Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 120 (2010) 2412–2431
stochastic processes and their applications
www.elsevier.com/locate/spa

Kernel estimation for time series: An asymptotic theory

Wei Biao Wu*, Yinxiao Huang, Yibi Huang

Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, United States

Received 18 June 2009; received in revised form 4 August 2010; accepted 4 August 2010
Available online 8 August 2010

Abstract

We consider kernel density and regression estimation for a wide class of nonlinear time series models. Asymptotic normality and uniform rates of convergence of kernel estimators are established under mild regularity conditions. Our theory is developed under the new framework of predictive dependence measures which are directly based on the data-generating mechanisms of the underlying processes. The imposed conditions are different from the classical strong mixing conditions and they are related to the sensitivity measure in the prediction theory of nonlinear time series.

Keywords: Kernel estimation; Nonlinear time series; Regression; Central limit theorem; Martingale; Markov chains; Linear processes; Sensitivity measure; Prediction theory; Mean concentration function; Fejér kernel

1. Introduction

Since the seminal work of Engle [13] on ARCH models (autoregressive models with conditional heteroscedasticity) and Tong [37] on TAR (threshold autoregressive) models, nonlinear time series has received considerable attention. Since then a variety of new nonlinear time series models have been proposed. Empirical evidence has been found in many disciplines including computer networks, communication, econometrics, electrical engineering, finance, geology, hydrology and other areas that the underlying random processes exhibit nonlinearity and so the classical ARMA and ARIMA (autoregressive integrated moving-average) based models would be inappropriate. See the excellent monographs of Priestley [29], Tong [38], Fan and Yao [16] and Tsay [40] for examples of nonlinear time series and the related statistical inference.

* Corresponding author.
E-mail addresses: wbwu@galton.uchicago.edu (W.B. Wu), yhuang@galton.uchicago.edu (Y. Huang), yibi@galton.uchicago.edu (Y. Huang).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2010.08.001

A fundamental problem in the study of nonlinear time series is that of unveiling the data-generating mechanisms that govern the observed time series. Nonparametric methods provide a powerful way to infer the underlying mechanisms and only mild structural assumptions are needed. An important nonparametric procedure is the kernel method. There is an extensive literature on the kernel estimation theory for independent and identically distributed (iid) observations; see, for example, [36,8,43,28,25,14]. Further references can be found in [16].

In time series analysis, however, observations are typically dependent. The dependence is the rule rather than the exception and is actually one of the primary goals of study. In the literature a commonly adopted framework for dependence is the strong mixing condition which asserts that the observations are asymptotically independent as the lags increase. Specifically, a stationary process $\{X_t\}_{t\in\mathbb{Z}}$ is said to be *strong mixing* if the strong mixing coefficients

$$\alpha_n := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}_{-\infty}^0,\ B \in \mathcal{A}_n^\infty\} \to 0, \tag{1}$$

where $\mathcal{A}_i^j = \sigma(X_i, \ldots, X_j)$, $i \le j$. Variants of strong mixing conditions include ρ-mixing, ψ-mixing, β-mixing conditions among others [4]. A variety of asymptotic results have been derived under various mixing rates. It is impossible to give a complete list of references here. Representative results are [32,35,5,17] and [3] among others. Rosenblatt [31], Yu [49], Neumann [26] and Neumann and Kreiss [27] deal with β-mixing processes. Further references are given in the excellent reviews by Hardle et al. [20] and Tjostheim [39]. A comprehensive account of nonparametric time series analysis is presented in [16] where numerous asymptotic results are presented under various strong mixing conditions.

This paper advances the nonparametric estimation theory for nonlinear time series under a new framework which is different from the one based on the classical strong mixing conditions. In particular, we shall implement the *dependence measures* proposed in [46] and present a unified asymptotic theory for kernel density and regression estimators. A huge class of time series models can be represented in the form

$$X_n = J(\ldots, \varepsilon_{n-1}, \varepsilon_n), \tag{2}$$

where J is a measurable function and ε_n, $n \in \mathbb{Z}$, are iid random variables; see [41,33,22,29,38]. Clearly (2) defines a stationary and causal process. We interpret (2) as a physical system with $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$ being the input, J being a filter and X_n being the output. Then it is natural to interpret the dependence as the degree of dependence of the output X_n on the input $\mathcal{F}_n$, which is a sequence of innovations that drive the system.

The paper is organized as follows. Section 2 introduces predictive dependence measures [cf. (8) and (10)], which basically quantify the degree of dependence of outputs on inputs. With those dependence measures, we present an asymptotic theory in Sections 2 and 3 for kernel density and regression estimation of time series. Section 4 contains applications to linear and nonlinear processes. Proofs are given in Section 6.

Our results have several interesting features: (i) the predictive dependence measures have nice input/output interpretations and they are directly related to the data-generating mechanisms; (ii) with the martingale theory, the predictive dependence measures are easy to work with; (iii) on the basis of the dependence measures, sharp results can be obtained and (iv) our conditions have a close connection with the sensitivity measure, an important quantity appearing in the prediction theory of stochastic processes. We expect our method and framework to be useful for other problems in time series analysis.

We now introduce some notation. For a random variable X, write $X \in \mathcal{L}^p$, $p > 0$, if $\|X\|_p := [\mathbb{E}(|X|^p)]^{1/p} < \infty$, and $\|\cdot\| := \|\cdot\|_2$. We say that a function g is Lipschitz continuous on a set A with index $0 < \iota \le 1$ if there exists a constant $C_g < \infty$ such that $|g(x) - g(x')| \le C_g|x - x'|^\iota$ for all $x, x' \in A$. In this case, write $g \in \mathcal{C}^\iota(A)$. The notation C denotes a constant whose value may vary from line to line. For a sequence of random variables (η_n) and a sequence of positive numbers (d_n), write $\eta_n = o_{\text{a.s.}}(d_n)$ if η_n/d_n converges to 0 almost surely and $\eta_n = \mathcal{O}_{\text{a.s.}}(d_n)$ if η_n/d_n is almost surely bounded. We can similarly define the relations $o_{\mathbb{P}}$ and $\mathcal{O}_{\mathbb{P}}$. Let $N(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2.

2. Kernel density estimation

A prerequisite for density estimation is that the marginal density of the process $\{X_t\}$ exists. Unfortunately, X_t given in (2) does not always have a density. A simple sufficient condition for the existence of marginal density is that the conditional density exists. Recall that $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$. For $i \in \mathbb{Z}$, $l \in \mathbb{N}$, let $F_l(x|\mathcal{F}_i) = \mathbb{P}(X_{i+l} \le x|\mathcal{F}_i)$ be the l-step-ahead conditional distribution function of X_{i+l} given $\mathcal{F}_i$ and $f_l(x|\mathcal{F}_i) = \frac{\mathrm{d}}{\mathrm{d}x} F_l(x|\mathcal{F}_i)$ be the conditional density.

Condition 1. *There exists a constant $c_0 < \infty$ such that*

$$\sup_{x\in\mathbb{R}} f_1(x|\mathcal{F}_0) \le c_0 \quad \textit{almost surely.} \tag{3}$$

Under Condition 1, it is easily seen that X_i has a density f satisfying the relation $f(x) = \mathbb{E}[f_1(x|\mathcal{F}_0)] \le c_0$. Following [30], given the data $X_1, \ldots, X_n$, the kernel density estimator of f at x_0 is

$$f_n(x_0) = \frac{1}{nb_n} \sum_{t=1}^{n} K\left(\frac{x_0 - X_t}{b_n}\right) = \frac{1}{n} \sum_{t=1}^{n} K_{b_n}(x_0 - X_t), \tag{4}$$

where the kernel K satisfies $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$, $K_b(x) = K(x/b)/b$ and $b = b_n$ is a sequence of bandwidths satisfying the natural condition

$$b_n \to 0 \quad \text{and} \quad nb_n \to \infty. \tag{5}$$

2.1. Dependence measures

To study asymptotic properties of the density estimate f_n, it is necessary to impose appropriate dependence conditions on the underlying process $\{X_t\}$. Instead of the traditional strong mixing conditions, we shall use a different dependence measure.

Let $\{\varepsilon_i'\}$ be an iid copy of $\{\varepsilon_i\}$, $\mathcal{F}_i^* = (\ldots, \varepsilon_{-2}, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i)$ if $i \ge 0$ and $\mathcal{F}_i^* = \mathcal{F}_i$ if $i < 0$, and $X_i^* = J(\mathcal{F}_i^*)$. Namely $\mathcal{F}_i^*$ (resp. X_i^*) is a coupled process of $\mathcal{F}_i$ (resp. X_i) with ε_0 replaced by an iid copy ε_0'. If $f_k(x|\mathcal{F}_0)$ does not depend on ε_0, then $f_k(x|\mathcal{F}_0) = f_k(x|\mathcal{F}_0^*)$. So the quantity $\sup_x \|f_k(x|\mathcal{F}_0) - f_k(x|\mathcal{F}_0^*)\|$, a distance between the two conditional (predictive) distributions $[X_k|\mathcal{F}_0]$ and $[X_k^*|\mathcal{F}_0^*]$, measures the contribution of the innovation ε_0 in predicting the future output X_k given $\mathcal{F}_0$ by perturbing the input via coupling. For a formal definition, let $p > 1$, $k \ge 0$ and

$$\theta_{k,p}(x) = \|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|_p \tag{6}$$

Note that $\mathbb{E}[f_1(x|\mathcal{F}_k)|\mathcal{F}_0] = f_{1+k}(x|\mathcal{F}_0)$. Then

$$\theta_{k,p}(x) = \|\mathbb{E}[f_1(x|\mathcal{F}_k) - f_1(x|\mathcal{F}_k^*)|\mathcal{F}_0, \mathcal{F}_0^*]\|_p \le \|f_1(x|\mathcal{F}_k) - f_1(x|\mathcal{F}_k^*)\|_p. \tag{7}$$

Define the sup-distance

$$\theta_p(k) = \sup_{x\in\mathbb{R}} \theta_{k,p}(x) \tag{8}$$

and the $\mathcal{L}^p$ integral distance

$$\bar{\theta}_p(k) = \left[\int_{\mathbb{R}} \theta_{k,p}^p(x)\mathrm{d}x\right]^{1/p}. \tag{9}$$

Certainly there are other kinds of distances between probability densities, like total variational distance, Hellinger distance and Kullback–Leibler divergence etc. It turns out that in our problem it is more convenient to use the supremum distance (8) and the $\mathcal{L}^p$ distance (9). If $f_1(\cdot|\mathcal{F}_0) \in \mathcal{C}^1$, we define the following distance on the derivatives:

$$\begin{aligned} \psi_{k,p}(x) &= \|f'_{1+k}(x|\mathcal{F}_0) - f'_{1+k}(x|\mathcal{F}_0^*)\|_p, \\ \psi_p(k) &= \sup_{x\in\mathbb{R}} \psi_{k,p}(x), \\ \bar{\psi}_p(k) &= \left[\int_{\mathbb{R}} \psi_{k,p}^p(x)\mathrm{d}x\right]^{1/p}. \end{aligned} \tag{10}$$

These quantities play an important role in the study of asymptotic properties of f_n and they allow us to derive central limit theorems, uniform convergence rates and $\mathcal{L}^p$ distances of $f_n(x) - f(x)$ in a very natural way. They are easy to work with since they are directly related to the data-generating mechanism of X_k. In Section 4 we calculate them for the widely used linear processes and some nonlinear time series. In defining our dependence measures, we require that the processes are of form (2). Such a requirement is not needed in the classical strong mixing conditions and the one in [10].

