Recursive Nonparametric Estimation For Time Series

Yinxiao Huang, Xiaohong Chen, and Wei Biao Wu

***Abstract*—The paper considers online kernel estimation for both short- and long-range dependent time series data. Utilizing the predictive dependence measure of Wu (2005), we carefully study the asymptotic properties of recursive kernel density and regression estimators for a general class of stationary processes. In particular, we prove that the proposed estimators have the asymptotic normality and the corresponding central limit theorems are provided. In addition, we establish the sharp laws of the iterated logarithms that precisely characterize the asymptotic almost sure behavior of the proposed estimators.**

***Index Terms*—Almost sure convergence; kernel estimation; law of the iterated logarithm; long-range dependence; recursive estimation; wavelet estimation.**

I. INTRODUCTION

WITH the advance of modern computing and data acquisition techniques, voluminous data are collected in various applied areas including astronomy, computer networks, remote sensing, weather monitoring, high frequency trading among others. New challenges arise as to how to interpret and model the data. Nonparametric methods are widely used as a powerful tool for data analysis, modeling and inference. They are especially useful to explore relationships between explanatory variables and dependent variables in situations in which the associated functional forms are not known. Popular nonparametric estimators such as the kernel density estimator of Rosenblatt [40] and the kernel regression estimator of Nadaraya [32] and Watson [55], however, suffer a serious computational drawback in that they are non-recursive. When a new data item arrives, one has to re-calculate any non-recursive estimator based on all of the observations. In situations where data items arrive sequentially and a large amount of data can be generated at a rapid rate, recursive updating algorithms are much preferred. Namely the value of the estimator at current time can be updated from its immediate past and the current data item, so that the update can be computed instantly regardless of the sample size. This approach is also termed as the *online* or *real-time updating* method. The traditional non-recursive procedure (such as the Rosenblatt estimate and the Nadaraya-Watson estimate) is called the *off-line* or *batch* method.

There are already huge amount of published papers on asymptotic properties of on-line kernel density and regression estimators for independent and identically distributed (i.i.d.) data. In particular, recursive kernel density estimators were introduced by Wolverton and Wagner [58], Yamato [63] and Deheuvels [6], and have been studied by Wegman and Davies [56], Wertz [57], Roussas [41], Hall and Patil [18], Mokkadem et al. [30] and many others. Different versions of recursive kernel regression estimators were introduced by Révész [35], [36], Ahmad and Lin [2] and Devroye and Wagner [7], and have been studied by Krzyżak and Pawlak [24], Greblicki and Pawlak [11], Krzyżak [23], Walk [52], Györfi et al. [13], Mokkadem et al. [31], among many others.

There are some published papers on asymptotic properties of recursive kernel density and regression estimators for short-range (i.e., weakly) dependent data. See, e.g., Takahata [47], Masry [27], [28], Masry and Györfi [29], Györfi and Masry [15], Tran [49], [50] and others for on-line kernel density estimator, and Roussas and Tran [42], Györfi et al. [14], Walk and Yakowitz [53], Wang and Liang [54], Amiri [3] and others for on-line kernel regression estimator for stationary short-range dependent data.

In this paper we shall study the on-line kernel estimators for density and regression functions allowing for both short- and long-range dependent data. We model the temporal dependence using the functional and predictive dependence measure of Wu [59]. The predictive dependence concept has a natural interpretation for on-line procedures since it measures how a dynamic system reacts to a change in the underlying exogenous shock. This notion of dependence is very general and easy-to-verify. It covers a large class of linear and nonlinear time series models, which could be short-range (i.e., weakly) or long-range (i.e., strongly) dependent. To the best of our knowledge, there is no published work on asymptotic properties of recursive nonparametric estimators for time series models using the predictive dependence measure. In this paper, for a large class of stationary time series that are both short- and long-range dependent, we establish the asymptotic properties for both the recursive density and the recursive regression estimators. Due to the online nature, the study of the almost sure convergence properties becomes more relevant than the weaker form of in probability convergence. Here we shall give a precise characterization of the almost sure convergence by proving the sharp laws of the iterated logarithm. We also establish the asymptotic normality, and the

Manuscript received Nov 26, 2012; revised Nov 03, 2013; accepted Nov 14, 2013. Date of current version Nov 16, 2013. W. B. Wu is supported by the NSF under Grant DMS-0906073 and DMS-1106970. X. Chen is supported by Cowles Foundation for Research in Economics, USA.

Y. Huang is with the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA. Email:yinxiao@illinois.edu.

X. Chen is with Department of Economics, Yale University, New Haven, CT 06520 USA. Email:xiaohong.chen@yale.edu.

W. B. Wu is with Department of Statistics, The University of Chicago, Chicago, IL 60637 USA. Email:wbwu@galton.uchicago.edu.

This paper was presented in part at the Joint Statistical Meetings 2012.

almost sure version of the optimal uniform convergence rates. Our results substantially extend the applicability of the online methods to time series estimation and forecasting.

In the literature, there are also plenty of work on other on-line nonparametric procedures. For example, Vilar and Vilar [51] and Gu and Lafferty [12] considered recursive local polynomial regression. Smale and Yao [44] and Yao [64] studied an on-line estimate based on stochastic gradient descent algorithms (Robbins and Monro [39]; Kiefer and Wolfowitz [22]; Duflo [8]; Kushner and Yin [25]. Smale and Zhou [45] studied on-line learning with reproducing kernel Hilbert space under Markov sampling. Chen and White [4] studied a class of projection based Robbins-Monro procedures in a Hilbert space for weakly dependent data. The mathematical tools used in our paper could also be extended to the settings of these papers. For example, in Section III-C we provide a law of the iterated logarithm and an asymptotic normality for recursive wavelet density estimators allowing for both short- and long-range dependent data.

The rest of paper is structured as follows. Section II presents the recursive kernel estimates and the functional and the predictive dependence measures. The latter allows us to consider both linear and non-linear processes under short- and long-range dependence. Under this framework, we investigate the limit properties of the recursive density estimator and the recursive regression estimator in Sections III and IV respectively. Besides the usual central limit theorem and strong (weak) convergence, a sharp law of the iterated logarithm is derived in each section. Section V provides some simulation results that corroborate the theoretical findings of the article. The appendix contains all of the proofs.

The following notation is used throughout the paper. For two real numbers a and b, we write $a \wedge b = \min(a, b)$. For two sequences of real numbers $\{s_n\}$ and $\{t_n\}$, we write $s_n \sim t_n$ if $\lim_{n\to\infty} s_n/t_n = 1$; we write $s_n \asymp t_n$ if there exists a constant $c > 0$ such that $c \le s_n/t_n \le 1/c$ hold for all large n. We say that a function g is Lipschitz continuous on a set A if there exists a constant $C < \infty$ such that $|g(x) - g(x')| \le C|x - x'|$ for all $x, x' \in A$. The symbol C denotes a positive constant which may differ in different occurrences. For a random variable X write $X \in \mathcal{L}^p$, $p \ge 1$, if $\|X\|_p := [\mathbb{E}(|X|^p)]^{1/p} < \infty$ and write $\|X\| := \|X\|_2$. The symbol $\Rightarrow$ means convergence in distribution. Denote by $\mathcal{C}^k$ the class consists of all continuous functions that are k-times differentiable with continuous k-th derivatives.

II. Recursive Estimates and Dependence Conditions

A. Recursive Kernel Estimators

Given the data $(X_i, Y_i)_{i=1}^n$, following the classic Nadaraya-Watson [32], [55] procedure, we can estimate the regression function $g(x) = \mathbb{E}(Y|X = x)$ by

$$\hat{g}_n(x) = \frac{\sum_{i=1}^n Y_i K_{b_n}(x, X_i)}{\sum_{i=1}^n K_{b_n}(x, X_i)}, \tag{1}$$

where $K_{b_n}(x, u) = K((x - u)/b_n)$, K is a kernel function and b_n is the bandwidth sequence. There is a huge literature on properties of the estimator (1); see for example Györfi et al. [13] and Fan and Yao [9]. A drawback of $\hat{g}_n$ is that it is non-recursive. There is no simple algebraic relationship between $\hat{g}_n$ and $\hat{g}_{n+1}$. If a new data item (X_{n+1}, Y_{n+1}) arrives and $b_n \ne b_{n+1}$, then one needs to re-calculate all $K_{b_{n+1}}(x, X_i)$ for $i = 1, \ldots, n$, thus having computational complexity $\mathcal{O}(n)$. The latter problem becomes more severe if one wants to evaluate g at multiple points $x_1, \ldots, x_l$ since in this case the computational complexity becomes $\mathcal{O}(nl)$. The memory requirement is also $\mathcal{O}(n)$ since one has to store all the past data. These drawbacks seriously restrict the applicability of the Nadaraya-Watson method in applications that involve a large amount of data and require frequent updates.

As a simple modification of (1), we can have the following online version:

$$\tilde{g}_n(x) = \frac{\sum_{i=1}^n Y_i K_{b_i}(x, X_i)/\sum_{i=1}^n b_i}{\sum_{i=1}^n K_{b_i}(x, X_i)/\sum_{i=1}^n b_i} := \frac{J_n(x)}{\tilde{f}_n(x)}, \tag{2}$$

where $\tilde{f}_n(x)$ is the online estimator for the marginal density of X_i. The estimators admit the following recursive update scheme: letting $B_n = \sum_{i=1}^n b_i$,

$$\begin{aligned} \tilde{f}_n(x) &= \tilde{f}_{n-1}(x) + B_n^{-1} b_n [K_{b_n}(x, X_n)/b_n - \tilde{f}_{n-1}(x)], \\ \tilde{g}_n(x) &= \tilde{g}_{n-1}(x) + \gamma_n [Y_n - \tilde{g}_{n-1}(x)], \end{aligned} \tag{3}$$

where $\gamma_n = \gamma_n(x) = (\tilde{f}_n(x) B_n)^{-1} K_{b_n}(x, X_n)$ is the step size that is a stochastic function of x. Note that the update can be computed within $\mathcal{O}(1)$ step and the memory complexity for the algorithm is also $\mathcal{O}(1)$. When the quantity $\tilde{f}_n(x) B_n$ in (3) is replaced by $n b_n$ then the estimator $\tilde{g}_n$ is very similar to the recursive estimator obtained by the method of stochastic approximation (Révész [35]). In (3), $\tilde{e}_n := Y_n - \tilde{g}_{n-1}(x)$ can be regarded as the *one-step-ahead prediction error*. The new information at time n comes in the form of $\tilde{e}_n$ which is used to update g together with the previous estimator $\tilde{g}_{n-1}$. The residuals $\tilde{e}_n$ so generated are called *nonparametric recursive residuals*.

