde{g}'(x_j).$$

Thus,

$$\mathrm{III} \le \mathbb{E}(\mathrm{II}) \le \frac{1}{h\nu}\mathbb{E}\tilde{g}'(x_j).$$

Finally, by centering

$$\mathrm{II} \le \frac{1}{h\nu}|\tilde{g}'(x_j) - \mathbb{E}\tilde{g}'(x_j)| + \frac{1}{h\nu}\mathbb{E}\tilde{g}'(x_j).$$

This proves (6.23).

Therefore, for any $\lambda > 0$,

$$\mathbb{P}\left(\sup_{x \in [-M,M]} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| \ge \frac{2\lambda}{\sqrt{nh}} + \frac{1}{h}h^{1/3}n^{-M_0} + \mathrm{C}\,\frac{\log(n)}{h\nu}\right)$$
$$\le Cn^{1-M_0} + CL_\nu \mathbb{P}\left(|\tilde{g} - \mathbb{E}\tilde{g}|(x_1) \ge \frac{\lambda}{\sqrt{nh}}\right)$$
$$+ CL_\nu \mathbb{P}\left(|\tilde{g}' - \mathbb{E}\tilde{g}'|(x_1) \ge \frac{\lambda}{\sqrt{nh}}\right).$$

Following the same ideas as in [12], the proof of assertion (i) will be complete with the inequality in Lemma 1 below.

The proof of assertion (ii) in Theorem 3 is then straightforward, using Collomb's expansion in (6.20). □

LEMMA 1. *Under the conditions in Theorem* 3, *there exist positive constants* F, G *such that, for any* $\lambda > 0$ *and any bandwidth* $h \to 0$ *with* $nh \ge 1$,

$$\sup_{x \in \mathbb{R}} \mathbb{P}\left(|\tilde{g}(x) - \mathbb{E}\tilde{g}(x)| > \frac{\lambda}{\sqrt{nh}}\right) \le F\exp(-G\sqrt{\lambda}).$$

The same exponential inequality holds for $\tilde{g}'$, *too*.

SKETCH OF THE PROOF OF LEMMA 1. Denote by $K_\infty = 2\|K\|_\infty$. We now specify the constants in the covariance inequalities (6.15), (6.16) and (6.17) for a

$(\theta, \psi_1, \mathcal{L})$-weakly dependent process:

$$C_1(q) = 2\|f\|_\infty R_0 K_\infty^q, \qquad C_2(q) = K_\infty^{q-2}\|f_{0,r}\|_\infty \left(\int |K(u)|\,du\right)^2,$$
$$C_3(q) = q^2 K_\infty^q \operatorname{Lip}(K)/2.$$

Then for some constant $D > 0$, which neither depends on r nor on q,

$$(nh)^{-q/2} \sum_{r=1}^{n} n(r+1)^{q-2} C_{r,q} \le (qD)^q .$$

Together with the previous inductive relation on $A_q(g)$ in Section 6.2 this yields

$$A_q(\tilde{g}) \le \sum_{m=2}^{q-2} A_m(\tilde{g}) A_{q-m}(\tilde{g}) + (qD)^q,$$

which holds for all integers $q \ge 2$. Thus, the numbers $B_l = A_l(\tilde{g})/(Dl\sqrt{nh})^l$ satisfy the relation $B_q \le \sum_{m=2}^{q-2} B_m B_{q-m} + 1 \le (2q-2)!/(q!(q-1)!)$. Therefore,

$$\mathbb{P}\left(|\tilde{g}(x) - \mathbb{E}\tilde{g}(x)| \ge \frac{\lambda}{\sqrt{nh}}\right) \le \lambda^{-q} q! A_q(\tilde{g})(nh)^{-q/2} \le \left(\frac{q^2}{De\lambda}\right)^q$$
$$\le \exp\left(-D\sqrt{\lambda}\,\gamma\left(\frac{q}{\sqrt{D\lambda}}\right)\right),$$

where $\gamma(t) = t\log(t)$. Now by optimizing the value of $\gamma(q/\sqrt{D\lambda})$ we obtain the exponential inequality (for more details see [12] or [13]). $\square$

6.5. *Local bootstrap.*

PROOF OF THEOREM 4. We follow the proof of Theorem 2.1(i) in [24]. Their conditions (A3) and (A8) hold by assumption. The Lipschitz property of the kernel $K(\cdot)$ with compact support is sufficient for the part in their assumption (A4) which is used to achieve Theorem 2.1(i) in [24]. Asymptotic normality in (A6) of [24] holds by our Proposition 6. Then Lemma 5.1 for $\varphi(x) = x^s$ $(s = 1, 2, 3, 4)$ in [24] follows from our Theorem 3(ii); see also Remark 5. All that remains to do for assertion (i) is to check the T_1 term in the proof of Lemma 5.2 in [24].