2.2. $\mathcal{L}^p$ bounds

Let $p > 1$ and $p' = \min(2, p)$. For a real sequence $a = \{a_i\}_{i\in\mathbb{Z}}$, define

$$S_p(n; a) = \sum_{j\in\mathbb{Z}} \left(\sum_{i=1-j}^{n-j} |a_i|\right)^{p'}. \tag{11}$$

Let $\theta_p = \{\theta_p(k)\}_{k\in\mathbb{Z}}$, where $\theta_p(k) = 0$ if $k < 0$. We similarly define $\bar{\theta}_p$, ψ_p and $\bar{\psi}_p$. Let

$$\begin{aligned} \Theta_p(n) &= S_p(n; \theta_p), \qquad \bar{\Theta}_p(n) = S_p(n; \bar{\theta}_p) \\ \Psi_p(n) &= S_p(n; \psi_p), \qquad \bar{\Psi}_p(n) = S_p(n; \bar{\psi}_p). \end{aligned} \tag{12}$$

Theorem 1 provides upper bounds for the sup-norm and integral $\mathcal{L}^p$ norms of $f_n(x) - \mathbb{E}[f_n(x)]$ in terms of $\Theta_p(n)$ and $\bar{\Theta}_p(n)$, respectively.

Theorem 1. *Let $p > 1$. Assume Condition 1, $\int_{\mathbb{R}} |K(v)|\mathrm{d}v < \infty$ and $\sup_v |K(v)| < \infty$. Then*

$$\sup_x \|f_n(x) - \mathbb{E}[f_n(x)]\|_p = O[(nb_n)^{-1/2} + \Theta_p^{1/p'}(n)/n], \tag{13}$$

where $p' = \min(2, p)$, *and*

$$\left[\int_{\mathbb{R}} \| f_n(x) - \mathbb{E}[f_n(x)]\|_p^p \mathrm{d}x\right]^{1/p} = O[(nb_n)^{1/p'-1} + \bar{\Theta}_p^{1/p'}(n)/n]. \tag{14}$$

In Theorem 1, the presence of $\Theta_p(n)$ and $\bar{\Theta}_p(n)$ is due to the dependence. In the special case of $p = 2$, the quantity $S_2(n; a)$ is interestingly related to Fejér's kernel in Fourier analysis. Let $\sqrt{-1}$ be the imaginary unit. For a nonnegative sequence (a_j), let

$$g(u) = \sum_{j\in\mathbb{Z}} a_j \mathrm{e}^{\sqrt{-1}ju}, \quad u \in \mathbb{R},$$

be its Fourier transform. Clearly the Fourier transform of the sequence $a_{j+1} + \cdots + a_{j+n}$, $j \in \mathbb{Z}$, is $g(u)\sum_{k=1}^n \mathrm{e}^{-\sqrt{-1}ku}$. By Parseval's identity, we have the Fejér kernel representation

$$2\pi S_2(n; a) = \int_0^{2\pi} \left| g(u) \sum_{k=1}^n \mathrm{e}^{-\sqrt{-1}ku}\right|^2 \mathrm{d}u = \int_0^{2\pi} |g(u)|^2 \frac{\sin^2(nu/2)}{\sin^2(u/2)} \mathrm{d}u. \tag{15}$$

If the nonnegative sequence (a_j) is summable, assume $\sum_j a_j = 1$ and let the random variable U have the distribution $\mathbb{P}(U = j) = a_j$. Then $S_2(n; a)/n = \int \mathbb{P}^2(t < U \le t + n)\mathrm{d}t/n$. The latter quantity is the *mean concentration function* of U [21]. So it is natural to view $S_p(n; a)$ as a generalized mean concentration function. Corollary 1 provides the magnitude of $S_p(n; a)$ for short- and long-range dependent processes, respectively.

Lemma 1. *Let* $a = \{a_i\}_{i\in\mathbb{Z}}$ *be a real sequence,* $p > 1$ *and* $p' = \min(2, p)$. (i) *If* $\sum_{i\in\mathbb{Z}} |a_i| < \infty$, *then* $S_p(n; a) = O(n)$ (ii) *If* $a_i = O[|i|^{-\beta}\ell(|i|)]$, *where* $1/p' < \beta < 1$ *and* ℓ *is a slowly varying function, then* $S_p(n; a) = O\{n[n^{1-\beta}\ell(n)]^{p'}\}$.

Proof. (i) Let $c_1 = \sum_{i\in\mathbb{Z}} |a_i|$. Then $(\sum_{i=1-j}^{n-j} |a_i|)^{p'} \le c_1^{p'-1} \sum_{i=1-j}^{n-j} |a_i|$, so $S_p(n; a) \le nc_1^{p'}$. (ii) Write $S_n(n; a) = I_n + II_n$, where $I_n = \sum_{j:|j|\ge 2n}$ and $II_n = \sum_{j:|j|<2n}$ (cf (11)). By Karamata's theorem, $\sum_{l=1}^n a_l = O[n^{1-\beta}\ell(n)]$. So $II_n = O\{n[n^{1-\beta}\ell(n)]^{p'}\}$. If $|j| \ge 2n$, then $\sum_{i=1-j}^{n-j} |a_i| = nO[|j|^{-\beta}\ell(|j|)]$ and hence $I_n = O\{n[n^{1-\beta}\ell(n)]^{p'}\}$ by another application of Karamata's theorem. □

Corollary 1. *Assume that the conditions in Theorem 1 hold.* (i) *If*

$$\sum_{k=0}^{\infty} \theta_p(k) < \infty, \tag{16}$$

then $\Theta_p(n) = O(n)$ *and the bound in* (13) *becomes* $O((nb_n)^{-1/2})$. *Similarly, if*

$$\sum_{k=1}^{\infty} \bar{\theta}_p(k) < \infty, \tag{17}$$

then $\bar{\Theta}_p(n) = O(n)$. (ii) *Let* $\theta_p(k) = k^{-\beta}\ell(k)$, *where* $1/p' < \beta < 1$ *and* ℓ *is a slowly varying function. Then* $\Theta_p(n) \sim n[n^{1-\beta}\ell(n)]^{p'}$ *and the bound in* (13) *becomes* $O[(nb_n)^{-1/2} + n^{1/p'-\beta}\ell(n)]$. *The same bound holds for* $\bar{\Theta}_p^{1/p'}(n)$ *if* $\bar{\theta}_p(k) = k^{-\beta}\ell(k)$.

Note that the quantity $\theta_p(k) = \sup_x \|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|_p$ measures the contribution of the innovation ε_0 in predicting X_{k+1} given $\mathcal{F}_0$. Condition (16) indicates that the cumulative contribution of the input ε_0 in predicting future values $\{X_k\}_{k\geq 1}$ is finite, thus suggesting short-range dependence. See [46] for more discussions. The other condition (17) can be similarly interpreted in terms of the $\mathcal{L}^p$ integral norm.

Corollary 1(ii) shows the interesting dichotomous phenomenon [7,44]: If $b_n = o[n^{2\beta-1-2/p'}\ell^{-2}(n)]$, then the first term $(nb_n)^{-1/2}$ dominates and it is same as the one obtained under short-range dependence. On the other hand, however, if we have a large bandwidth b_n such that $n^{2\beta-1-2/p'}\ell^{-2}(n) = o(b_n)$, then the second term $n^{1/p'-\beta}\ell(n)$ dominates. The overall bound depends on the interplay between the bandwidth b_n and the long-range dependence parameter β.

Corollary 2. *Let the conditions in Theorem 1 be satisfied. Assume $f \in \mathcal{C}^2$, $\int_{\mathbb{R}} u^2|K(u)|\mathrm{d}u < \infty$ and $\int_{\mathbb{R}} uK(u)\mathrm{d}u = 0$.* (i) *If $\sup_x |f''(x)| < \infty$, then*

$$\sup_x \|f_n(x) - f(x)\|_p = O[b_n^2 + (nb_n)^{-1/2} + \Theta_p^{1/p'}(n)/n]. \tag{18}$$

(ii) *If $\int_{\mathbb{R}} |f''(u)|^p\mathrm{d}u < \infty$, then*

$$\left[\int_{\mathbb{R}} \|f_n(x) - f(x)\|_p^p\mathrm{d}x\right]^{1/p} = O[b_n^2 + (nb_n)^{1/p'-1} + \bar{\Theta}_p^{1/p'}(n)/n]. \tag{19}$$

Proof. Let the bias $B_n(x) = \mathbb{E}f_n(x) - f(x)$. Since $\int_{\mathbb{R}} uK(u)\mathrm{d}u = 0$,

$$B_n(x) = \int_{\mathbb{R}} K(u)[f(x - b_n u) - f(x) + b_n u f'(x)]\mathrm{d}u.$$

Case (i) is well-known and it easily follows from Taylor's expansion. For (ii), we have

$$\begin{aligned} |B_n(x)| &\leq \int_{\mathbb{R}} |K(u)|\frac{b_n^2 u^2}{2}\int_0^1 |f''(x - b_n ut)|\mathrm{d}t\mathrm{d}u \\ &= O(b_n^2)\int_0^1\int_{\mathbb{R}} |f''(x - b_n ut)|\tilde{K}(u)\mathrm{d}u\mathrm{d}t, \end{aligned}$$

where $\tilde{K}(v) = v^2K(v)/\int_{\mathbb{R}} u^2|K(u)|\mathrm{d}u$. Then (19) follows from

$$\int_{\mathbb{R}} |B_n(x)|^p\mathrm{d}x = O(b_n^{2p})\int_{\mathbb{R}}\int_0^1\int_{\mathbb{R}} |f''(x - b_n ut)|^p\tilde{K}(u)\mathrm{d}u\mathrm{d}t\mathrm{d}x = O(b_n^{2p}). \quad \square$$

2.3. Uniform bounds

Theorem 2. *Assume that, for some $\iota, a > 0$, $K \in \mathcal{C}^\iota$ is a bounded function with bounded support, and that $X_i \in \mathcal{L}^a$. Further assume Condition 1, $\bar{\Theta}_2(n) + \bar{\Psi}_2(n) = O[n^\alpha\ell(n)]$, where $\alpha \geq 1$ and ℓ is a slowly varying function, and $\log n = o(nb_n)$. Then*

$$\sup_{x\in\mathbb{R}} |f_n(x) - \mathbb{E}[f_n(x)]| = \mathcal{O}\left[\sqrt{\frac{\log n}{nb_n}} + n^{\alpha/2-1}\tilde{\ell}(n)\right] \quad \textit{almost surely.} \tag{20}$$

Here $\tilde{\ell}$ is another slowly varying function.