B. Causal Time Series, Dependence Measures

Throughout the paper we assume that X_i is a (strictly) stationary process of the form

$$X_i = R(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{4}$$

where ε_i, $i \in \mathbf{Z}$, are i.i.d. random variables and R is a measurable function such that X_i is well-defined with a marginal density function f. In our regression model we assume

$$Y_i = G(X_i, e_i), \tag{5}$$

where e_i, $i \in \mathbf{Z}$, are i.i.d. and e_i is independent of $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Then the regression function $g(x) = \mathbb{E}(Y_i|X_i = x) = \mathbb{E}[G(x, e_i)]$.

The class of processes that (4) represents is huge and it includes many commonly used linear and nonlinear processes; see Wu [60] for a review. Section III concerns properties of the on-line estimate $\tilde{f}_n(x)$ for the marginal density function f of X_i. An important special case of our regression model (5) is the nonlinear time series model

$$X_{i+1} = G(X_i, \varepsilon_{i+1}), \tag{6}$$

where ε_i are i.i.d., by letting $Y_i = X_{i+1}$ and $e_i = \varepsilon_{i+1}$. Under suitable condition on G, (6) has a (strictly) stationary and ergodic solution (Wu and Shao [62]).

Based on the representation (4), we can introduce our dependence conditions. For $l \ge 1$ let $F_l(x|\mathcal{F}_k) = P(X_{k+l} \le x|\mathcal{F}_k)$ be the l-step ahead conditional or predictive distribution function of X_{k+l} given $\mathcal{F}_k$. Assume that the conditional density $f_l(x|\mathcal{F}_k) = dF_l(x|\mathcal{F}_k)/dx$ exists, hence the marginal density $f(x) = \mathbb{E} f_l(x|\mathcal{F}_k)$. Assume throughout the paper that there exists a constant $C > 0$ such that

$$\operatorname*{ess\,sup}_{\mathcal{F}_k} \sup_{x\in\mathbf{R}} f_1(x|\mathcal{F}_k) \le C. \tag{7}$$

Let $\mathcal{P}_l(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_l) - \mathbb{E}(\cdot|\mathcal{F}_{l-1})$ be a projection operator. Following Wu [59], for $k \ge 1$, define the *predictive dependence measure*

$$\theta_k = \sup_{x\in\mathbf{R}} \|\mathcal{P}_0 f_1(x|\mathcal{F}_{k-1})\|. \tag{8}$$

Let $\varepsilon_i', \varepsilon_k, i, k \in \mathbf{Z}$, be i.i.d. If $l \le k$, let $\mathcal{F}_{k,\{l\}} = (\ldots, \varepsilon_{l-1}, \varepsilon_l', \varepsilon_{l+1}, \ldots, \varepsilon_k)$ be a coupled version of $\mathcal{F}_k$ with ε_l in the latter replaced by ε_l'. If $l > k$ let $\mathcal{F}_{k,\{l\}} = \mathcal{F}_k$. Note that for $k \ge 1$, $\mathbb{E}(f_1(x|\mathcal{F}_{k-1})|\mathcal{F}_0) = f_k(x|\mathcal{F}_0)$ and $\mathbb{E}(f_1(x|\mathcal{F}_{k-1})|\mathcal{F}_{-1}) = f_{1+k}(x|\mathcal{F}_{-1}) = \mathbb{E}(f_k(x|\mathcal{F}_{0,\{0\}})|\mathcal{F}_0)$. We can then interpret θ_k as the contribution of ε_0 in predicting X_k. If $f_1(x|\mathcal{F}_{k-1})$ does not functionally depend on ε_0, then $\theta_k = 0$. A simple upper bound for θ_k is the following *functional dependence measure*

$$\delta_k = \sup_{x\in\mathbf{R}} \|f_1(x|\mathcal{F}_{k-1}) - f_1(x|\mathcal{F}_{k-1,\{0\}})\|. \tag{9}$$

Note that $\mathbb{E}(f_1(x|\mathcal{F}_{k-1,\{0\}})|\mathcal{F}_0) = \mathbb{E}(f_1(x|\mathcal{F}_{k-1})|\mathcal{F}_{-1})$, we have $\theta_k \le \delta_k$. For $k < 0$, we have $\mathcal{F}_{k,\{0\}} = \mathcal{F}_k$ and thus $\theta_k = \delta_k = 0$.

In many cases it is easier to deal with δ_k. Proposition 1 provides bounds for δ_k for linear processes, which can have infinite variances, for example, α-stable processes with $0 < \alpha < 2$. Proposition 2 concerns nonlinear time series that can be expressed as iterated random functions.

PROPOSITION **1**. *Let ε_j be i.i.d. with density $f_\varepsilon \in \mathcal{C}^1$ and finite α-th moment, $\alpha > 0$. Assume that $(a_j)_{j=0}^\infty$ are real coefficients satisfying $\sum_{j=0}^\infty |a_j|^{2\wedge\alpha} < \infty$ where $a\wedge b = \min(a,b)$. Then the linear process*

$$X_t = \sum_{j=0}^\infty a_j \varepsilon_{t-j} \tag{10}$$

is well-defined, and, if $\sup_x(f_\varepsilon(x) + |f_\varepsilon'(x)|) < \infty$, then $\delta_k^2 = \mathcal{O}(|a_k|^{\alpha\wedge 2})$.

PROPOSITION **2**. *Consider the nonlinear time process $X_{i+1} = G(X_i, \varepsilon_{i+1})$ in (6), where ε_i are i.i.d. and $G(\cdot,\cdot)$ satisfies $G(x,\varepsilon_0) \in \mathcal{L}^\tau$, $\tau \ge 1$ and*

$$L := \sup_{x\ne x'} \frac{\|G(x,\varepsilon_0) - G(x',\varepsilon_0)\|_\tau}{|x-x'|} < 1. \tag{11}$$

Assume that the conditional density $f(x|u)$ of X_t given $X_{t-1} = u$ is uniformly bounded with $\sup_{x,u} f(x|u) \le C_0$ and satisfies the uniform Hölder continuous condition

$$\sup_{u\ne u'} \frac{\sup_x |f(x|u) - f(x|u')|}{|u-u'|^\alpha} \le H, \ \alpha > 0. \tag{12}$$

Then $\delta_k = \mathcal{O}(\rho^k)$ for some $\rho \in (0,1)$.

EXAMPLE **1**. *(Thresholded Time Series). Following Tong [48], we consider the thresholded autoregressive process*

$$X_t = a_1 X_{t-1}^+ + a_2 X_{t-1}^- + \varepsilon_t, \tag{13}$$

where $u^+ = \max(u,0)$, $u^- = \min(u,0)$, $|a_1| < 1$, $|a_2| < 1$ and ε_t are i.i.d. with density function f_ε and have a finite τth moment, $\tau > 0$. Let $g(u) = a_1 u^+ + a_2 u^-$. Then $f(x|u) = f_\varepsilon(x - g(u))$. Assume $H = \sup_w |f_\varepsilon'(w)| < \infty$. Clearly (11) and (12) hold with $L = \max(|a_1|, |a_2|) < 1$, $\alpha = 1$ and this H. Note that the moment condition on ε_t is very mild. It holds, for example, if ε_t is Cauchy, since in this case $\mathbb{E}(|\varepsilon_t|^\tau) < \infty$ if $0 < \tau < 1$.

EXAMPLE **2**. *(ARCH Processes). Assume that ε_t are i.i.d. with density f_ε and have a finite τth moment, $\tau > 0$. Define*

$$X_t = (a_1^2 + a_2^2 X_{t-1}^2)^{1/2} \varepsilon_t, \tag{14}$$

where a_1 and a_2 are real parameters such that $\|a_2\varepsilon_0\|_\tau < 1$. Then (11) holds and (14) has a stationary solution. Let $g(u) = (a_1^2 + a_2^2 u^2)^{1/2}$. Then the conditional density has the form $f(x|u) = f_\varepsilon(x/g(u))/g(u)$. Assume that $\sup_w |f_\varepsilon'(w)| < \infty$. Elementary calculations show that $\sup_{x,u} |\partial f(x|u)/\partial u| < \infty$, implying (12) with $\alpha = 1$. As in Example 1 the moment condition here on ε_t can also be very mild.

Our dependence measure is very different from the widely-used strong mixing conditions and their variants such as absolute regularity and near epoch dependence. In the context of asymptotic theory for recursive kernel estimates, it turns out that our predictive and functional dependence measures are very useful and convenient. Optimal and nearly optimal results can be established with elegant and concise proofs.

In our asymptotic theory we can allow both short- and long-range dependent processes:

(D1). (Short-range dependence) $\Delta_m := \sum_{k=m}^\infty \delta_k < \infty$. Hence $\Theta_m := \sum_{k=m}^\infty \theta_k \le \Delta_m < \infty$;

(D2). (Long-range dependence) $\delta_k = \mathcal{O}(k^{-\gamma}\ell(k))$, where $1/2 < \gamma < 1$ and $\ell(\cdot)$ is a slowly-varying function, namely $\lim_{z\to\infty} \ell(cz)/\ell(z) = 1$ for any $c > 0$.

Condition (**D1**) implies that the cumulative contribution of the input ε_0 in predicting future values $\{X_k\}_{k\ge 1}$ is finite, thus suggesting short-range dependence. It is satisfied for many nonlinear time series models (see Proposition 2 and Examples 1 and 2) and short-range dependent linear processes (see Proposition 1). The other condition (**D2**) indicates that ε_0 can have a long-range predictive capability, due to the slow decay of δ_k. An important example is the fractionally integrated autoregressive moving average process.