For the conditional variance $V(x) = \operatorname{Var}(Y_0|X_0 = x)$ and fourth moment $M(x) = \mathbb{E}\{(Y_0 - \mathbb{E}(Y_0|X_0 = x))^4|X_0 = x\}$, assumption (5.1) with $p = 4$ implies that, in a neighborhood U_x of x,

$$V(\cdot) \text{ is differentiable with } \sup_{v\in U_x} |V'(v)| < \infty, \tag{6.24}$$

$$M(\cdot) \text{ is differentiable with } \sup_{v\in U_x} |M'(v)| < \infty. \tag{6.25}$$

It then suffices to show

$$(nh)^{-1}\sum_{t=1}^{n} V(X_t)K^2\left(\frac{x-X_t}{h}\right) = f(x)V(x)\int K^2(w)\,dw + o_P(1), \tag{6.26}$$

$$(nh)^{-1}\sum_{t=1}^{n} M(X_t)K^4\left(\frac{x-X_t}{h}\right) = f(x)M(x)\int K^4(w)\,dw + o_P(1). \tag{6.27}$$

Consider first (6.26). Since we do not assume that V is in $\mathcal{L}$, that is, a bounded Lipschitz function, there is not a direct way to achieve (6.26). By the compact support of K, the term $V(X_t)K((x-X_t)/h)$ is zero for X_t outside a neighborhood of x and if n is sufficiently large. Thus, consider a modification $\widetilde{V}(\cdot)$ which is bounded, is Lipschitz, satisfies (6.24) and has the requirement that $\widetilde{V}(x) = V(x)$ at the point x. Then write the left-hand side of (6.26) as

$$\begin{aligned}(nh)^{-1}\sum_{t=1}^{n} V(X_t)K^2\left(\frac{x-X_t}{h}\right) &= (nh)^{-1}\sum_{t=1}^{n} \widetilde{V}(X_t)K^2\left(\frac{x-X_t}{h}\right) + \Delta_n,\\ \Delta_n &= (nh)^{-1}\sum_{t=1}^{n}\big(\widetilde{V}(X_t) - V(X_t)\big)K^2\left(\frac{x-X_t}{h}\right).\end{aligned} \tag{6.28}$$

By (6.24) applied to $\widetilde{V}(\cdot)$, the fact that $\widetilde{V}(x) = V(x)$ and the continuity of $f(\cdot)$,

$$\mathbb{E}\left((nh)^{-1}\sum_{t=1}^{n} \widetilde{V}(X_t)K^2\left(\frac{x-X_t}{h}\right)\right) = f(x)V(x)\int K^2(w)\,dw + o(1). \tag{6.29}$$

Moreover, since $\widetilde{V}(\cdot)$ and $K(\cdot)$ are bounded Lipschitz functions it follows from the weak-dependence assumption that

$$\operatorname{Var}\left((nh)^{-1}\sum_{t=1}^{n} \widetilde{V}(X_t)K^2\left(\frac{x-X_t}{h}\right)\right) = o(1) \qquad (n\to\infty). \tag{6.30}$$

Finally,

$$\begin{aligned}\mathbb{E}|\Delta_n| &\le \mathbb{E}\left((nh)^{-1}\sum_{t=1}^{n} |V(X_t) - \widetilde{V}(X_t)|K^2\left(\frac{x-X_t}{h}\right)\right)\\ &= \int |V(v) - \widetilde{V}(v)|K^2\left(\frac{x-v}{h}\right)f(v)\,dv = O(h^2),\end{aligned} \tag{6.31}$$

where the last bound follows by assumption (6.24) for $V(\cdot)$ and $\widetilde{V}(\cdot)$. By (6.28)–(6.31) we have shown (6.26). The proof of (6.27) is analogous. This then completes the proof of the first assertion.

The second assertion is an immediate consequence of the first one. $\square$

Acknowledgments. Comments by an Associate Editor and two referees were helpful.

REFERENCES

[1] ANGO NZE, P., BÜHLMANN, P. and DOUKHAN, P. (2001). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.* **30** 397–430.

[2] ANGO NZE, P. and DOUKHAN, P. (1998). Functional estimation for time series: uniform convergence properties. *J. Statist. Plann. Inference* **68** 5–29.