If $\int_{\mathbb{R}} uK(u)\mathrm{d}u = 0$ and $\sup_x |f''(x)| < \infty$, then the bias $B_n(x) = \mathbb{E}f_n(x) - f(x)$ satisfies $\sup_x |B_n(x)| = O(b_n^2)$. If additionally $\alpha < 8/5$, then for $b_n \asymp (n^{-1}\log n)^{1/5}$, by Theorem 2, we have

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x)| = \mathcal{O}(n^{-1}\log n)^{2/5} \quad \text{almost surely.} \tag{21}$$

Stute [34] showed that, if X_i are iid, then $(n/\log n)^{2/5} \sup_{|x|\le c} |f_n(x) - f(x)|/f(x)$ converges almost surely to a non-zero constant if $\inf_{|x|\le c} f(x) > 0$, $c > 0$. So (21) gives the optimal convergence rate $(n^{-1}\log n)^{2/5}$. Section 5 contains a comparison study of Theorem 2 and results obtained under strong mixing conditions. Bickel and Rosenblatt [2] obtained a deep result on asymptotic distributional properties of $\sup_{0\le x\le 1} |f_n(x) - \mathbb{E}[f_n(x)]|$ for iid random variables X_i. Their result is generalized by Neumann [26] to geometrically β-mixing processes; see also [23] for some recent contributions.

3. Kernel regression estimation

Nonparametric techniques play an important role in assessing the relationship between predictors and responses if the form of the functional relation is unknown. A popular nonparametric procedure is the Nadaraya–Watson estimator. To formulate the regression problem, we consider the model

$$Y_n = G(X_n, \eta_n), \tag{22}$$

where η_n, $n \in \mathbb{Z}$, are also iid and η_n is independent of $\mathcal{F}_{n-1} = (\ldots, \varepsilon_{n-2}, \varepsilon_{n-1})$. An important special example of (22) is the autoregressive model

$$X_{n+1} = R(X_n, \varepsilon_{n+1}) \tag{23}$$

on letting $\eta_n = \varepsilon_{n+1}$ and $Y_n = X_{n+1}$. Given the data (X_i, Y_i), $0 \le i \le n$, let

$$T_n(x) = \frac{1}{n}\sum_{t=1}^{n} Y_t K_{b_n}(x - X_t). \tag{24}$$

Then the Nadaraya–Watson estimator of the regression function

$$g(x_0) = \mathbb{E}(Y_n | X_n = x_0) = \mathbb{E}[G(x_0, \eta_0)] \tag{25}$$

has the form

$$g_n(x_0) = \frac{T_n(x_0)}{f_n(x_0)}. \tag{26}$$

Here we shall present an asymptotic theory for $g_n(x_0)$. In particular, under mild regularity conditions on G and f, we shall provide a central limit theorem and a uniform convergence rate for $g_n(x_0) - g(x_0)$. Let

$$V_p(x) = \mathbb{E}[|G(x, \eta_n)|^p] \quad \text{and} \quad \sigma^2(x) = V_2(x) - g^2(x). \tag{27}$$

The following regularity conditions on K are needed.

Condition 2. *The kernel K is symmetric and bounded on $\mathbb{R}$: $\sup_{u\in\mathbb{R}} |K(u)| \le K_0$, $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$ and K has bounded support; namely, $K(x) = 0$ if $|x| \ge c$ for some $c > 0$.*

Theorem 3. *Let $p > 2$. Assume Conditions 1 and 2, V_2, $g \in \mathcal{C}(\mathbb{R})$ and that $V_p(x)$ is bounded on a neighborhood of x_0. Further assume that*

$$b_n \Theta_2(n) = o(n) \quad \text{and} \quad nb_n \to \infty. \tag{28}$$

Let $\kappa = \int K^2(u)\mathrm{d}u$. Then

$$\sqrt{nb_n}\{T_n(x_0) - \mathbb{E}[T_n(x_0)]\} \Rightarrow N[0,\ V_2(x_0) f(x_0)\kappa]. \tag{29}$$

In Theorem 3, (29) can be used to prove central limit theorems for kernel density and Nadaraya–Watson estimates (cf. Corollary 3). For $G \equiv 1$, $T_n(x) = f_n(x)$ is the kernel density estimate and one has (29) with $V_2 \equiv 1$. Wu [46] obtained asymptotic normality of f_n under the condition $\sum_{k=0}^{\infty} \theta_2(k) < \infty$. In this case $\Theta_2(n) = O(n)$ and (28) is automatically satisfied under the natural bandwidth condition (5). Clearly condition (28) also allows long-memory processes. Wu and Mielniczuk [44] considered the special cases of short- and long-memory linear processes.

Corollary 3. *Let $f(x_0) > 0$. Then under the conditions of Theorem 3, we have*

$$\sqrt{nb_n}\left\{g_n(x_0) - \frac{\mathbb{E}T_n(x_0)}{\mathbb{E}f_n(x_0)}\right\} \Rightarrow N[0, \sigma^2(x_0)\kappa/f(x_0)]. \tag{30}$$

Proof. Let $\nu_n = \nu_n(x_0) = \mathbb{E}T_n(x_0)$ and $\mu_n = \mu_n(x_0) = \mathbb{E}f_n(x_0)$. Since $f(x_0) > 0$, K has bounded support and g is continuous, $\nu_n/\mu_n \to g(x_0)$. Observe that

$$\begin{aligned} T_n(x_0) - f_n(x_0)\frac{\nu_n}{\mu_n} &= \{T_n(x_0) - f_n(x_0)g(x_0) - \nu_n + \mu_n g(x_0)\} \\ &\quad + [f_n(x_0) - \mu_n][g(x_0) - \nu_n/\mu_n] =: A_n + B_n. \end{aligned} \tag{31}$$

Applying this time Theorem 3 with $G \equiv 1$ instead of G, we have $B_n\sqrt{nb_n} = o_{\mathbb{P}}(1)$ and $f_n(x_0) \to f(x_0)$ in probability. Hence again by Theorem 3, $\sqrt{nb_n}A_n \Rightarrow N[0, \sigma^2(x_0)\kappa f(x_0)]$, which by Slutsky's theorem yields (30). □

Theorem 4. *Let $p > 1$ and $p' = \min(2, p)$. Assume that $Y_i \in \mathcal{L}^p$, $g \in \mathcal{C}(\mathbb{R})$, $V_{p'}(\cdot)$ is bounded in an open interval containing $[-m, m]$, $m > 0$, and the kernel $K \in \mathcal{C}^\iota$, $\iota > 0$, satisfies Condition 2.* (i) *Let $z_n = n^{1/p}\log n + (nb_n \log n)^{1/2}$. Then*

$$\sup_{x\in[-m,m]} |T_n(x) - \mathbb{E}[T_n(x)]| = \frac{\mathcal{O}_{\text{a.s.}}(z_n)}{nb_n} + \frac{\mathcal{O}_{\mathbb{P}}[\sqrt{\Theta_2(n) + \Psi_2(n)}]}{n}. \tag{32}$$

(ii) *If additionally $\Theta_2(n) + \Psi_2(n) = O[n^\alpha \ell(n)]$, where $\alpha \geq 1$ and ℓ is a slowly varying function, then (32) has the bound $\mathcal{O}_{\text{a.s.}}[z_n/(nb_n)] + o_{\text{a.s.}}[n^{\alpha/2-1}\tilde{\ell}(n)]$, where $\tilde{\ell}$ is a slowly varying function depending on ℓ.*

In the kernel estimation theory it is routine to compute the bias $\nu_n(x)/\mu_n(x) - g(x)$. If f, $g \in \mathcal{C}^2$ and K satisfies Condition 2, then it is easily seen that the bias is of the order $\mathcal{O}(b_n^2)$. Clearly $\sup_{x\in[-m,m]} |f_n(x) - \mu_n(x)|$ has the same bound as the one in (32). By (31) and Theorem 4, we have:

Corollary 4. *Assume that f, $g \in \mathcal{C}^2$, $\inf_{|x|\leq m} f(x) > 0$, $\sup_{|x|\leq m} |g''(x)| < \infty$, and that $K \in \mathcal{C}^\iota$, $\iota > 0$ satisfies Condition 2. Then for any $m > 0$,*

$$\sup_{|x|\leq m} |g_n(x) - g(x)| = \frac{\mathcal{O}_{\text{a.s.}}(z_n)}{nb_n} + \frac{\mathcal{O}_{\mathbb{P}}[\sqrt{\Theta_2(n) + \Psi_2(n)}]}{n} + O(b_n^2). \tag{33}$$

Corollary 4 allows long-memory as well as heavy-tailed processes. As in (ii) of Lemma 1, if $\theta_2(k) + \psi_2(k) = k^{-\beta}\ell(n)$, where $1/2 < \beta < 1$ and ℓ is a slowly varying function, then $\Theta_2(n) + \Psi_2(n) = O\{n[n^{1-\beta}\ell(n)]^2\}$. The term that dominates the sum will vary with different choices of b_n, suggesting dichotomy. Now consider the short-memory case in which $\sum_{k=0}^{\infty}[\theta_2(k) + \psi_2(k)] < \infty$. Then $\Theta_2(n) + \Psi_2(n) = O(n)$ and the bound in (33) becomes $\mathcal{O}_{\mathbb{P}}[z_n/(nb_n)] + O_{\mathbb{P}}(b_n^2)$. If $p \le 5/2$, then the latter bound achieves a minimum if $b_n \asymp (n^{1/p-1}\log n)^{1/3}$. On the other hand, if $p > 5/2$, then the minimal bound is achieved if $b_n \asymp (n^{-1}\log n)^{1/5}$.