III. RECURSIVE DENSITY ESTIMATION

Let f be the marginal density function of the stationary process (X_i) given in (4). As in (2), we consider the recursive kernel density estimator

$$\tilde{f}_n(x) = \frac{1}{B_n} \sum_{i=1}^n K_{b_i}(x, X_i), \tag{15}$$

where (b_i) is the bandwidth sequence, $B_n = \sum_{i=1}^n b_i$, $K_b(x,u) = K((x-u)/b)$ and $K(\cdot)$ is a kernel function. Throughout the paper, we assume that the kernel function satisfies the following condition:

(**K**). The kernel K has a bounded support $[-M, M]$; there exists $C_K < \infty$ such that $\sup_u |K(u)| + \int_{\mathbf{R}} u^2 |K(u)| du \le C_K$. Let $\kappa := \int_{\mathbf{R}} K^2(s) \mathrm{d}s < \infty$.

Condition (**K**) is satisfied by many popular choices of kernels including the rectangle kernel $K(v) = \mathbf{1}_{|v|<1}/2$, the Epanechnikov kernel $K(v) = 3(1-v^2)\mathbf{1}_{|v|<1}/4$, the triangular kernel $K(v) = (1-|v|)\mathbf{1}_{|v|<1}$, the biweight kernel $K(v) = 15(1-v^2)^2\mathbf{1}_{|v|<1}/16$ and the triweight kernel $K(v) = 35(1-v^2)^3\mathbf{1}_{|v|<1}/32$ among others, where $\mathbf{1}.$ is the indicator function.

A. Asymptotic Properties

Let $B_{n,l} = \sum_{i=1}^n b_i^l$, $l \in \mathbf{Z}$, and write $B_n = B_{n,1}$. The following theorem provides the central limit theorem of the recursive kernel density estimator (15).

THEOREM **1**. *Assume condition* (**K**), $\sup_u |f'(u)| < \infty$, *and either one of the following:*

(i) *Condition* (**D1**), $B_n \to \infty$, $\sum_{k=1}^\infty B_{2^{1+k},2}/B_{2^k} < \infty$, *and* $B_{n,2} = o(B_n)$;

(ii) *Condition* (**D2**), $b_i \asymp i^{-\beta}\ell_0(i)$ *where* $2(1-\gamma) < \beta < 1$ *and* $\ell_0(\cdot)$ *is a slowly varying function.*

Then we have the central limit theorem

$$\sqrt{B_n}[\tilde{f}_n(x_0) - \mathbb{E}\tilde{f}_n(x_0)] \Rightarrow N(0, \kappa f(x_0)). \tag{16}$$

If in addition $B_{n,3}^2 = o(B_n)$, *then (16) holds with* $\mathbb{E}\tilde{f}_n(x_0)$ *replaced by* $f(x_0)$.

We shall briefly discuss the bandwidth condition for the short-range dependence case (i) and the long-range dependence case (ii) in Theorem 1. Motivated by the traditional kernel estimates, a popular choice of the bandwidth sequence is $b_i \asymp i^{-\beta}$, where $0 < \beta < 1$. Then under (**D1**), $B_n = \sum_{i=1}^n b_i \asymp n^{1-\beta}$, and $B_{n,2} = \sum_{i=1}^n b_i^2 \asymp n^{1-2\beta}$ (resp. $\log(n)$ or $\mathcal{O}(1)$) for $\beta \in (0, 1/2)$ (resp. $\beta = 1/2$ or $\beta \in (1/2, 1)$). Thus the bandwidth condition in Theorem 1(i) is satisfied. On the other hand, if the underlying process exhibits long-range dependence as in case (ii), then a more restrictive condition $2(1-\gamma) < \beta < 1$ is needed.

The following theorem provides the law of iterated logarithm (LIL) for the recursive kernel density estimator (15). For a sequence of random variables W_n and a real number c, we write $\overline{\underline{\lim}}_{n\to\infty} W_n = \pm c$ if $\limsup_{n\to\infty} W_n = c$ and $\liminf_{n\to\infty} W_n = -c$.

THEOREM **2**. *Assume condition* (**K**), $\sup_u |f'(u)| < \infty$, *and either one of the following:*

(i) *Condition* (**D1**), $B_n \to \infty$, $(\log n)^4 + B_{n,2} \log n = \mathcal{O}(B_n)$ *and* $\sum_{k=1}^\infty B_{2^{1+k},2}/B_{2^k} < \infty$;

(ii) *Condition* (**D2**), $b_i \asymp i^{-\beta}\ell_0(i)$ *where* $2(1-\gamma) < \beta < 1$ *and* $\ell_0(\cdot)$ *is a slowly varying function.*

Then we have the law of the iterated logarithm

$$\overline{\underline{\lim}}_{n\to\infty} \frac{\sqrt{B_n}[\tilde{f}_n(x_0) - \mathbb{E}\tilde{f}_n(x_0)]}{\sqrt{\log\log B_n}} = \pm\sqrt{2\kappa f(x_0)}. \tag{17}$$

We remark that generally there is no LIL result for non-recursive estimators unless the bandwidths $\{b_n\}$ vary very slowly. For i.i.d. random variables, Hall [16] proved a LIL for the non-recursive kernel density estimator $\hat{f}_n(x) = (nb_n)^{-1}\sum_{i=1}^n K_{b_n}(x, X_i)$ under

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_{m:|m-n|\le n\varepsilon} |b_m/b_n - 1| = 0. \tag{18}$$

See also Härdle [19]. If (18) is violated, for example, $b_n = (2+(-1)^n)n^{-\beta}$, then one can only obtain an almost sure upper bound of the form $\sqrt{B_n^{-1}\log n}$ for $\hat{f}_n(x_0) - \mathbb{E}\hat{f}_n(x_0)$.

B. An Alternative Form

As an alternative of the recursive kernel density estimate (15), one can use the form

$$\tilde{f}_n^*(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{b_i} K_{b_i}(x, X_i). \tag{19}$$

Under the assumption that X_i are i.i.d., the above estimator were studied independently by Wolverton and Wagner [58] and Yamato [63]. Takahata [47] and Masry and Györfi [29] obtained almost sure rates of $\tilde{f}_n^*$ to f for the class of asymptotically uncorrelated processes. Nguyen [33], [34] proved asymptotic normality of $\tilde{f}_n^*(x)$ for Markovian processes that satisfy the G2 geometric beta-mixing condition. Robinson [37], [38] derived weak consistency and asymptotic normality of $\tilde{f}_n^*(x)$ under strong mixing conditions. Masry [27] obtained quadratic mean convergence and asymptotic normality under various mixing conditions. Masry [28] and Györfi and Masry [15] derived strong consistency and almost sure convergence rates of $\tilde{f}_n^*(x)$ under various mixing conditions; see Tran [49], [50] for improved results. Recall that $B_{n,l} = \sum_{i=1}^n b_i^l$, $l \in \mathbf{R}$. We shall here provide the central limit theorem of the recursive estimator (19) under our framework (4).

THEOREM **3**. *Assume condition* (**K**), $\sup_u |f'(u)| < \infty$, *and either one of the following:*

(i) *Condition* (**D1**), $B_{n,-3} = o(B_{n,-1}^2)$;

(ii) *Condition* (**D2**), $b_i \asymp i^{-\beta}\ell_0(i)$ *where* $2(1-\gamma) < \beta < 1$ *and* $\ell_0(\cdot)$ *is a slowly varying function.*

Then we have the central limit theorem

$$\frac{n[\tilde{f}_n^*(x_0) - \mathbb{E}\tilde{f}_n^*(x_0)]}{\sqrt{B_{n,-1}}} \Rightarrow N(0, \kappa f(x_0)).$$

Besides the central limit theorem, one can also have the LIL result as in Theorem 2 that concerns the almost sure limiting behavior of the estimator. The proofs are similar to those of Theorems 1 and 2, and we shall here omit the details.

C. Recursive Wavelet Density Estimation

Another popular approach is the wavelet density estimate. To introduce it, we shall use the language of multi-resolution analysis (Daubechies [5]). Let $\{V_j\}_{j\in\mathbf{Z}} \subset \mathcal{L}_2(\mathbf{R})$ be a collection of nested subspaces such that $\bigcap_{j\in\mathbf{Z}} V_j = \{\emptyset\}$, $\bigcup_{j\in\mathbf{Z}} V_j$ is dense in $\mathcal{L}_2(\mathbf{R})$, $V_{j+1} \subset V_j$, $f(2x) \in V_{j+1} \Leftrightarrow f(x) \in V_j$, V_j has orthonormal basis $\{\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k),\ k \in \mathbf{Z}\}$ for some function $\phi \in \mathcal{L}_2(\mathbf{R})$ with $\int_{\mathbf{R}} \phi^2(x) dx = 1$. Let the detail space W_j be given by $V_{j+1} = V_j \bigoplus W_j$, which has the orthonormal basis $\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), k \in \mathbf{Z}\}$ for some function ψ. Then $V_m \bigoplus W_m \bigoplus W_{m+1} \bigoplus W_{m+2} \dots$ is dense in $\mathcal{L}_2(\mathbf{R})$. Let

$$f(x) = \sum_{k\in\mathbf{Z}} c_{m,k}\phi_{m,k}(x) + \sum_{j\geq m}\sum_{k\in\mathbf{Z}} d_{j,k}\psi_{j,k}(x), \quad (20)$$

be the orthogonal representation of f, where m is the resolution level, the coefficients $c_{m,k} = \int_{\mathbf{R}} f(u)\phi_{m,k}(u)du$ and $d_{j,k} = \int_{\mathbf{R}} f(u)\psi_{j,k}(u)du$. Since $c_{m,k} = \mathbb{E}\phi_{m,k}(X_1)$, we can empirically estimate it by $\hat{c}_{m,k} = n^{-1}\sum_{t=1}^n \phi_{m,k}(X_t)$, and the wavelet-based estimator

$$\begin{aligned}\hat{f}_n^{\mathrm{wv}}(x) &= \sum_{k\in\mathbf{Z}} \hat{c}_{m,k}\phi_{m,k}(x) = \frac{1}{n}\sum_{k\in\mathbf{Z}}\phi_{m,k}(x)\sum_{t=1}^n \phi_{m,k}(X_t) \\ &= \frac{1}{n}\sum_{t=1}^n\sum_{k\in\mathbf{Z}} 2^{\frac{m}{2}}\phi(2^m x - k)2^{\frac{m}{2}}\phi(2^m X_t - k) \\ &= (nb_n)^{-1}\sum_{t=1}^n K\left(\frac{x}{b_n}, \frac{X_t}{b_n}\right),\end{aligned}$$

where $K(u,v) = \sum_{k\in\mathbf{Z}}\phi(u-k)\phi(v-k)$ has a bivariate representation with bandwidth $b_n = 2^{-m}$. Here the kernel $K(\cdot,\cdot)$ is bivariate. As in (2), we can naturally define the recursive wavelet density estimator

$$\tilde{f}_n^{\mathrm{wv}}(x) = \frac{\sum_{i=1}^n K(x/b_i, X_i/b_i)}{\sum_{i=1}^n b_i} \quad (21)$$