[3] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

[4] BICKEL, P. J. and BÜHLMANN, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5** 413–446.

[5] BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148.

[6] BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* To appear.

[7] COLLOMB, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrsch. Verw. Gebiete* **66** 441–460.

[8] COMTET, L. (1970). *Analyse combinatoire* **1**. Masson, Paris.

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

[10] DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statist.* **85**. Springer, New York.

[11] DOUKHAN, P. and LOUHICHI, S. (1998). Functional estimation of the density of a sequence under a new weak dependence condition. *C. R. Acad. Sci. Paris Sér. I Math.* **327** 989–992.

[12] DOUKHAN, P. and LOUHICHI, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342.

[13] DOUKHAN, P. and PORTAL, F. (1987). Principe d'invariance faible pour la fonction de répartition empirique dans un cadre multidimensionnel et mélangeant. *Probab. Math. Statist.* **8** 117–132.

[14] DUFLO, M. (1996). *Algorithmes stochastiques. Math. Appl.* **23**. Springer, Paris.

[15] ESARY, J., PROSCHAN, F. and WALKUP, D. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466–1474.

[16] HÄRDLE, W. and TSYBAKOV, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics* **81** 223–242.

[17] KREISS, J.-P. (1992). Bootstrap procedures for AR(∞)-processes. *Bootstrapping and Related Techniques*. *Lecture Notes in Econom. and Math. Systems* **376** 107–113. Springer, New York.

[18] LIEBSCHER, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Process. Appl.* **65** 69–80.

[19] MASRY, E. and TJØSTHEIM, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11** 258–289.

[20] MCKEAGUE, I. W. and ZHANG, M.-J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.* **22** 495–514.

[21] NEUMANN, M. H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26** 2014–2048.

[22] NEUMANN, M. H. and KREISS, J.-P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26** 1570–1613.

[23] NEWMAN, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In *Inequalities in Statistics and Probability* (Y. L. Tong, ed.) 127–140.

[24] PAPARODITIS, E. and POLITIS, D. N. (2000). The local bootstrap for kernel estimators under general dependence conditions. *Ann. Inst. Statist. Math.* **52** 139–159.

[25] PAWLAK, M. (1991). On the almost everywhere properties of the kernel regression estimate. *Ann. Inst. Statist. Math.* **43** 311–326.

[26] PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

[27] RIO, E. (1996). Sur le théorème de Berry–Esseen pour les suites faiblement dépendantes. *Probab. Theory Related Fields* **104** 255–282.

[28] RIO, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants. Math. Appl.* **31**. Springer, Paris.

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

[30] ROSENBLATT, M. (1980). Linear Processes and bispectra. *J. Appl. Probab.* **17** 265–270.

[31] ROSENBLATT, M. (1991). *Stochastic Curve Estimation.* IMS, Hayward, CA.

[32] ROUSSAS, G. G. (1991). Kernel estimates under association: strong uniform consistency. *Statist. Probab. Lett.* **12** 393–403.

[33] SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164.

[34] TONG, H. (1981). A note on a Markov bilinear stochastic process in discrete time. *J. Time Ser. Anal.* **2** 279–284.

[35] TRAN, L. T. (1990). Recursive kernel density estimators under a weak dependence condition. *Ann. Inst. Statist. Math.* **42** 305–329.

[36] TRAN, L. T. (1992). Kernel density estimation for linear processes. *Stochastic Process. Appl.* **41** 281–296.

[37] TRAN, L., ROUSSAS, G., YAKOWITZ, S. and TRUONG VAN, B. (1996). Fixed-design regression for linear time series. *Ann. Statist.* **24** 975–991.

[38] TRUONG, Y. K. and STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.

[39] ZHAO, L. C. and FANG, Z. (1985). Strong convergence of kernel estimates of nonparametric regression functions. *Chinese Ann. Math. Ser. B* **6** 147–155.

UNIVERSITÉ LILLE 3
E.A. GREMARS, UFR AES, BP 149
59653 VILLENEUVE D'ASCQ CEDEX
FRANCE
E-MAIL: angonze@univ-lille3.fr

SEMINAR FÜR STATISTIK
ETH ZENTRUM
CH-8092 ZÜRICH
SWITZERLAND
E-MAIL: buhlmann@stat.math.ethz.ch

UNIVERSITÉ CERGY PONTOISE
J.E. 2070 MATHEMATICS, DEPT. OF ECONOMY
33 BOULEVARD DU PORT
95011 CERGY-PONTOISE CEDEX
FRANCE
E-MAIL: doukhan@math.u-cergy.fr