4. Applications

To apply the results of Sections 2 and 3, we need to calculate $\theta_p(k)$, $\bar{\theta}_p(k)$ and $\psi_p(k)$ defined in (8)–(10). It is usually not difficult to calculate them since they are directly related to the underlying data-generating mechanism. Sections 4.1–4.3 consider linear processes, iterated random functions and chains with infinite memory, respectively.

4.1. Linear processes

Let ε_i be iid random variables with density f_ε; let (a_i) be real coefficients such that

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i} \tag{34}$$

is a well-defined random variable. Important special cases of (34) include ARMA and fractional ARIMA models. Assume that $\varepsilon_i \in \mathcal{L}^q$, $q > 0$, and that f_ε satisfies

$$c_2 := \sup_x [|f_\varepsilon(x)| + |f'_\varepsilon(x)| + |f''_\varepsilon(x)|] < \infty. \tag{35}$$

Then both $\theta_p(k)$ and $\psi_p(k)$ are of order $O[|a_{k+1}|^{\min(1,q/p)}]$; see Lemma 3 in [47]. For completeness we include that simple argument here. Let $a_0 = 1$, $Y_k = X_{k+1} - \varepsilon_{k+1}$ and $D_k = a_{k+1}(\varepsilon_0 - \varepsilon'_0)$, $k \ge 0$. Then $f_1(x|\mathcal{F}_k) = f_\varepsilon(x - Y_k)$ and

$$\begin{aligned}\theta_{k,p}(x) &= \|\mathbb{E}[f_1(x|\mathcal{F}_k)|\mathcal{F}_0] - \mathbb{E}[f_1(x|\mathcal{F}_k^*)|\mathcal{F}_0^*]\|_p \\ &\le 2\|f_\varepsilon(x - Y_k) - f_\varepsilon(x - Y_k^*)\|_p \le 2c_2\|\min(1, |D_k|)\|_p \\ &\le 2c_2\{\mathbb{E}[|D_k|^{\min(q,p)}]\}^{1/p} = O[|a_{k+1}|^{\min(1,q/p)}]\end{aligned} \tag{36}$$

since $\sup_x[|f_\varepsilon(x)| + |f'_\varepsilon(x)|] < \infty$. If, additionally, $\sup_x |f''_\varepsilon(x)| < \infty$, then the same bound holds for $\psi_p(k)$. It is worthwhile to mention that in our setting heavy-tailed distributions are allowed. To deal with $\bar{\theta}_p(k)$, we shall impose the following analogue of (35):

$$I_0 := \int_{\mathbb{R}} [|f_\varepsilon(x)|^p + |f'_\varepsilon(x)|^p + |f''_\varepsilon(x)|^p]\mathrm{d}x < \infty. \tag{37}$$

Let $t \in \mathbb{R}$ and $p > 1$. By Hölder's inequality, since $f_\varepsilon(x+t) - f_\varepsilon(x) = \int_0^t f'_\varepsilon(x+u)\mathrm{d}u$,

$$\int_{\mathbb{R}} |f_\varepsilon(x+t) - f_\varepsilon(x)|^p \mathrm{d}x \le \int_{\mathbb{R}} \left| |t|^{p-1} \int_0^t |f'_\varepsilon(x+u)|^p \mathrm{d}u \right| \mathrm{d}x \le I_0|t|^p.$$

It is easily seen that the above integral is also bounded by $2^p I_0$. Then, by (36),

$$\int_{\mathbb{R}} \theta_{k,p}^p(x)\mathrm{d}x = \mathbb{E}\int_{\mathbb{R}} |f_\varepsilon(x - Y_k) - f_\varepsilon(x - Y_k^*)|^p \mathrm{d}x$$
$$\le \mathbb{E}[\min(2^p I_0, I_0|a_{k+1}\varepsilon_0 - a_{k+1}\varepsilon_0'|^p)] = O[|a_{k+1}|^{\min(p,q)}]. \tag{38}$$

With (36) and (38), we are able to give bounds for $\Theta_p(n) = S_p(n; \theta_p)$ and $\bar{\Theta}_p(n) = S_p(n; \bar{\theta}_p)$. Consider the special case $p = q = 2$. If the short-range dependence condition

$$\sum_{i=0}^{\infty} |a_i| < \infty \tag{39}$$

holds, then $\bar{\Theta}_2(n) + \bar{\Psi}_2(n) = O(n)$, and, under the mild bandwidth condition $b_n + 1/(nb_n) = O(n^{-\delta})$, $\delta > 0$, the bound in (20) becomes $\mathcal{O}[(\log n)^{1/2}/(nb_n)^{1/2}]$. Note that the optimal bound (21) continues to hold for long-range dependent processes with $a_n = O(n^{-\beta})$, $9/10 < \beta < 1$, in which case by Corollary 1 we have $\bar{\Theta}_2(n) + \bar{\Psi}_2(n) = O(n^{3-2\beta})$ and (21) follows from elementary calculations.

For short-memory linear processes, Wu and Mielniczuk [44] proved a central limit theorem for $f_n(x)$ by assuming that f_ε is Lipschitz continuous and ε_i has finite second moment. The former condition is weaker than (35) while in our setting we allow $\mathbb{E}(\varepsilon_0^2) = \infty$. For long-memory linear processes, using Ho and Hsing's [19] empirical process theory, Wu and Mielniczuk [44] discovered the dichotomous and trichotomous phenomena for $f_n(x)$ for various choices of bandwidths. Since there is no empirical process theory for long-memory nonlinear processes, our general approach here is unable to produce Wu and Mielniczuk's dichotomy and trichotomy results.

4.2. *Iterated random functions*

Consider the nonlinear time series defined by the recursion

$$X_n = R_{\varepsilon_n}(X_{n-1}), \tag{40}$$

where R is a bivariate measurable function. For different forms of R, one can get threshold autoregressive models [38], AR with conditionally heteroscedasticity [13], random coefficient models [24] and exponential autoregressive models [18] among others.

Diaconis and Freedman [9] showed that (40) has a unique and stationary distribution if there exist $\alpha > 0$ and x_0 such that

$$L_{\varepsilon_0} + |R_{\varepsilon_0}(x_0)| \in \mathcal{L}^\alpha \quad \text{and} \quad \mathbb{E}[\log(L_{\varepsilon_0})] < 0, \quad \text{where}$$
$$L_{\varepsilon_0} = \sup_{x \ne x'} \frac{|R_{\varepsilon_0}(x) - R_{\varepsilon_0}(x')|}{|x - x'|}. \tag{41}$$

In this case, by iterating (40), we have that X_n is of form (2). Due to the Markovian structure, we can write $f_k(x|\mathcal{F}_0) = f_k(x|X_0)$, where $f_k(x|X_0)$ is the conditional density of X_k at x given X_0. Let $f_k'(y|x) = \partial f_k(y|x)/\partial y$. For $p > 1$, $k \in \mathbb{N}$ define

$$I_{k,p}(x) = \left[\int_{\mathbb{R}} \left|\frac{\partial}{\partial x} f_k(y|x)\right|^p \mathrm{d}y\right]^{1/p} \quad \text{and} \quad J_{k,p}(x) = \left[\int_{\mathbb{R}} \left|\frac{\partial}{\partial x} f_k'(y|x)\right|^p \mathrm{d}y\right]^{1/p}. \tag{42}$$

In the nonlinear prediction theory, a key problem is the sensitivity of initial values. In particular, one needs to study the distance between the k-step-ahead predictive distributions $[X_k|X_0 = x]$ and $[X_k|X_0 = x + \delta]$, which results from a drift δ in the initial value. A natural way to quantify the sensitivity is to use the $\mathcal{L}^p$ distance

$$\Delta_k(x, \delta) := \left[\int_{\mathbb{R}} |f_k(y|x + \delta) - f_k(y|x)|^p \mathrm{d}y\right]^{1/p}. \tag{43}$$

Fan and Yao [16, p. 466] considered the case $p = 2$. Under certain regularity conditions, $\lim_{\delta\to 0} \Delta_k(x, \delta)/|\delta| = I_{k,p}(x)$. $J_{k,p}(x)$ can be similarly interpreted as a prediction sensitivity measure. Wu [48] applied the sensitivity measure to empirical processes. Proposition 1 shows the relation between $\bar{\theta}_p(k)$ and $I_{k,p}$.

Proposition 1. *Let* $\tau_{k,p}(a, b) = \int_a^b I_{k,p}(x)\mathrm{d}x$, $k \geq 1$. *Then* (i) $\bar{\theta}_p(k - 1) \leq \|\tau_{k,p}(X_0, X_0^*)\|_p$ *and* (ii) $\bar{\theta}_p(k - 1) \leq 2\|\tau_{1,p}(X_{k-1}, X_{k-1}^*)\|_p$.

Proof. Let $q = p/(p - 1)$ and $\lambda(x) = I_{k,p}^{1/q}(x)$. By Hölder's inequality,

$$\left|\int_{X_0}^{X_0^*} \frac{\partial}{\partial x} f_k(y|x)\mathrm{d}x\right|^p \leq \left|\int_{X_0}^{X_0^*} \frac{|\partial f_k(y|x)/\partial x|^p}{\lambda^p(x)}\mathrm{d}x\right| \times |\tau_{k,p}(X_0, X_0^*)|^{p/q}.$$

Hence $\int_{\mathbb{R}} |f_k(y|X_0) - f_k(y|X_0^*)|^p \mathrm{d}y \leq |\tau_{k,p}(X_0, X_0^*)|^p$ and (i) follows. By (7), (ii) similarly follows. □

If $r = \|L_{\varepsilon_0}\|_p < 1$, then $\|X_n - X_n^*\|_p = O(r^n)$ [45]. If additionally $\sup_x I_{1,p}(x) < \infty$, then by Proposition 1(ii), $\bar{\theta}_p(n) = O(r^n)$.