THEOREM **4**. *Assume that $K(\cdot,\cdot)$ is a symmetric bivariate kernel satisfying $\sup_{u,v}|K(u,v)| < \infty$, $\sup_u \int_{\mathbf{R}}|K(u,v)|dv < \infty$, $K(u,v)\to 0$ as $|u-v|\to\infty$, and for $\sigma^2(u) := \int_{\mathbf{R}} K^2(u,v)dv$, $\inf_u \sigma^2(u) > 0$. Let $B_n(x) = \sum_{i=1}^n b_i\sigma^2(x/b_i)$. Then under (i) or (ii) of Theorem 1, we have the central limit theorem*

$$\frac{B_n[\tilde{f}_n^{\mathrm{wv}}(x_0) - \mathbb{E}\tilde{f}_n^{\mathrm{wv}}(x_0)]}{\sqrt{B_n(x_0)}} \Rightarrow N(0,\, f(x_0)). \quad (22)$$

Under (i) or (ii) of Theorem 2, we have the law of the iterated logarithm

$$\overline{\underline{\lim}}_{n\to\infty}\frac{B_n[\tilde{f}_n^{\mathrm{wv}}(x_0) - \mathbb{E}\tilde{f}_n^{\mathrm{wv}}(x_0)]}{\sqrt{B_n(x_0)\log\log B_n(x_0)}} = \pm\sqrt{2f(x_0)}. \quad (23)$$

If the bias $b^{-1}\mathbb{E}K(x_0/b, X/b) - f(x_0) = \mathcal{O}(b^p)$ as $b\to 0$, $p\geq 1$, then (22) [resp. (23)] hold with $\mathbb{E}\tilde{f}_n^{\mathrm{wv}}(x_0)$ therein replaced by $f(x_0)$ if $B_{n,p+1} = o(B_n^{1/2})$ [resp. $B_{n,p+1} = \mathcal{O}(B_n^{1/2})$].

D. Bandwidth Selection

For the traditional kernel density estimate

$$\hat{f}_n(x) = \frac{1}{nb_n}\sum_{i=1}^n K_{b_n}(x, X_i), \quad (24)$$

we have the bias $\mathbb{E}\hat{f}_n(x) - f(x) \sim b_n^2\nu_2 f''(x)/2$, where $\nu_2 = \int u^2K(u)du$, and the variance $\mathrm{var}(\hat{f}_n(x)) \sim (nb_n)^{-1}f(x)\kappa$ (Silverman [43]). The optimal bandwidth b_n minimizing the asymptotic mean squared error (AMSE) has the form $b_n = \hat{c}n^{-1/5}$ with $\hat{c} = (f(x)\kappa)^{1/5}(\nu_2 f''(x))^{-2/5}$. For our recursive estimate $\tilde{f}_n(x)$, letting $b_i = ci^{-\beta}$ where $0<\beta<1$ and c is a constant, we have by elementary calculations that $\mathbb{E}\tilde{f}_n(x) - f(x) \sim \nu_2 f''(x)B_{n,3}/B_n \sim (1-\beta)/(2-6\beta)b_n^2\nu_2 f''(x)$ if $\beta < 1/3$, or $2\log(n)b_n^2\nu_2 f''(x)/3$ if $\beta = 1/3$, or $\mathcal{O}((nb_n)^{-1})$ if $\beta > 1/3$. And $\mathrm{var}(\tilde{f}_n(x)) \sim (1-\beta)(nb_n)^{-1}f(x)\kappa$. Let $\beta = 1/5$. Then the AMSE-optimal c satisfies

$$c = \hat{c}\left\{\frac{(1-3/5)^2}{(1-1/5)}\right\}^{1/5} = 5^{-1/5}\hat{c}. \quad (25)$$

Hence we can first conduct a pilot study and, based on part of the sample $X_1,\dots,X_N$ with a relatively large N, we can apply the classical bandwidth selection procedure (for example, Jones et al [21]) and obtain an estimate of $\hat{c}$ for the non-recursive estimate. Then we use $b_i = ci^{-\beta}$ with $c = 5^{-1/5}\hat{c}$ for our recursive estimates. With this choice, we have $\mathrm{AMSE}(\tilde{f}_n(x))/\mathrm{AMSE}(\hat{f}_n(x)) = (2\times 5^{-2/5})^2 \approx 1.1$, suggesting that the recursive estimator has a quite satisfactory performance. A simulation study is carried out in Section V.

IV. RECURSIVE KERNEL REGRESSION ESTIMATION

If we observe a bivariate stationary process (X_i, Y_i) according to model (5), we can consider the problem of estimating the conditional regression function $g(x) = \mathbb{E}(Y_i|X_i = x)$. Let $V_p(x) = \mathbb{E}(|Y_1|^p|X_1 = x)$, and $\sigma^2(x) = V_2(x) - g^2(x)$ be the conditional variance function. Recall that the conditional regression function can be recursively estimated by $\tilde{g}_n(x)$ defined in (2). We shall here provide the asymptotic properties of the recursive kernel regression estimator $\tilde{g}_n(\cdot)$. Section IV-A presents a central limit theorem for $\tilde{g}_n$. A LIL is given in Section IV-B and Section IV-C provides an optimal uniform almost sure convergence rate.

A. Asymptotic Normality

Recall (2) for $J_n(x)$. We shall first state a central limit theorem for $J_n(x) - \mathbb{E}J_n(x)$. Then the asymptotic normality of $\tilde{g}_n(x)$ follows by Slutsky's theorem. Theorem 1 is a special case of Theorem 5 with $Y_i \equiv 1$.

THEOREM **5**. *Let $p>2$. Assume that g and V_2 are Lipschitz continuous, V_p is bounded on a neighborhood of x_0. Under conditions of Theorem 1, we have*

$$\sqrt{B_n}(J_n(x_0) - \mathbb{E}J_n(x_0)) \Rightarrow N(0, V_2(x_0)f(x_0)\kappa). \quad (26)$$

Hence it implies that for $f(x_0)>0$, we have

$$\sqrt{B_n}\left(\tilde{g}_n(x_0) - \frac{\mathbb{E}J_n(x_0)}{\mathbb{E}\tilde{f}_n(x_0)}\right) \Rightarrow N(0, \sigma^2(x_0)\kappa/f(x_0)). \quad (27)$$

B. A Law of the Iterated Logarithm

THEOREM **6**. *Let $2<p\leq 4$. Assume condition* (**K**), *$\sup_u|f'(u)| < \infty$, g and V_2 are Lipschitz continuous, V_p is bounded on a neighborhood of x, $g_n(x) := \mathbb{E}J_n(x)/\mathbb{E}\tilde{f}_n(x) = g(x) + \mathcal{O}(B_n^{-1/2})$, and either*

(i) *Condition* (**D1**), *$(\log n)^{2p/(p-2)} + B_{n,2}\log n = \mathcal{O}(B_n)$ and $\sum_{k=1}^\infty B_{2^{1+k},2}/B_{2^k} < \infty$; or*
(ii) *Condition (ii) of Theorem 2.*

Then for any x with $f(x) > 0$, we have the LIL

$$\overline{\lim}_{n\to\infty} \frac{B_n^{1/2}(\tilde{g}_n(x) - g_n(x))}{\sqrt{\log\log B_n}} = \pm\frac{\sqrt{2\kappa}}{\sqrt{f(x)}}\sigma(x). \quad (28)$$

C. Uniform Almost Sure Convergence

In this section we shall study the uniform almost sure convergence for $\tilde{g}(\cdot)$ over compact intervals. To this end we need to introduce the uniform functional dependence measure

$$\delta_k^\diamond = \left\| \sup_{x\in\mathbf{R}} |f_1(x|\mathcal{F}_{k-1}) - f_1(x|\mathcal{F}_{k-1,\{0\}})| \right\|. \quad (29)$$

We shall modify conditions (**D1**) and (**D2**) accordingly with δ_k therein replaced by $\delta_k^\diamond$, and call the new dependence conditions as (**D1**$^\diamond$) and (**D2**$^\diamond$), respectively.

THEOREM **7**. *Assume* (**K**)*, $Y_i \in \mathcal{L}^p$, $p > 2$, $V_p(\cdot)$ is bounded on $[-a-\psi, a+\psi]$ for some $\psi > 0$. (i) Assume* (**D1**$^\diamond$) *and $\sum_{k=1}^\infty B_{2^{1+k},2}/(kB_{2^k}) < \infty$. Then*

$$\sup_{x\in[-a,a]} B_n|J_n(x) - \mathbb{E}J_n(x)| = \mathcal{O}_{a.s.}(\chi_n), \quad (30)$$

where $\chi_n = n^{1/p}\log n + \sqrt{B_n \log n}$.

(ii) Assume (**D2**$^\diamond$)*, $b_i \asymp i^{-\beta}\ell_0(i)$ where $\beta < 1$ and $\ell_0(\cdot)$ is a slowly varying function. Then there exists a slowly varying function $\ell_2(n)$ such that*

$$\sup_{x\in[-a,a]} B_n|J_n(x) - \mathbb{E}J_n(x)| = \mathcal{O}_{a.s.}(\eta_n), \quad (31)$$

where $\eta_n = \chi_n + n^{3/2-\beta-\gamma}\ell_2(n)$.