When $p = 1$, the quantity $\tau_{k,1}(X_0, X_0^*)$ in Proposition 1 is closely related to the τ-dependent coefficient in [11]. Let $\Lambda_1(\mathbb{R})$ be the set of 1-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$. Then their τ coefficient $\tau(\sigma(X_0), X_k)$ is

$$\mathbb{E} \sup_{g\in\Lambda_1(\mathbb{R})} |\mathbb{E}[g(X_k)|X_0] - \mathbb{E}g(X_k)| = \mathbb{E} \sup_{g\in\Lambda_1(\mathbb{R})} \left|\int g(y)[f_k(y|X_0) - f(y)]\mathrm{d}y\right|.$$

If $\sup_{x,y}[f_k(y|x) + f(y)] < \infty$, then $\tau(\sigma(X_0), X_k) \leq C\mathbb{E}\int |f_k(y|X_0) - f(y)|\mathrm{d}y$. Note that $\mathbb{E}f_k(y|X_0^*) = f(y)$. Then $\tau(\sigma(X_0), X_k) \leq C\mathbb{E}[\tau_{k,1}(X_0, X_0^*)]$.

4.3. Chains with infinite memory

Doukhan and Wintenberger [12] introduced a model for chains with infinite memory:

$$X_{k+1} = F(X_k, X_{k-1}, \ldots; \varepsilon_{k+1}), \tag{44}$$

where ε_k are iid innovations. Here we consider a special form of (44):

$$X_{k+1} = G(X_k, X_{k-1}, \ldots) + \varepsilon_{k+1}, \tag{45}$$

where, as in [12], we assume that G satisfies

$$|G(x_{-1}, x_{-2}, \ldots) - G(x'_{-1}, x'_{-2}, \ldots)| \leq \sum_{j=1}^{\infty} \omega_j |x_{-j} - x'_{-j}|, \quad \text{where } w_j \geq 0. \tag{46}$$

Under suitable conditions on $(\omega_j)_{j\geq 1}$, iterations of (45) lead to a stationary solution X_k of form (2). For such processes we are also able to provide a bound for $\Theta_p(n)$ and other similar quantities, so our theorems are applicable. For simplicity let $p = 2$ and assume $\varepsilon_k \in \mathcal{L}^2$. Let $\delta_2(k) = \|X_k - J(\mathcal{F}_k^*)\|$. For $k \geq 0$, by (45) and (46), we have

$$\delta_2(k+1) \leq \sum_{i=1}^{k+1} \omega_i \delta_2(k+1-i). \tag{47}$$

Define the sequence $(a_k)_{k\geq 0}$ recursively by $a_0 = \delta_2(0)$ and

$$a_{k+1} = \sum_{i=1}^{k+1} \omega_i a_{k+1-i}. \tag{48}$$

Then $S_2(n; \delta_2(\cdot)) \leq S_2(n; a)$. Let $A(s) = \sum_{k=0}^{\infty} a_k s^k$ and $\Omega(s) = \sum_{i=1}^{\infty} \omega_i s^i$, $|s| \leq 1$. By (48), we have $A(s) = a_0 + A(s)\Omega(s)$. Hence $A(s) = a_0(1 - \Omega(s))^{-1}$. Assume that, as $s \uparrow 1$, $1 - \Omega(s) \sim (1-s)^d$ with $d \in (0, 1/2)$. Note that in our setting $\Omega(1) = \sum_{j=1}^{\infty} \omega_j = 1$, while $\Omega(1) < 1$ is required in [12]. Hence we can allow stronger dependence. As in (15), with elementary manipulations, we have

$$2\pi S_2(n; a) = \int_0^{2\pi} |A(e^{\sqrt{-1}u})|^2 \frac{\sin^2(nu/2)}{\sin^2(u/2)} du = O(n^{1+2d}).$$

Assume that the density function of ε_i satisfies (35). As in Section 4.1, let $Y_k = X_{k+1} - \varepsilon_{k+1}$. Following the calculation in (36), we have $\theta_{k,2} = O(\|Y_k - Y_k^*\|) = O(\delta_2(k+1))$. Hence $\Theta_2(n) = O(n^{1+2d})$. If, as in [12], $\Omega(1) < 1$, then $A(1) < \infty$ and $\Theta_2(n) = O(n)$. Other quantities $\bar{\Theta}_2(n)$, $\Psi_2(n)$ and $\bar{\Psi}_2(n)$ can be similarly dealt with.

5. A comparison with earlier results

For strong mixing processes, uniform error bounds of kernel density estimates $\sup_x |f_n(x) - f(x)|$ have been discussed by Bosq [3] and Fan and Yao [16] among others. Bosq obtained a bound of the form $(n^{-1}\log n)^{2/5} \log\ldots\log n$ under the assumption that the process is exponentially strong mixing. Fan and Yao [16, p. 208] improved Bosq's results by showing that, if the strong mixing coefficient $\alpha(n) = O(n^{-\chi})$ with $\chi > 5/2$,

$$n^{2\chi-5} b_n^{2\chi+5} (\log n)^{-(2\chi+1)/4} \to \infty \quad \text{and} \quad b_n \to 0, \tag{49}$$

then over a compact interval $[c_1, c_2]$, the following weak upper bound holds:

$$\sup_{x\in[c_1,c_2]} |f_n(x) - \mathbb{E}[f_n(x)]| = \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nb_n}}\right). \tag{50}$$

We now compare our results with that of Fan and Yao for linear processes. It is not easy to obtain a sharp bound for the strong mixing coefficient α_k. Consider the special case in which $a_k \sim k^{-\delta}$, $k \in \mathbb{N}$. By the result of Withers [42], one can get $\alpha_k = O(k^{4/3-2\delta/3})$. Restrictive conditions on δ are needed to ensure strong mixing. To apply Fan and Yao's result, one needs to have $\delta > 23/4$ to ensure the strong mixing condition $\alpha(n) = O(n^{-\chi})$ with $\chi > 5/2$. In comparison, however, our Theorem 2 only requires $\delta > 1$ (cf Condition (39)).

Doukhan and Louhichi [10] introduced an interesting weak dependence measure for stationary processes. In their sense a sequence $\{Z_t\}$ is weakly dependent if there exists a sequence $\tau(n) \downarrow 0$ such that, for any u-tuple $(i_1, \dots, i_u) \in \mathbb{Z}^u$ and v-tuple $(j_1, \dots, j_v) \in \mathbb{Z}^v$ with $i_1 \le \dots \le i_u < i_u + n \le j_1 \le \dots \le j_v$, $u, v \in \mathbb{N}$, and any h, k with finite Lipschitz modulus and $\|h\|_\infty, \|k\|_\infty \le 1$,

$$|\mathrm{cov}(h(Z_{i_1}, \dots, Z_{i_u}), k(Z_{j_1}, \dots, Z_{j_v}))| \le [uLip(h) + vLip(k)]\tau(n). \tag{51}$$

Here a function $h : \mathbb{R}^u \to \mathbb{R}$ is said to have finite Lipschitz modulus if

$$Lip(h) := \sup_{x \ne y} \frac{|h(x) - h(y)|}{\|x - y\|_1} < \infty.$$

Ango Nze et al. [1] discussed asymptotic properties of nonparametric estimates under (51) and showed that if the weak dependence coefficient $\tau(n) = O(n^{-r})$ with $r > 4$, then (29) holds, and if $\tau(n) \le a^n$ for some $0 < a < 1$, then

$$\sup_{x \in [-m,m]} |T_n(x) - \mathbb{E}[T_n(x)]| = \mathcal{O}\left(\frac{(\log n)^2}{\sqrt{nb_n}}\right) \text{ almost surely.}$$

Note that the bound in Theorem 4 is sharper. For the linear process, $Z_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$, where ε_t are iid with compact support $[-1, 1]$ (say) and $A_0 := \sum_{i=0}^{\infty} |a_i| < \infty$. Assume $A_0 \le 1$ and let $h(x) = k(x) = \min(\max(x, -1), 1)$. Then $h(Z_t) = k(Z_t) = Z_t$ and $2\tau(n) \ge |\mathrm{cov}(Z_0, Z_n)| = |\sum_{i=0}^{\infty} a_i a_{n+i}|$. As a special case, let $a_n \sim n^{-\beta}$; then $\tau(n) \ge cn^{1-2\beta}$ for some $c > 0$. The condition of Ango Nze et al. requires $\beta > 2.5$, where $\beta > 1$ is sufficient for Theorem 3. For the almost sure convergence, the result of Ango Nze et al. requires exponential decay of $\tau(n)$, which forces a_n to decay exponentially as well, while Theorem 4 only requires that the a_n are summable.

6. Proofs

This section provides proofs of the results in previous sections. Lemma 2 easily follows from Burkholder's inequality, so we omit the details. Lemma 3 below gives a bound for $H_n(x) = \sum_{i=1}^{n} [f_1(x|\mathcal{F}_i) - f(x)]$. It allows long-memory processes.