REMARK **1**. *In (ii) of Theorem 7, we do not impose the condition $2(1-\gamma) < \beta$, which is needed in previous theorems. Consequently, in (31) there is the term $n^{3/2-\beta-\gamma}\ell_2(n)$ which accounts for dependence. Note that if $3/2-\beta-\gamma < 1/p$ or $2(1-\gamma) < \beta$, then the second term in η_n is absorbed into the first term χ_n.* $\diamondsuit$

COROLLARY **1**. *Assume $\min_{x\in[-a,a]} |f(x)| > 0$ and $f, g \in \mathcal{C}^2[-a-\psi, a+\psi]$. Let ρ_n be the right hand side of (30) and (31). Then under conditions of Theorem 7, we have*

$$\max_{x\in[-a,a]} |\tilde{g}_n(x) - g(x)| = \mathcal{O}_{a.s.}(\rho_n/B_n + B_{n,3}/B_n). (32)$$

PROPOSITION **3**. *Let $X_i, e_i, i \in \mathbf{Z}$, be i.i.d. $N(0,1)$ and consider the regression model*

$$Y_i = g(X_i) + e_i, \quad (33)$$

where $g \in \mathcal{C}^1[-\psi, 1+\psi]$ for some $\psi > 0$. Let $b_i \asymp i^{-\beta}$, $0 < \beta < 1$. Assume (**K**)*. Let $g_n(x) = \mathbb{E}J_n(x)/\mathbb{E}\tilde{f}_n(x)$. Then for any $\alpha \in (0,1)$, there exists a constant $c > 0$ such that*

$$\limsup_{n\to\infty} P\left(\sup_{x\in[0,1]} |\tilde{g}_n(x) - g_n(x)| > c\sqrt{B_n^{-1}\log n}\right) \geq \alpha. \quad (34)$$

Proposition 3 suggests that the almost sure bounds in Theorem 7 and Corollary 1 are generally un-improvable, since under model (33), χ_n in Theorem 7 can be reduced to $\sqrt{B_n \log n}$, and the stochastic part of (32) $\max_{x\in[0,1]} |\tilde{g}_n(x) - g_n(x)| = \mathcal{O}_{a.s.}(\sqrt{B_n^{-1}\log n})$.

V. A SIMULATION STUDY

We shall here carry out a simulation study and examine the finite sample performance of the recursive kernel estimators. We consider two models:

(i) Nonlinear AR(1) process $X_i = \theta|X_{i-1}| + \varepsilon_i$, where ε_i are i.i.d. $N(0,1)$ and $|\theta| < 1$. Then the stationary distribution is skew-normal with density $2\phi(x(1-\theta^2)^{1/2})\Phi(\theta x)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions (cf. Anděl et al. [1] and Tong [48]);

(ii) Fractionally integrated process $(1-B)^d X_i = \varepsilon_i$ where ε_i are i.i.d. $N(0,\sigma^2)$ with $\sigma^2 = \Gamma^2(1-d)/\Gamma(1-2d)$, $0 < d < 1/2$. Then the marginal distribution of (X_i) is standard normal, and $X_i = \sum_{j=0}^\infty \psi_j\varepsilon_{i-j}$ with $\psi_j = \Gamma(j+d)/(\Gamma(j+1)\Gamma(d)) \sim j^{d-1}/\Gamma(d)$ (cf. Hosking [20]).

Let $\theta = 0.3$ in model (i) and $d = 0.09$ in model (ii), and $n = 10^3$. Using the bandwidth selector in Jones et al [21], we obtain $\hat{c} \approx 0.92$ for model (i), and $\hat{c} \approx 0.90$ for model (ii), respectively. By (25), for the recursive estimator we choose $c = 0.67$ and 0.65 for model (i) and (ii), respectively. Figure 1 presents the comparison of the empirical bias/variance with the asymptotic values averaged over 10000 replications for both recursive and traditional kernel density estimators given by (15) and (24), respectively. It shows that, for both models, the empirical bias and variance of the recursive kernel density estimator are quite close to those of the classical non-recursive version.

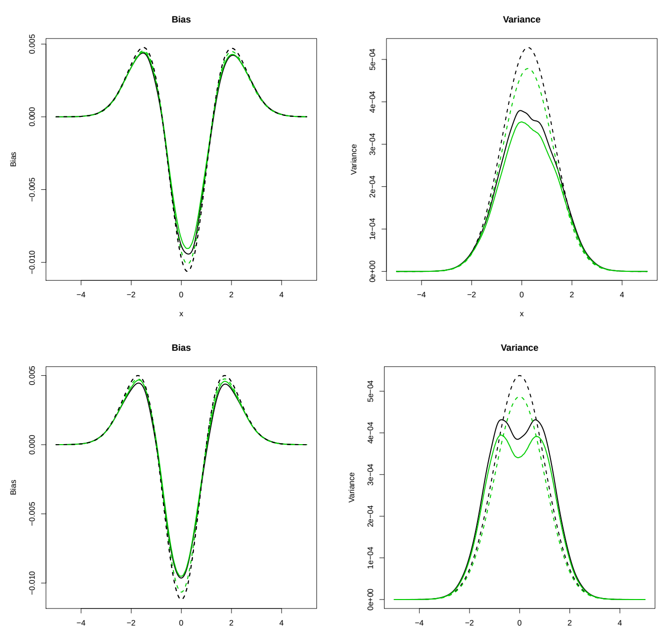


Fig. 1. Plots of the bias (left) and variance (right) of both recursive (black) and traditional (green) kernel density estimates with sample sizes $n = 1000$ from Model (i) (top) or Model (ii) (bottom). The form of density estimators are given by given by (15) and (24) for recursive and traditional ones. Each result is averaged over 10^4 replications. The solid lines are empirical estimates, while the dashed lines are average asymptotic values for either bias or variance discussed in Section III-D.

APPENDIX

Proof of Proposition 1. It follows from equation (36) in Wu et al. [61] by letting $p = 2$ and $q = \alpha$ therein. $\diamondsuit$

Proof of Proposition 2. By Theorem 2 in Wu and Shao [62], (11) implies that the Markov chain (6) has a stationary solution of form (4) such that the following geometric moment contraction property holds:

$$\|X_k - X_{k,\{0\}}\|_\tau = \mathcal{O}(L^k), \text{ where } X_{k,\{0\}} = R(\mathcal{F}_{k,\{0\}}). \tag{35}$$

Let $\delta_{k,p} = \sup_{x\in\mathbf{R}} \|f_1(x|\mathcal{F}_{k-1}) - f_1(x|\mathcal{F}_{k-1,\{0\}})\|_p$. Then by (12),

$$|f_1(x|\mathcal{F}_{k-1}) - f_1(x|\mathcal{F}_{k-1,\{0\}})| \le \min(2C_0,\, H|X_k - X_{k,\{0\}}|^\alpha). \tag{36}$$

Hence, if $\alpha p \le \tau$, by (35), we have $\delta_{k,p} = \mathcal{O}(\rho^k)$ with $\rho = L^\alpha$. If $\alpha p > \tau$, we have

$$\min(2C_0,\, H|X_k - X_{k,\{0\}}|^\alpha)^p \le (2C_0)^{p-\frac{\tau}{\alpha}} H^{\frac{\tau}{\alpha}} |X_k - X_{k,\{0\}}|^\tau, \tag{37}$$

which by (35) and (36) also implies $\delta_{k,p} = \mathcal{O}(\rho^k)$ with $\rho = L^{\tau/p}$. $\diamondsuit$

LEMMA **1**. *Assume that (X_i) is short-range dependent satisfying condition* (**D1**) *and*

$$\sum_{k=1}^\infty \frac{B_{2^{1+k},2}}{B_{2^k}} < \infty. \tag{38}$$

Let

$$R_n(x) := \sum_{i=1}^n \left(\mathbb{E}(K_{b_i}(x, X_i)|\mathcal{F}_{i-1}) - \mathbb{E}(K_{b_i}(x, X_i))\right). \tag{39}$$

Then $B_n^{-1/2} R_n(x) \to 0$ almost surely, namely $R_n(x) = o_{a.s.}(B_n^{1/2})$.

Proof of Lemma 1. Recall $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Define

$$\Lambda_n(x, v) := \sum_{i=1}^n b_i[f_1(x - b_i v|\mathcal{F}_{i-1}) - f(x - b_i v)]. \tag{40}$$

Notice that $\mathbb{E}(K_{b_i}(x, X_i))|\mathcal{F}_{i-1}) = b_i \int_{\mathbf{R}} K(u) f_1(x - b_i u|\mathcal{F}_{i-1}) du$, we have $R_n(x) = \int_{\mathbf{R}} K(u)\Lambda_n(x, u) du$. Since $f_1(x|\mathcal{F}_{i-1}) - f(x) = \sum_{l=1}^\infty \mathcal{P}_{i-l} f_1(x|\mathcal{F}_{i-1})$ and $\mathcal{P}_{i-l} f_1(x|\mathcal{F}_{i-1})$, $i = 1, 2, \ldots$, are martingale differences, by the triangle and Doob's inequalities, and the orthogonality of the projection operators $\mathcal{P}_l\cdot$,

$$\begin{aligned}
&\left\|\max_{n\le 2^k} |\Lambda_n(x, u)|\right\| \qquad (41)\\
\le&\sum_{l=1}^\infty \left\|\max_{n\le 2^k} \left|\sum_{i=1}^n b_i \mathcal{P}_{i-l} f_1(x - b_i v|\mathcal{F}_{i-1})\right|\right\|\\
\le&\sum_{l=1}^\infty 2\left\|\sum_{i=1}^{2^k} b_i \mathcal{P}_{i-l} f_1(x - b_i v|\mathcal{F}_{i-1})\right\|\\
=&\sum_{l=1}^\infty 2\left(\sum_{i=1}^{2^k} b_i^2 \|\mathcal{P}_{i-l} f_1(x - b_i v|\mathcal{F}_{i-1})\|^2\right)^{1/2}\\
\le&\sum_{l=1}^\infty 2\left(\sum_{i=1}^{2^k} b_i^2\theta_l^2\right)^{1/2} = 2\Theta_1 B_{2^k,2}^{1/2} \qquad (42)
\end{aligned}$$

holds uniformly over x and u. Let $W_k = \max_{n\le 2^k} |R_n(x)|^2/B_{2^{k-1}}$. By (38),

$$\sum_{k=2}^\infty \mathbb{E}W_k \le \sum_{k=2}^\infty \frac{\int_{\mathbf{R}} |K(u)| \left\|\max_{n\le 2^k} |\Lambda_n(x, u)|\right\|^2 du}{B_{2^{k-1}}} < \infty. \tag{43}$$

Hence $\sum_{k=2}^\infty W_k < \infty$ almost surely, implying $W_k = o_{a.s.}(1)$ and $R_n(x) = o_{a.s.}(B_n^{1/2})$. $\diamondsuit$

Lemma 1 concerns the short-range dependent case with $\Theta_1 < \infty$. Under long-range dependence, Θ_1 in (41) is infinite and the bound therein is no longer useful. In this case we need the following Lemma 2, where an upper bound of a different type is given.

LEMMA **2**. *We have*

$$\left\|\max_{k\le n} |\Lambda_k(x, u)|\right\|^2 \le \sum_{l\in\mathbf{Z}} \left(\sum_{i=1}^n b_i \delta_{i-l}\right)^2 =: \Xi_n. \tag{44}$$

Assume condition (**D2**) *and $b_i \asymp i^{-\beta}\ell_0(i)$ where $0 < \beta < 1$ and $\ell_0(\cdot)$ is another slowly varying function. If $\gamma + \beta/2 > 1$, then $R_n(x) = o_{a.s.}(B_n^{1/2})$ and $\Xi_n = \mathcal{O}(\Psi_n)$. Here $\Psi_n = \mathcal{O}(n^{3-2\beta-2\gamma}\ell^2(n)\ell_0^2(n))$ if $3/2 > \beta+\gamma$; $\Psi_n = \mathcal{O}(1)$ if $3/2 < \beta + \gamma$; and $\Psi_n = \mathcal{O}(\ell_1(n))$ if $3/2 = \beta + \gamma$, where $\ell_1(n) = \sum_{m=1}^n m^{-1}\ell_0^2(m)\ell^2(m)$.*

Proof of Lemma 2. Write $\Lambda_n = \Lambda_n(x, u)$. Let $\Lambda_{n,\{l\}}$ be a coupled version of Λ_n with ε_l in the latter replaced by the i.i.d. copy ε'_l. By Jensen's and the triangle inequalities,

$$\begin{aligned}
&\left\|\mathcal{P}_l \max_{k\le n} |\Lambda_k|\right\|\\
\le&\left\|\max_{k\le n} |\Lambda_k| - \max_{k\le n} |\Lambda_{k,\{l\}}|\right\|\\
\le&\sum_{i=1}^n |b_i| \|f_1(x - b_i v|\mathcal{F}_{i-1}) - f_1(x - b_i v|\mathcal{F}_{i-1,\{l\}})\|\\
\le&\sum_{i=1}^n |b_i|\delta_{i-l},
\end{aligned} \tag{45}$$

which implies (44) since the projection operators $\mathcal{P}_l\cdot$ yield martingale differences.

We now prove the second assertion. Let $b_i^* = i^{-\beta}\ell_0(i)$, $\delta_i^* = i^{-\gamma}\ell(i)$, $i \ge 1$, $b_0^* = \delta_0^* = 1$ and $b_i^* = \delta_i^* = 0$ if $i < 0$, and $\Xi_n^* = \sum_{l=-\infty}^n (\sum_{i=1}^n b_i^*\delta_{i-l}^*)^2$. By Karamata's Theorem, $\sum_{i=1}^n b_i^* \asymp n b_n^*$ and

$$\begin{aligned}
\sum_{l=-\infty}^{-n} \left(\sum_{i=1}^n b_i^*\delta_{i-l}^*\right)^2 =& \sum_{l=-\infty}^{-n} \mathcal{O}(\delta_{-l}^*)^2 \left(\sum_{i=1}^n b_i^*\right)^2\\
=&\mathcal{O}(n(\delta_n^*)^2(nb_n^*)^2).
\end{aligned} \tag{46}$$

Elementary calculations show that, since $\beta + \gamma > 1$, we have $\sum_{i=0}^\infty b_i^*\delta_{i+m}^* \asymp m b_m^*\delta_m^*$ as $m \to \infty$. Again by Karamata's Theorem, if $2(1 - \beta - \gamma) > -1$, or $3/2 > \beta + \gamma$, we obtain $\sum_{m=1}^n (m b_m^*\delta_m^*)^2 \asymp n^3(b_n^*\delta_n^*)^2$. If $3/2 < \beta + \gamma$, then $\sum_{m=1}^n (m b_m^*\delta_m^*)^2 < \sum_{m=1}^\infty (m b_m^*\delta_m^*)^2 < \infty$. At the borderline case with $3/2 = \beta + \gamma$, then $\sum_{m=1}^{n}(mb_m^*\delta_m^*)^2 = \ell_1(n)$, which is also a slowly varying function. Combining all these three cases, we have $\sum_{l=-n+1}^{1}(\sum_{i=0}^{n} b_i^*\delta_{i-l}^*)^2 \asymp \sum_{m=1}^{n}(\sum_{i=0}^{\infty} b_i^*\delta_{i+m}^*)^2 \asymp \sum_{m=1}^{n}(mb_m^*\delta_m^*)^2 = \mathcal{O}(\Psi_n)$. And $\sum_{l=0}^{n}(\sum_{i=1}^{n} b_i^*\delta_{i-l}^*)^2 \leq \sum_{l=0}^{n}(\sum_{m=1}^{n} b_{m+l}^*\delta_m^*)^2$ can be bounded in the same way. Notice that $\Xi_n = \left(\sum_{l=-\infty}^{-n} + \sum_{l=-n+1}^{1} + \sum_{l=0}^{n}\right)\left(\sum_{i=1}^{n} b_i^*\delta_{i-l}^*\right)^2$, the result follows and we have $\sum_{k=1}^{\infty} \Xi_{2^k}/B_{2^k} < \infty$ if $\gamma + \beta/2 > 1$, which, as in (43), implies $R_n(x) = o_{a.s.}(B_n^{1/2})$. ◇

Proof of Theorem 1. By Lemmas 1 and 2, it follows that $R_n = o_p(B_n^{1/2})$ since almost sure convergence implies convergence in probability, hence it suffices to show that

$$\frac{\sum_{i=1}^{n} D_i}{B_n^{1/2}} \Rightarrow N(0,\, 2\kappa f(x_0)), \tag{47}$$

where $D_i = K_{b_i}(x_0, X_i) - \mathbb{E}[K_{b_i}(x_0, X_i)|\mathcal{F}_{i-1}]$.

To this end, we shall apply the martingale central limit theorem; see Hall and Heyde [17]. Since $|D_i|$ is bounded by $2C_K$ and $B_n \to \infty$, for any $\epsilon > 0$, we have $D_i^2 \mathbf{1}_{|D_i| \geq B_n^{1/2}\epsilon} = 0$ for all large n. It remains to show that

$$B_n^{-1}\sum_{i=1}^{n}\mathbb{E}(D_i^2|\mathcal{F}_{i-1}) \to \kappa f(x_0) \text{ in probability.} \tag{48}$$

Since

$$\mathbb{E}(D_i^2|\mathcal{F}_{i-1}) = \mathbb{E}(K_{b_i}^2(x_0, X_i)|\mathcal{F}_{i-1}) - \mathbb{E}^2[K_{b_i}(x_0, X_i)|\mathcal{F}_{i-1}],$$

and elementary calculation readily shows that $\mathbb{E}(K_{b_i}^2(x_0, X_i)|\mathcal{F}_{i-1}) \to \kappa f(x_0) b_i$ as $i \to \infty$; and $\mathbb{E}(K_{b_i}(x_0, X_i)|\mathcal{F}_{i-1}) = \mathcal{O}(b_i)$. Notice that $B_n \to \infty$, it follows by Toeplitz lemma that,

$$\begin{aligned}
&B_n^{-1}\sum_{i=1}^{n}\mathbb{E}(D_i^2|\mathcal{F}_{i-1})\\
=&B_n^{-1}\sum_{i=1}^{n}[\mathbb{E}(K_{b_i}^2(x_0, X_i)|\mathcal{F}_{i-1}) - \mathbb{E}^2(K_{b_i}(x_0, X_i)|\mathcal{F}_{i-1})]\\
\to&B_n^{-1}\sum_{i=1}^{n}\left(\kappa f(x_0) b_i + \mathcal{O}(b_i^2)\right)\\
\to&\kappa f(x_0).
\end{aligned} \tag{49}$$

Hence (47), and then (16) follows. ◇

Proof of Theorem 2. Recall $R_n(x)$ in equation (39). We can write

$$B_n(\tilde{f}_n(x) - \mathbb{E}\tilde{f}_n(x)) = M_n(x) + R_n(x), \tag{50}$$

where

$$M_n(x) = \sum_{i=1}^{n}(K_{b_i}(x, X_i) - \mathbb{E}(K_{b_i}(x, X_i)|\mathcal{F}_{i-1})) \tag{51}$$

is a martingale relative to the filtration $\mathcal{F}_n$. Then the Theorem follows from Lemmas 1, 2 above and 3 below. ◇

LEMMA **3**. *Under conditions of Theorem 2, we have*

$$\overline{\underline{\lim}}_{n\to\infty}\frac{M_n(x_0)}{\sqrt{B_n \log\log B_n}} = \pm\sqrt{2\kappa f(x_0)} \tag{52}$$

Proof of Lemma 3. Write $M_0 = 0$, $M_n = M_n(x_0)$ and $D_n = M_n - M_{n-1}$. Then D_n are martingale differences. By Strassen's [46] martingale representation theorem (see also the Appendix in Hall and Heyde [17]), on a possibly richer probability space (here, for presentational simplicity, we take it to be the same as the original space), there exists a standard Brownian motion $\mathbb{B}$ and a sequence of nonnegative random variables $\tau_1, \tau_2, \cdots$ such that

$$\begin{aligned}
&\{M_k,\, k \geq 1\} =_{\mathcal{D}} \{\mathbb{B}(T_k),\, k \geq 1\}, \text{ where } T_k = \sum_{i=1}^{k}\tau_i,\\
&\mathbb{E}(\tau_k|\mathcal{F}_{k-1}) = \mathbb{E}(D_k^2|\mathcal{F}_{k-1}) \text{ a.s.,}
\end{aligned} \tag{53}$$

and for $C_q = 2(8/\pi^2)^{q-1}\Gamma(q+1)$,

$$\mathbb{E}(\tau_k^{q/2}|\mathcal{F}_{k-1}) \leq C_q \mathbb{E}(D_k^q|\mathcal{F}_{k-1}) \text{ a.s. for } q > 2. \tag{54}$$