Lemma 2. *Let D_i, $i \in \mathbb{Z}$, be martingale differences with $D_i \in \mathcal{L}^p$, $p > 1$, $p' = \min(2, p)$. Then there exists a constant $c_p > 0$ such that $\|\sum_i D_i\|_p^{p'} \le c_p \sum_i \|D_i\|_p^{p'}$.*

Lemma 3. *Let $H_n(x) = \sum_{i=1}^{n} [f_1(x|\mathcal{F}_i) - f(x)]$, $H_n'(x) = \frac{\mathrm{d}}{\mathrm{d}x} H_n(x)$, $p > 1$ and $p' = \min(2, p)$. Then we have:*

(i) $\sup_x \|H_n(x)\|_p^{p'} \le c_p \Theta_p(n)$,
(ii) $\int_{\mathbb{R}} \|H_n(x)\|_p^p \mathrm{d}x = O[\bar{\Theta}_p^{p/p'}(n)]$,
(iii) $\sup_x \|H_n'(x)\|_p^{p'} \le c_p \Psi_p(n)$,
(iv) $\int_{\mathbb{R}} \|H_n'(x)\|_p^p \mathrm{d}x = O[\bar{\Psi}_p^{p/p'}(n)]$,
(v) $\mathbb{E}\left[\sup_x H_k^2(x)\right] \le \int_{\mathbb{R}} \|H_k(u)\|^2 + \|H_k'(u)\|^2 \mathrm{d}u = O[\bar{\Theta}_2(k) + \bar{\Psi}_2(k)]$, *and*
(vi) *if* $\bar{\Theta}_2(n) + \bar{\Psi}_2(n) = O(n^\alpha \ell(n))$, *where ℓ is slowly varying, then* $\sup_x |H_n(x)| = \mathcal{O}_{a.s.}(n^{\alpha/2} \tilde{\ell}(n))$, *where $\tilde{\ell}$ is another slowly varying function.*

Proof. Define projection operators $\mathcal{P}_k$, $k \in \mathbb{Z}$, by $\mathcal{P}_k V = \mathbb{E}(V|\mathcal{F}_k) - \mathbb{E}(V|\mathcal{F}_{k-1})$, $V \in \mathcal{L}^1$. Note that $\mathcal{P}_k H_n(x)$, $k = \dots, n-1, n$, are martingale differences. By Theorem 1 in [46], $\|\mathcal{P}_0 f_1(x|\mathcal{F}_i)\|_p \le \theta_{i,p}(x)$. So (i) follows from Lemma 2 in view of

$$\frac{\|H_n(x)\|_p^{p'}}{c_p} \le \sum_{k=-\infty}^{n} \|\mathcal{P}_k H_n(x)\|_p^{p'} \le \sum_{k=-\infty}^{n} \left(\sum_{i=1}^{n} \|\mathcal{P}_k f_1(x|\mathcal{F}_i)\|_p \right)^{p'}$$
$$\le \sum_{k=-\infty}^{n} \left(\sum_{i=1-k}^{n-k} \theta_{i,p}(x) \right)^{p'} \le \Theta_p(n). \tag{52}$$

To prove (ii), let $\delta_k = \sum_{i=1}^{n} \bar{\theta}_p(i-k)$. By Hölder's inequality,

$$\int_{\mathbb{R}} \left(\sum_{i=1-k}^{n-k} \theta_{i,p}(x) \right)^{p} \mathrm{d}x \le \int_{\mathbb{R}} \left[\sum_{i=1}^{n} \frac{\theta_{i-k,p}^{p}(x)}{\bar{\theta}_{i-k,p}^{p-1}} \right] \left[\sum_{i=1}^{n} \bar{\theta}_{i-k,p} \right]^{p-1} \mathrm{d}x = \delta_k^p.$$

If $1 < p \le 2$, (ii) follows from (52). If $p > 2$, then $p' = 2$. Let $q = 1/(1-2/p)$. Again by Hölder's inequality and (52),

$$\int_{\mathbb{R}} \frac{\|H_n(x)\|_p^p}{c_p^{p/2}} \mathrm{d}x \le \int_{\mathbb{R}} \left[\sum_{k=-\infty}^{n} \frac{\left(\sum_{i=1-k}^{n-k} \theta_{i,p}(x) \right)^{p}}{\delta_k^{p-2}} \right] \left[\sum_{k=-\infty}^{n} \delta_k^2 \right]^{p/2-1} \mathrm{d}x = O[\bar{\Theta}_p^{p/2}(n)].$$

Cases (iii) and (iv) can be similarly proved. Case (v) easily follows from the inequality $\sup_x H_k^2(x) \le \int_{\mathbb{R}} |H_k(u)|^2 + |H_k'(u)|^2 \mathrm{d}u$. For (vi), define $\tilde{H}_n = \max_{1 \le k \le n} \sup_x |H_k(x)|$. Let $\ell_0(n) = \ell(n)$ if $\alpha > 1$, and $\ell_0(n) = \sum_{j:2^j \le n} \ell^{1/2}(2^j)$ if $\alpha = 1$; let $\tilde{\ell}(n) = (\log n)^{1/2+\epsilon} \ell_0(n)$, where $\epsilon > 0$. Then $\tilde{\ell}$ is also slowly varying. By Lemma 4 in [47], we get

$$\|\tilde{H}_{2^d}\| \le \sum_{j=0}^{d} 2^{\frac{d-j}{2}} \left\| \sup_{x \in \mathbb{R}} |H_{2^j}(x)| \right\|$$
$$= \sum_{j=0}^{d} \mathcal{O}(1) 2^{\frac{d+j(\alpha-1)}{2}} \ell^{1/2}(2^j) = \mathcal{O}(2^{d\alpha/2} \ell_0(2^d)), \tag{53}$$

which by the Borel–Cantelli lemma implies $\tilde{H}_{2^d} = o_{\text{a.s.}}[2^{d\alpha/2} \tilde{\ell}(2^d)]$ as $d \to \infty$ since

$$\sum_{d=1}^{\infty} \mathbb{P}(\tilde{H}_{2^d} > 2^{d\alpha/2} \tilde{\ell}(2^d)) \le \sum_{d=1}^{\infty} \frac{\|\tilde{H}_{2^d}\|^2}{2^{d\alpha} \tilde{\ell}^2(2^d)} \le \sum_{d=1}^{\infty} \frac{1}{\log^{1+2\epsilon} 2^d} < \infty. \tag{54}$$

Hence $\tilde{H}_n = o_{\text{a.s.}}[n^{\alpha/2} \tilde{\ell}(n)]$ since $\tilde{H}_n$ is non-decreasing and $\tilde{\ell}$ is slowly varying. □

6.1. *Proof of Theorem 1*

Let $H_n(x) = \sum_{t=1}^{n} [f_1(x|\mathcal{F}_{t-1}) - f(x)]$. Write $n\{f_n(x) - \mathbb{E}[f_n(x)]\} = P_n(x) + Q_n(x)$, where

$$P_n(x) = \sum_{t=1}^{n} \{K_{b_n}(x - X_t) - \mathbb{E}[K_{b_n}(x - X_t)|\mathcal{F}_{t-1}]\}, \tag{55}$$

$$Q_n(x) = \sum_{t=1}^{n}\{\mathbb{E}[K_{b_n}(x-X_t)|\mathcal{F}_{t-1}] - \mathbb{E}[K_{b_n}(x-X_t)]\}$$
$$= \int_{\mathbb{R}} K_{b_n}(x-u)H_n(u)\mathrm{d}u = \int_{\mathbb{R}} K(u)H_n(x-b_nu)\mathrm{d}u. \tag{56}$$

By Hölder's inequality,

$$|Q_n(x)|^p \le \int_{\mathbb{R}} |K(u)H_n(x-b_nu)|^p\mathrm{d}u\left[\int_{\mathbb{R}}|K(u)|\mathrm{d}u\right]^{p-1}. \tag{57}$$

By (57) and Lemma 3(i),

$$\sup_x \|Q_n(x)\|_p^p = \sup_x \|H_n(x)\|_p^p O(1) = O[\Theta_p^{p/p'}(n)]. \tag{58}$$

Similarly, by Lemma 3(ii),

$$\int_{\mathbb{R}} \|Q_n(x)\|_p^p\mathrm{d}x \le O(1)\int_{\mathbb{R}}\|H_n(x)\|_p^p\mathrm{d}x = O[\bar{\Theta}_p^{p/p'}(n)]. \tag{59}$$

It remains to deal with the martingale part $P_n(x)$. Let $D_{i,1}(x) = K((x-X_i)/b_n) - \mathbb{E}[K((x-X_i)/b_n)|\mathcal{F}_{i-1}]$ and define recursively $D_{i,k+1}(x) = D_{i,k}^2(x) - \mathbb{E}[D_{i,k}^2(x)|\mathcal{F}_{i-1}]$, $k \in \mathbb{N}$. Then $D_{i,k}(x)$, $i = 1, 2, \ldots$, are martingale differences. Let $l \in \mathbb{N}$ be fixed. We now show that, for any k,

$$A_{n,k}(2^l) := \int_{\mathbb{R}} \mathbb{E}\left|\sum_{i=1}^{n} D_{i,k}(x)\right|^{2^l}\mathrm{d}x = O\left((nb_n)^{2^{l-1}}\right) \tag{60}$$

as $n \to \infty$. To this end, we shall apply the induction method. If $l = 1$, (60) easily follows since the $D_{i,k}(x)$ are orthogonal and $\mathbb{E}[D_{i,k}^2(x)] \le C\mathbb{E}[K^{2^k}((x-X_i)/b_n)] = Cb_n\int_{\mathbb{R}} K^{2^k}(u)f(x-b_nu)\mathrm{d}u$. Let $l \ge 2$. By the Burkholder–Davis–Gundy inequality [6],

$$A_{n,k}(2^l) \le C\int_{\mathbb{R}}\mathbb{E}\left|\sum_{i=1}^{n} D_{i,k}^2(x)\right|^{2^{l-1}}\mathrm{d}x$$
$$\le C\int_{\mathbb{R}}\mathbb{E}\left|\sum_{i=1}^{n} D_{i,1+k}(x)\right|^{2^{l-1}}\mathrm{d}x + C\int_{\mathbb{R}}\mathbb{E}\left|\sum_{i=1}^{n}\mathbb{E}[D_{i,k}^2(x)|\mathcal{F}_{i-1}]\right|^{2^{l-1}}\mathrm{d}x. \tag{61}$$

By the induction hypothesis, the first integral in (61) is of order $O((nb_n)^{2^{l-2}})$. For the second one, note that $\mathbb{E}[D_{i,k}^2(x)|\mathcal{F}_{i-1}] \le C\mathbb{E}[K^{2^k}((x-X_i)/b_n)|\mathcal{F}_{i-1}] \le Cb_n\int_{\mathbb{R}} K^{2^k}(u)f_1(x-b_nu|\mathcal{F}_{i-1})\mathrm{d}u$. Under Condition 1, since $\int_{\mathbb{R}}|K(v)|\mathrm{d}v < \infty$ and $\sup_v|K(v)| < \infty$, the second term in (61) is of order $O((nb_n)^{2^{l-1}})$ in view of the inequality $|\sum_{i=1}^n a_i|^p \le n^{p-1}\sum_{i=1}^n|a_i|^p$, $p \ge 1$. Hence (60) follows and it further implies that, for all $p \in (2^l, 2^{l+1})$, we have

$$A_{n,k}(p) \le [A_{n,k}(2^l)]^{2-p/2^l}[A_{n,k}(2^{l+1})]^{p/2^l-1} = O((nb_n)^{p/2}) \tag{62}$$

by Hölder's inequality. So, by (59), (14) follows if $p \ge 2$. If $1 < p < 2$, by Lemma 2,

$$A_{n,1}(p) \le nC\int_{\mathbb{R}}\|D_{1,1}(x)\|_p^p\mathrm{d}x \le nC\int_{\mathbb{R}}\mathbb{E}|K((x-X_i)/b_n)|^p\mathrm{d}x = O(nb_n).$$

So (14) holds if $1 < p < 2$ in view of (59).