Let $t_n = \mathbb{E}(T_n) = \sum_{i=1}^{n}\mathbb{E}(D_i^2)$. By (7) and condition (**K**), $|\mathbb{E}(K_{b_i}(x_0 - X_i))|\mathcal{F}_{i-1})| \leq Cb_i$, which by $\sup_u |f'(u)| < \infty$ implies $\mathbb{E}(D_i^2) = \kappa f(x_0) b_i(1 + \mathcal{O}(b_i))$. Hence

$$t_n = \kappa f(x_0) B_n + \mathcal{O}(B_{n,2}). \tag{55}$$

We shall now apply (53) and (54) to prove (52). Write

$$\begin{aligned}
T_n - t_n =& \sum_{i=1}^{n}(\tau_i - \mathbb{E}(\tau_i|\mathcal{F}_{i-1})) + \sum_{i=1}^{n}(\mathbb{E}(\tau_i|\mathcal{F}_{i-1}) - \mathbb{E}(\tau_i))\\
:=& P_n + Q_n.
\end{aligned} \tag{56}$$

Note that P_n is also a martingale. By Doob's inequality and (54),

$$\begin{aligned}
\left\|\max_{i\leq n}|P_i|\right\|^2 \leq& 4\|P_n\|^2 \leq 4\sum_{i=1}^{n}\mathbb{E}(\tau_i^2)\\
\leq& 4C_4\sum_{i=1}^{n}\mathbb{E}(D_i^4) = \mathcal{O}(B_n).
\end{aligned} \tag{57}$$

As in (43), by the Borel-Cantelli Lemma, $\max_{i\leq 2^k}|P_i|^2 = o_{a.s.}(B_{2^k}k^2)$, implying the almost sure bound $\max_{i\leq n}|P_i| = o_{a.s.}(B_n^{1/2}\log n)$. For Q_n, recall (40) for $\Lambda_n(x, v)$, we have

$$\begin{aligned}
Q_n =& \sum_{i=1}^{n}\left(\mathbb{E}(D_i^2|\mathcal{F}_{i-1}) - \mathbb{E}(D_i^2)\right)\\
=& \sum_{i=1}^{n}\left(\mathbb{E}(K_{b_i}^2(x_0, X_i)|\mathcal{F}_{i-1}) - \mathbb{E}(K_{b_i}^2(x_0, X_i)) + \mathcal{O}(b_i^2)\right)\\
=& \int_{\mathbf{R}} K^2(v)\Lambda_n(x_0, v)dv + \mathcal{O}(B_{n,2}).
\end{aligned}$$

Hence, under (i) (resp. (ii)), we have by Lemma 1 (resp. Lemma 2) that

$$T_n - t_n = O_{a.s.}(\Delta_n), \text{ where } \Delta_n = B_n^{1/2}\log n + B_{n,2}.$$

By Lévy's modulus of continuity for Brownian motions, we have almost surely that

$$\begin{aligned}
\max_{k\leq n}|\mathbb{B}(T_k) - \mathbb{B}(t_k)| \leq& \sup_{|x-y|\leq C\Delta_n,\, 0\leq y\leq t_n}|\mathbb{B}(x) - \mathbb{B}(y)|\\
\leq& C(\Delta_n \log(B_n/\Delta_n))^{1/2}.
\end{aligned} \tag{58}$$

Under conditions on (b_i) in (i) or (ii), we have

$$\lim_{n\to\infty}\frac{\Delta_n \log(B_n/\Delta_n)}{B_n \log\log B_n} = 0. \tag{59}$$

Since $t_n \to \infty$, by the law of the iterated logarithm for Brownian motions, we have

$$\overline{\lim}_{n\to\infty} \frac{\mathbb{B}(t_n)}{\sqrt{2t_n \log\log t_n}} = \pm 1. \tag{60}$$

Hence by (53), (58) and (59), we have (52). $\diamondsuit$

Proof of Theorem 4. To simplify notation we let $x = x_0$. Let $D_i = K(x/b_i, X_i/b_i) - \mathbb{E}(K(x/b_i, X_i/b_i)|\mathcal{F}_{i-1})$ and $M_n = \sum_{i=1}^n D_i$. Similarly as (50), we write

$$B_n(\tilde{f}_n^{\rm wv}(x) - \mathbb{E}\tilde{f}_n^{\rm wv}(x)) = M_n + R_n, \tag{61}$$

where, noting that

$$\mathbb{E}(K(x/b_i, X_i/b_i)|\mathcal{F}_{i-1}) = b_i \int_{\mathbf{R}} K(x/b_i, u) f_1(b_i u|\mathcal{F}_{i-1}) du,$$

the term

$$R_n = \sum_{k=1}^{\infty} R_{n,k}, \tag{62}$$

where $R_{n,k} = \sum_{i=1}^n b_i \int_{\mathbf{R}} K(x/b_i, u)\mathcal{P}_{i-k} f_1(b_i u|\mathcal{F}_{i-1}) du$. By the Cauchy-Schwarz inequality and conditions on $K(\cdot,\cdot)$, we have

$$\begin{aligned}\|R_{n,k}\|^2 =& \sum_{i=1}^n b_i^2 \mathbb{E}\left(\int_{\mathbf{R}} K(x/b_i, u)\mathcal{P}_{i-k} f_1(b_i u|\mathcal{F}_{i-1}) du\right)^2 \\ \le& \sum_{i=1}^n b_i^2 \int_{\mathbf{R}} |K(x/b_i, u)| du \\ &\times \mathbb{E}\int_{\mathbf{R}} |K(x/b_i, u)||\mathcal{P}_{i-k} f_1(b_i u|\mathcal{F}_{i-1})|^2 du \\ =& \sum_{i=1}^n b_i^2 \mathcal{O}(\theta_k^2). \end{aligned} \tag{63}$$

Hence under the short-range dependent case (**D1**), $\|R_n\| = \mathcal{O}(\Theta_1) B_{n,2}^{1/2}$. Using a similar argument as the one in Lemma 2, we can deal with the long-range dependent case (**D2**). By the continuity of f, we have

$$\begin{aligned}&\mathbb{E}K^2(x/b, X_i/b) \\ =& \int_{\mathbf{R}} K^2(x/b, u) f(bu) du \\ =& \left(\int_{|ub-x|\ge b^{1/2}} + \int_{|ub-x|<b^{1/2}}\right) K^2(x/b, u) f(bu) du \\ =& \sigma^2(x/b) f(x) + o(1) \end{aligned} \tag{64}$$

as $b \to 0$, due to the conditions on the kernel $K(\cdot,\cdot)$. Then the martingale part M_n can be similarly dealt with by using the argument in Lemma 3. $\diamondsuit$

Proof of Theorem 5. To simplify notation we let $x = x_0$. Let $\mathcal{G}_i = (\cdots, e_{i-1}, e_i, \mathcal{F}_i)$, $\xi_i = Y_i K_{b_i}(x, X_i)$ and

$$\Lambda_n^\circ(x, v) := \sum_{i=1}^n g(x - b_i v) b_i \left(f_1(x - b_i v|\mathcal{F}_{i-1}) - f(x - b_i v)\right). \tag{65}$$

Note that ξ_i is $\mathcal{G}_i$-measurable and $\mathbb{E}(Y_i|X_i = x) = g(x)$. As in (39), we can write

$$L_n := \sum_{i=1}^n (\mathbb{E}(\xi_i|\mathcal{G}_{i-1}) - \mathbb{E}\xi_i) = \int K(v)\Lambda_n^\circ(x, v) dv. \tag{66}$$

By (**K**), since g is bounded at a neighbor of $x = x_0$, by the argument in Lemmas 1 and 2, we have $\|L_n\| = o(\sqrt{B_n})$. Let $d_i = \xi_i - \mathbb{E}(\xi_i|\mathcal{G}_{i-1})$. Then (26) follows from

$$\frac{1}{\sqrt{B_n}} \sum_{i=1}^n d_i \Rightarrow N[0, V_2(x) f(x)\kappa]. \tag{67}$$

To this end, we shall apply the martingale central limit theorem. Note that $|\mathbb{E}(\xi_i|\mathcal{G}_{i-1})| = \mathcal{O}(b_i)$. Hence $\|\mathbb{E}(d_i^2|\mathcal{G}_{i-1}) - \mathbb{E}(\xi_i^2|\mathcal{G}_{i-1})\| = \mathcal{O}(b_i^2)$. By Lemmas 1 and 2, we have

$$\begin{aligned}&\frac{1}{B_n}\sum_{i=1}^n \left(\mathbb{E}(\xi_i^2|\mathcal{G}_{i-1}) - \mathbb{E}(\xi_i^2)\right) \\ =& \frac{1}{B_n}\sum_{i=1}^n \int V_2(v) K^2\left(\frac{x-v}{b_i}\right)(f_1(v|\mathcal{G}_{i-1}) - f(v))\, \mathrm{d}v \\ =& \mathcal{O}({B_n}^{-1}) \int K^2(u)\Lambda_n(x, u) \mathrm{d}u = o_P(1)\end{aligned}$$

since $V_2(\cdot)$ is Lipschitz continuous. Note that K has bounded support,

$$\begin{aligned}&\frac{1}{B_n}\sum_{i=1}^n \mathbb{E}(d_i^2|\mathcal{G}_{i-1}) = \frac{1}{B_n}\sum_{i=1}^n \mathbb{E}(\xi_i^2) + o_P(1) \\ =& \frac{1}{B_n}\sum_{i=1}^n \int V_2(x - b_i v) K^2(v) f(x - b_i v) b_i \mathrm{d}v + o_P(1) \\ =& V_2(x) f(x) \int K^2(v)\mathrm{d}v + o_P(1).\end{aligned}$$

The proof is now complete since $\sum_{i=1}^n \|B_n^{-1/2} d_i\|_p^p \le 2^p B_n^{-p/2} \sum_{i=1}^n \|\xi_i\|_p^p = \mathcal{O}(B_n^{1-p/2}) \to 0$, Lindeberg's condition is satisfied. By Slutsky's Lemma, we have (27). $\diamondsuit$