Using an induction argument similar to that of (60) and (61), we have $\sup_x \|\sum_{i=1}^n D_{i,k}(x)\|_p = O((nb_n)^{1/2})$ for all $p = 2^l$, $l \in \mathbb{N}$. Hence by (58), (13) follows. □

6.2. Proof of Theorem 2

Let $R_n = \sup_{|x|\geq n^{5/a}} |f_n(x) - \mathbb{E} f_n(x)|$. Since K is bounded with bounded support and $X_1 \in \mathcal{L}^a$, we have by Markov's inequality that

$$b_n \mathbb{E}(R_n) \leq 2\mathbb{E}\left[\sup_{|x|\geq n^{5/a}} |K[(x - X_1)/b_n]|\right] = O(1)\mathbb{P}(|X_1| \geq n^{5/a}/2) = \frac{O(1)}{n^5}. \quad (63)$$

By the Borel–Cantelli lemma, since $\mathbb{E}(R_n)\sqrt{nb_n}$ is summable, $R_n\sqrt{nb_n} = o_{\text{a.s.}}(1)$. Write $n\{f_n(x) - \mathbb{E}[f_n(x)]\} = P_n(x) + Q_n(x)$ as in Theorem 1. From (56), $\sup_x |Q_n(x)| \leq O(1) \sup_x |H_n(x)|$. By Lemma 3(vi), $\sup_x |H_n(x)| = \mathcal{O}_{a.s.}(n^{\alpha/2}\tilde{\ell}(n))$.

It remains to consider the behavior of $P_n(x)$ over $x \in [-n^{5/a}, n^{5/a}]$. Let

$$Z_t(x) = K_{b_n}(x - X_t) - \mathbb{E}(K_{b_n}(x - X_t)|\mathcal{F}_{t-1}) \quad (64)$$

be the summands of $P_n(x)$. Let $\ell = \lfloor n^{1+5/a+1/\iota} \rfloor$ and $\lfloor x \rfloor_\ell = \lfloor x\ell \rfloor/\ell$. Observe that $|Z_t| \leq 2K_0/b_n$ and $\mathbb{E}(Z_t^2|\mathcal{F}_{t-1}) \leq b_n^{-1} \int_{\mathbb{R}} K^2(u) f(x - b_n u)\mathrm{d}u \leq b_n^{-1} c_1$, where $c_1 = c_0 \int_{\mathbb{R}} K^2(u)\mathrm{d}u$. Let $\tau_n = \sqrt{nb_n^{-1} \log n}$ and $\lambda = 30c_1(1/a + 1/\iota + 1)$. Since $\log n = o(nb_n)$, by the inequality of Freedman [15],

$$\mathbb{P}(|P_n(x)| \geq \sqrt{\lambda}\tau_n) \leq 2\exp\left(\frac{-\lambda\tau_n^2}{4K_0 b_n^{-1}\sqrt{\lambda}\tau_n + 2nb_n^{-1}c_1}\right) = O\left(n^{-\lambda/(3c_1)}\right).$$

Hence $\mathbb{P}(\max_{|x|\leq n^{5/a}} |P_n(\lfloor x \rfloor_\ell)| > \sqrt{\lambda}\tau_n) = O(n^{5/a}\ell n^{-\lambda/(3c_1)}) = o(n^{-2})$, which by the Borel–Cantelli lemma implies that $\max_{|x|\leq n^{5/a}} |P_n(x)| = O_{\text{a.s.}}(\tau_n)$ since $n[n^{5/a}/(\ell b_n)]^\iota = O(\sqrt{n})$, $K \in \mathcal{C}^\iota$ and $\sup_x |P_n(x) - P_n(\lfloor x \rfloor_\ell)| = O(n[n^{5/a}/(\ell b_n)]^\iota) = O(\tau_n)$. □

6.3. Proof of Theorem 3

Recall that $\mathcal{G}_i = (\ldots, \eta_{i-1}, \eta_i; \mathcal{F}_i)$. Let $\zeta_{n,t} = \sqrt{b_n/n} K_{b_n}(x_0 - X_t)$ and $\xi_{n,t} = \zeta_{n,t} Y_t$. Then $\xi_{n,t}$ is $\mathcal{G}_t$-measurable. Let $d_{n,t} = \xi_{n,t} - \mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1})$. Then (29) follows from

$$\sum_{t=1}^n d_{n,t} \Rightarrow N[0,\ V_2(x_0) f(x_0)\kappa] \quad (65)$$

and

$$L_n := \sum_{t=1}^n [\mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1}) - \mathbb{E}\xi_{n,t}] = o_{\mathbb{P}}(1). \quad (66)$$

For (66), since K satisfies Condition 2 and

$$\mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1}) = \sqrt{b_n/n} \int_{\mathbb{R}} g(x_0 - b_n u) K(u) f_1(x_0 - b_n u|\mathcal{F}_{t-1})\mathrm{d}u, \quad (67)$$

by Lemma 3(i), (28), and $\mathbb{E}|H_n(x)| \leq \|H_n(x)\|$,

$$\mathbb{E}|L_n| \leq \sqrt{\frac{b_n}{n}} \int_{\mathbb{R}} |K(u)g(x_0 - b_n u)|\mathbb{E}|H_n(x_0 - b_n u)|\mathrm{d}u$$
$$= O[\sqrt{b_n \Theta_2(n)/n}] = o(1).$$

Next we shall apply the martingale central limit theorem to show (65). Let $\delta = p/2 - 1$. By (67), $n\|\mathbb{E}(\xi_{n,1}|\mathcal{G}_0)\|^2 = O(b_n)$. Again by Lemma 3(i) and (28),

$$\sum_{t=1}^{n}[\mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-1}) - \mathbb{E}(\xi_{n,t}^2)] = \int_{\mathbb{R}} V_2(x_0 - b_n u)K^2(u)\frac{H_n(x_0 - b_n u)}{n}\mathrm{d}u$$
$$= \mathcal{O}_{\mathbb{P}}[\Theta_2^{1/2}(n)/n] = o_{\mathbb{P}}(1).$$

Since $\mathbb{E}(d_{n,t}^2|\mathcal{G}_{t-1}) = \mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-1}) - \mathbb{E}^2(\xi_{n,1}|\mathcal{G}_0)$ and

$$n\mathbb{E}(\xi_{n,t}^2) = \mathbb{E}[Y_t^2 K_{b_n}^2(x_0 - X_t)] = \mathbb{E}[V_2(X_t)K_{b_n}^2(x_0 - X_t)]$$
$$= \int_{\mathbb{R}} V_2(x_0 - b_n u)K^2(u) f(x_0 - b_n u)\mathrm{d}u \to V_2(x_0) f(x_0)\kappa,$$

we have $\sum_{t=1}^{n} \mathbb{E}(d_{n,t}^2|\mathcal{G}_{t-1}) \to V_2(x_0) f(x_0)\kappa$ in probability. Note that

$$n\|d_{n,1}\|_p^p \leq 2^p n\|\xi_{n,1}\|_p^p = 2^p b_n^{1+\delta} n^{-\delta}\mathbb{E}\{V_p(X_t)|K_{b_n}(x_0 - X_t)|^p\}$$
$$= O[(nb_n)^{-\delta}] \to 0.$$

Then Lindeberg's condition is fulfilled. □

6.4. *Proof of Theorem 4*

Write $n\{T_n(x) - \mathbb{E}[T_n(x)]\} = D_n(x) + M_n(x) + N_n(x)$, where

$$D_n(x) = \sum_{t=1}^{n}[Y_t - g(X_t)]K_{b_n}(x - X_t), \tag{68}$$

$$M_n(x) = \sum_{t=1}^{n}\{K_{b_n}(x - X_t)g(X_t) - \mathbb{E}[K_{b_n}(x - X_t)g(X_t)|\mathcal{F}_{t-1}]\}, \quad \text{and} \tag{69}$$

$$N_n(x) = \sum_{t=1}^{n}\{\mathbb{E}[K_{b_n}(x - X_t)g(X_t)|\mathcal{F}_{t-1}] - \mathbb{E}[K_{b_n}(x - X_t)g(X_t)]\}. \tag{70}$$

Then Theorem 4 follows from Proposition 2 and Lemma 4 below.

Proposition 2. *Let the conditions in Theorem 4 be fulfilled. Recall* $z_n = n^{1/p}\log n + (nb_n \log n)^{1/2}$. *Then* (i) $\sup_{x\in[-m,m]} |D_n(x)| = \mathcal{O}_{\text{a.s.}}(z_n/b_n)$ *and* (ii) $\sup_{x\in[-m,m]} |M_n(x)| = \mathcal{O}_{\text{a.s.}}(z_n/b_n)$.

Proof of Proposition 2. Let $\bar{n} = 2^{\lfloor \log n/\log 2\rfloor}$, $Y_i' = Y_i \mathbf{1}_{|Y_i|\leq \bar{n}^{1/p}}$, $Y_i'' = Y_i - Y_i'$. Recall $\mathcal{G}_i = (\ldots, \eta_{i-1}, \eta_i, \mathcal{F}_i)$. Let $Z_t(x) = Z_{t,n}(x) = K_{b_n}(x - X_t)[Y_t' - \mathbb{E}(Y_t'|\mathcal{G}_{t-1})]$,

$$I_n(x) = \sum_{t=1}^{n} Z_t(x) \quad \text{and} \quad II_n(x) = D_n(x) - I_n(x). \tag{71}$$

Elementary calculations show that there exists a constant C_p such that

$$\sum_{d=1}^{\infty} \frac{\mathbb{E} \sum_{t=1}^{2^d} [|Y_t''| + \mathbb{E}(|Y_t''||\mathcal{G}_{t-1})]}{2^{d/p}} = 2 \sum_{d=1}^{\infty} 2^{d-d/p} \mathbb{E}[|Y_1| \mathbf{1}_{|Y_1| \geq 2^{d/p}}] \leq C_p \mathbb{E}[|Y_1|^p].$$

By the Borel–Cantelli lemma, $\sum_{t=1}^{2^d} [|Y_t''| + \mathbb{E}(|Y_t''||\mathcal{G}_{t-1})] = o_{\text{a.s.}}(2^{d/p})$ as $d \to \infty$. Since K is bounded, we have $\sup_x |II_n(x)| = o_{\text{a.s.}}(n^{1/p}/b_n)$.

We shall now deal with $I_n(x)$. Observe that $\mathbb{E}[Z_t(x)|\mathcal{G}_{t-1}] = 0$ and Z_t is bounded by $2K_0 \bar{n}^{1/p}/b_n \leq c_1 n^{1/p}/b_n$, where $c_1 = 2K_0$. Also,

$$\begin{aligned} \mathbb{E}[Z_t^2(x)|\mathcal{F}_t, \eta_{t-1}, \eta_{t-2}, \ldots] &\leq K_{b_n}^2(x - X_t) \mathbb{E}[(Y_t')^2 | \mathcal{F}_t, \eta_{t-1}, \eta_{t-2}, \ldots] \\ &\leq K_{b_n}^2(x - X_t) V_{p'}(X_t) (n^{1/p})^{2-p'}. \end{aligned}$$

Since $V_{p'}(\cdot)$ is bounded in an open interval containing $[-m, m]$, $\mathbb{E}[K_{b_n}^2(x - X_t) V_{p'}(X_t)|\mathcal{G}_{t-1}] \leq c_2 b_n^{-1}$, where c_2 is chosen such that $c_0 \int_{\mathbb{R}} K^2(u) V_{p'}(x - b_n u) \mathrm{d}u \leq c_2$. For all $x \in [-m, m]$, by the inequality of Freedman [15],

$$\mathbb{P}[|I_n(x)| \geq \lambda z_n / b_n] \leq 2 \exp \left[\frac{-\lambda^2 z_n^2 / b_n^2}{2(c_1 n^{1/p}/b_n)(\lambda z_n/b_n) + 2c_2 n^{1+2/p-p'/p}/b_n} \right].$$

For $x \in \mathbb{R}$ let $\lfloor x \rfloor_\ell = \lfloor x\ell \rfloor / \ell$, where $\ell = \lfloor n^{8/\iota} \rfloor$. Then $\mathbb{E}[\sup_x |I_n(x) - I_n(\lfloor x \rfloor_\ell)|] = O(\ell^{-\iota}/b_n^2) \mathbb{E}|Y_0| = O(n^{-5})$. Using the argument in the proof of Theorem 2, we have (i) by letting $\lambda = 16(c_2^{1/2} + c_1)(1 + \iota^{-1})$. (ii) can be similarly proved and the truncation argument is not needed since $K_{b_n}(x - X_t) g(X_t)$ is bounded on $[-m, m]$. □

Lemma 4. *With Condition 2, for any $m > 0$, $\mathbb{E}[\sup_{|x| \leq m} N_n^2(x)] = O[\Theta_2(n) + \Psi_2(n)]$. Furthermore, if $\Theta_2(n) + \Psi_2(n) = O(n^\alpha \ell(n))$, where ℓ is a slowly varying function, then $\sup_{|x| \leq m} |N_n(x)| = \mathcal{O}_{a.s.}(n^{\alpha/2} \tilde{\ell}(n))$, where $\tilde{\ell}(n)$ can be chosen as the one in Lemma 3*(vi).

Proof. Let $H_n(x) = \sum_{i=1}^n [f_1(x|\mathcal{F}_{i-1}) - f(x)]$. Since K has bounded support, and g is bounded on compact sets, we have $\sup_x |N_n(x)| = O(1) \sup_x |H_n(x)|$ in view of

$$N_n(x) = \int_{\mathbb{R}} K(u) g(x - b_n u) H_n(x - b_n u) \mathrm{d}u, \tag{72}$$

Observe that $\sup_{|u| \leq m} |H_n(u)| \leq |H_n(0)| + \int_{-m}^m |H_n'(x)| \mathrm{d}x$. Then the result follows from Lemma 3(i), (iii) and (vi). □

Acknowledgements

We are grateful to two referees and an Associate Editor for their many helpful comments. This work was supported in part from DMS-0906073 and DMS-0448704.

References

[1] P. Ango Nze, P. Bühlmann, P. Doukhan, Weak dependence beyond mixing and asymptotics for nonparametric regression, Ann. Statist. 30 (2002) 397–430.

[2] P.J. Bickel, M. Rosenblatt, On some global measures of the deviations of density function estimates, Ann. Statist. 1 (1973) 1071–1095.

[3] D. Bosq, Nonparametric Statistics for Stochastic Processes : Estimation and Prediction, in: Lecture Notes in Statist., vol. 110, Springer, New York, 1996.
[4] R.C. Bradley, Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions, Probab. Surveys 2 (2005) 107–144.
[5] J.V. Castellana, M.R. Leadbetter, On smoothed probability density estimation, Stochastic Process. Appl. 21 (1986) 179–193.
[6] Y.S. Chow, H. Teicher, Probability Theory, second ed., Springer, New York, 1988.
[7] S. Csörgő, J. Mielniczuk, Density estimation under long-range dependence, Ann. Statist. 23 (1995) 990–999.
[8] L. Devroye, L. Györfi, Nonparametric Density Estimation: The L^1 View, Wiley, New York, 1985.
[9] P. Diaconis, D. Freedman, Iterated random functions, SIAM Rev. 41 (1999) 41–76.
[10] P. Doukhan, S. Louhichi, A new weak dependence condition and applications to moment inequalities, Stochastic Process. Appl. 84 (1999) 313–342.
[11] J. Dedecker, C. Prieur, Coupling for τ-dependent sequences and applications, J. Theor. Probab. 17 (2004) 861–885.
[12] P. Doukhan, O. Wintenberger, Weakly dependent chains with infinite memory, Stochastic Process. Appl. 118 (2008) 1997–2013.
[13] R. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50 (1982) 987–1007.
[14] R. Eubank, Nonparametric Regression and Spline Smoothing, second ed., Marcel Dekker, New York, 1999.
[15] D.A. Freedman, On tail probabilities for martingales, Ann. Probab. 3 (1975) 100–118.
[16] J.Q. Fan, Q. Yao, Nonlinear Time Series. Nonparametric and Parametric Methods, Springer, New York, 2003.
[17] L. Györfi, W. Härdle, P. Sarda, P. Vieu, Nonparametric Curve Estimation from Time Series, in: Lecture Notes in Statist., vol. 60, Springer, Berlin, 1989.
[18] V. Haggan, T. Ozaki, Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, Biometrika 68 (1981) 189–196.
[19] H.C. Ho, T. Hsing, On the asymptotic expansion of the empirical process of long memory moving averages, Ann. Statist. 24 (1996) 992–1024.
[20] W. Härdle, H. Lutkepohl, R. Chen, A review of nonparametric time series analysis, Int. Stat. Rev. 65 (1997) 49–72.
[21] T. Kawata, The function of mean concentration of a chance variable, Duke Math. J. 8 (1941) 666–677.
[22] G. Kallianpur, Some ramifications of Wiener's ideas on nonlinear prediction, in: P. Masani (Ed.), Norbert Wiener, Collected Works, 3, MIT Press, Cambridge, 1981, pp. 402–424.
[23] W. Liu, W.B. Wu, Simultaneous nonparametric inference of time series, Ann. Stat. 38 (2010) 2388–2421.
[24] D.F. Nicholls, B.G. Quinn, Random Coefficient Autoregressive Models: An Introduction, Springer, New York, 1982.
[25] E.A. Nadaraya, Nonparametric Estimation of Probability Densities and Regression Curves, Kluwer Academic Pub, 1989.
[26] M.H. Neumann, Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations, Ann. Statist. 26 (1998) 2014–2048.
[27] M.H. Neumann, J.P. Kreiss, Regression-type inference in nonparametric autoregression, Ann. Statist. 26 (1998) 1570–1613.
[28] B.L.S. Prakasa Rao, Nonparametric Functional Estimation, Academic Press, New York, 1983.
[29] M.B. Priestley, Non-linear and Non-stationary Time Series Analysis, Academic Press, New York, 1988.
[30] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Statist. 27 (1956) 832–837.
[31] M. Rosenblatt, Density estimates and Markov sequences, in: M.L. Puri (Ed.), Nonparametric Techniques in Statistical Inference, Cambridge Univ. Press, London, 1970, pp. 199–213.
[32] P.M. Robinson, Nonparametric estimators for time series, J. Time Ser. Anal. 4 (1983) 185–207.
[33] M. Rosenblatt, A comment on a conjecture of N. Wiener, Statist. Probab. Lett. 79 (2009) 347–348.
[34] W. Stute, A law of the logarithm for kernel density estimators, Ann. Probab. 10 (1982) 414–422.
[35] R.S. Singh, A. Ullah, Nonparametric time-series estimation of joint DGP, conditional DGP, and vector autoregression, Econometric Theory 1 (1985) 27–52.
[36] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
[37] H. Tong, Threshold models in non-linear time series analysis, in: Lecture Notes in Statistics, vol. 21, Springer, New York, 1983.
[38] H. Tong, Non-linear Time Series: A Dynamical System Approach, Oxford Univ. Press, 1990.
[39] D. Tjostheim, Nonlinear time series, a selective review, Scand. J. Statist. 21 (1994) 97–130.
[40] R. Tsay, Analysis of Financial Time Series, Wiley, New York, 2005.

[41] N. Wiener, Nonlinear Problems in Random Theory, MIT Press, John Wiley, 1958, pp. 118–128.
[42] C.S. Withers, Conditions for linear process to be strongly mixing, Z. Wahrsch. Verw. Gebiete. 57 (1981) 477–480.
[43] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman and Hall, London, 1995.
[44] W.B. Wu, J. Mielniczuk, Kernel density estimation for linear processes, Ann. Statist. 30 (2002) 1441–1459.
[45] W.B. Wu, X. Shao, Limit theorems for iterated random functions, J. Appl. Probab. 41 (2004) 425–436.
[46] W.B. Wu, Nonlinear system theory: another look at dependence, Proc. Natl Acad. Sci. 102 (2005) 14150–14154.
[47] W.B. Wu, On the Bahadur representation of sample quantiles for dependent sequences, Ann. Statist. 33 (2005) 1934–1963.
[48] W.B. Wu, Empirical processes of stationary sequences, Statist. Sinica 18 (2008) 313–333.
[49] B. Yu, Density estimation in the L^{∞} norm for dependent data with applications to the Gibbs sampler, Ann. Statist. 21 (1993) 711–735.