Proof of Theorem 6. As in the proof of Theorem 5, let $\mathcal{G}_i = (\cdots, e_{i-1}, e_i, \mathcal{F}_i)$, $\zeta_i = (Y_i - g(x)) K_{b_i}(x, X_i)$, $D_i^\circ = \zeta_i - \mathbb{E}(\zeta_i|\mathcal{G}_{i-1})$ and $M_n^\circ = \sum_{i=1}^n D_i^\circ$. We shall show that

$$\overline{\lim}_{n\to\infty} \frac{\sum_{i=1}^n D_i^\circ}{\sqrt{B_n \log\log B_n}} = \pm\sqrt{2\kappa f(x)}\sigma(x) \tag{68}$$

almost surely. To this end, we shall use the arguments in the proof of Lemma 3. There exists a standard Brownian motion $\mathbb{B}^\circ$ and a sequence of nonnegative random variables $\tau_1^\circ, \tau_2^\circ, \cdots$, with $T_k^\circ = \sum_{i=1}^k \tau_i^\circ$ such that $\{M_k^\circ\}_{k\ge1} =_{\mathcal{D}} \{\mathbb{B}(T_k^\circ)\}_{k\ge1}$, and

$$\begin{aligned}\mathbb{E}(\tau_k^\circ|\mathcal{G}_{k-1}) = \mathbb{E}((D_k^\circ)^2|\mathcal{G}_{k-1}) \text{ a.s.} \\ \mathbb{E}((\tau_k^\circ)^{p/2}|\mathcal{G}_{k-1}) \le C_p \mathbb{E}((D_k^\circ)^p|\mathcal{G}_{k-1}) \text{ a.s.}\end{aligned} \tag{69}$$

where $C_p = 2(8/\pi^2)^{p-1}\Gamma(p+1)$. Since g is Lipschitz continuous and $\mathbb{E}(Y_i - g(x)|\mathcal{F}_i) = g(X_i) - g(x)$, by (7), we have

$$\begin{aligned}|\mathbb{E}(\zeta_i|\mathcal{G}_{i-1})| =& |\mathbb{E}((g(X_i) - g(x)) K_{b_i}(x, X_i)|\mathcal{F}_{i-1})| \\ \le& \int |g(u) - g(x)| K\left(\frac{x-u}{b_i}\right) f_1(u|\mathcal{F}_{i-1}) du \\ \le& b_i C \int |g(x - b_i v) - g(x)| K(v) dv \\ \le& b_i C \int C|b_i v| K(v) dv = \mathcal{O}(b_i^2)\end{aligned} \tag{70}$$

since K has bounded support. Similarly, $\mathbb{E}\{[g(X_i) - g(x)]^2 K_{b_i}^2(x, X_i)\} = \mathcal{O}(b_i^3)$, and since $\sigma^2(\cdot)$ is also Lipschitz continuous, $\mathbb{E}\{[\sigma^2(X_i)-\sigma^2(x)]K_{b_i}^2(x,X_i)\}=O(b_i^2)$. By the condition $\sup_u|f'(u)|<\infty$, we have $\mathbb{E}K_{b_i}^2(x,X_i)=b_i\kappa f(x)+O(b_i^2)$. Note that $\mathbb{E}((Y_i-g(x))^2|\mathcal{F}_i)=\sigma^2(X_i)+(g(X_i)-g(x))^2$. Then

$$\begin{aligned}\mathbb{E}T_k^\circ=&\mathbb{E}\sum_{i=1}^k\tau_i^\circ=\mathbb{E}\sum_{i=1}^k(\zeta_i-\mathbb{E}(\zeta_i|\mathcal{G}_{i-1}))^2\\ =&\sum_{i=1}^k(\mathbb{E}\zeta_i^2+\mathcal{O}(b_i^4))\\ =&\sum_{i=1}^k\mathbb{E}\left(\{\sigma^2(X_i)+[g(X_i)-g(x)]^2\}K_{b_i}^2(x,X_i)\right)+\mathcal{O}(B_{k,4})\\ =&\sum_{i=1}^k(\sigma^2(x)b_i\kappa f(x)+\mathcal{O}(b_i^2))+\mathcal{O}(B_{k,4})\\ =&\sigma^2(x)B_k\kappa f(x)+\mathcal{O}(B_{k,2}).\end{aligned}\tag{71}$$

By (69) and the Burkholder inequality, note that $\|\zeta_i\|_p^p=\mathcal{O}(b_i)$,

$$\begin{aligned}&\|\max_{k\le n}|\sum_{i=1}^k[\tau_i^\circ-\mathbb{E}(\tau_i^\circ|\mathcal{G}_{i-1})]|\|_{p/2}^{p/2}\\ \le&c_p\|\sum_{i=1}^n[\tau_i^\circ-\mathbb{E}(\tau_i^\circ|\mathcal{G}_{i-1})]\|_{p/2}^{p/2}\\ \le&c_p\sum_{i=1}^n\|\tau_i^\circ-\mathbb{E}(\tau_i^\circ|\mathcal{G}_{i-1})\|_{p/2}^{p/2}\le c_p\sum_{i=1}^n\|\tau_i^\circ\|_{p/2}^{p/2}\\ \le&c_p\sum_{i=1}^n\|D_i^\circ\|_p^p\le c_p\sum_{i=1}^n\|\zeta_i\|_p^p=\mathcal{O}(B_n).\end{aligned}\tag{72}$$

Here the constant c_p may change from place to place and it only depends on p. As in the proof of Lemma 3, under either (i) or (ii), we have (68) in view of

$$\begin{aligned}&T_n^\circ-\sigma^2(x)B_n\kappa f(x)\\ =&O_{\text{a.s.}}(B_n^{2/p}\log n)+O_{\text{a.s.}}(B_n^{1/2})+\mathcal{O}(B_{n,2}).\end{aligned}\tag{73}$$

By the representation (70) and the arguments in Lemmas 1 and 2, we have $\sum_{i=1}^n(\mathbb{E}(\zeta_i|\mathcal{G}_{i-1})-\mathbb{E}\zeta_i)=o_{\text{a.s.}}(B_n^{1/2})$. So we have a LIL for $\sum_{i=1}^n(\zeta_i-\mathbb{E}\zeta_i)$ in the form of (68) with D_i° therein replaced by $\zeta_i-\mathbb{E}\zeta_i$. With the latter, we have (28) by Slutsky's theorem. $\diamondsuit$

Proof of Theorem 7. Let $\bar{n}=2^{\lfloor(\log n)/(\log 2)\rfloor}$, $Y_i'=Y_i\mathbf{1}_{|Y_i|\le\bar{i}^{1/p}}$, $\xi_i'(x)=Y_i'K_{b_i}(x,X_i)$, $Y_i''=Y_i-Y_i'$, $\xi_i''(x)=Y_i''K_{b_i}(x,X_i)$ and $R_d=\sum_{i=1}^{2^d}[|Y_i''|+\mathbb{E}(|Y_i''||\mathcal{G}_{i-1})]$. Elementary calculations show that there exists a constant C such that

$$\sum_{d=1}^\infty\frac{\mathbb{E}R_d}{2^{d/p}}\le C\sum_{d=1}^\infty 2^{d-d/p}\mathbb{E}[|Y_1|\mathbf{1}_{|Y_1|\ge 2^{d/p}}]\le C\mathbb{E}[|Y_1|^p].$$

By the Borel-Cantelli lemma, $R_d=o_{\text{a.s.}}(2^{d/p})$ as $d\to\infty$. Since K is bounded, we have

$$\sup_x\left|\sum_{i=1}^n[\xi_i''(x)-\mathbb{E}(\xi_i''(x)|\mathcal{G}_{i-1})]\right|\le C_KR_{d_n}=o_{\text{a.s.}}(n^{1/p}),\tag{74}$$

where $d_n=\lceil(\log n)/(\log 2)\rceil$. We shall now deal with

$$W_n(x)=\sum_{i=1}^n[\xi_i'(x)-\mathbb{E}(\xi_i'(x)|\mathcal{G}_{i-1})].\tag{75}$$

Let $N=n^5$ and $\lfloor u\rfloor_N=\lfloor uN\rfloor/N$. Since K is Lipschitz continuous,

$$\begin{aligned}&\sup_x\left|\sum_{i=1}^n(\xi_i'(x)-\xi_i'(\lfloor x\rfloor_N))\right|\\ \le&\sum_{i=1}^n\sup_x\left|Y_i\mathbf{1}_{|Y_i|\le\bar{i}^{1/p}}(K_{b_i}(x,X_i)-K_{b_i}(\lfloor x\rfloor_N,X_i))\right|\\ \le&\sum_{i=1}^n\bar{i}^{1/p}C\frac{1}{b_i}\frac{1}{N}\le Cn^{1/p}B_{n,-1}/N.\end{aligned}$$

Notice that the same bound holds for $\sup_x|\sum_{i=1}^n(\mathbb{E}(\xi_i'(x)|\mathcal{G}_{i-1})-\mathbb{E}(\xi_i'(\lfloor x\rfloor_N)|\mathcal{G}_{i-1}))|$. Therefore

$$\sup_{x\in\mathbf{R}}|W_n(x)-W_n(\lfloor x\rfloor_N)|\le 2Cn^{1/p}B_{n,-1}/N.\tag{76}$$

By (7), since $V_p(\cdot)$ is bounded on $[-a-\psi,a+\psi]$, there is a constant C_0 such that

$$\begin{aligned}\mathbb{E}[\xi_i'(x)^2|\mathcal{G}_{i-1}]\le\mathbb{E}[\xi_i(x)^2|\mathcal{G}_{i-1}]\le&\mathbb{E}[V_2(X_i)K_{b_i}^2(x,X_i)|\mathcal{F}_{i-1}]\\ \le&b_iC_0.\end{aligned}\tag{77}$$

By Freedman's [10] inequality and (77), we have

$$\begin{aligned}&P\left[\max_{|x|\le a}|W_n(\lfloor x\rfloor_N)|\ge\lambda\chi_n\right]\\ \le&4aN\exp\left[-\frac{\lambda^2\chi_n^2}{2\lambda\chi_nC_Kn^{1/p}+2B_nC_0}\right],\end{aligned}\tag{78}$$

which is of order $\mathcal{O}(n^{-2})$ by choosing a sufficiently large λ. Therefore, by (76) and since $B_{n,-1}n^{1/p}/N=o(\chi_n)$, we have by the Borel-Cantelli lemma that

$$\max_{|x|\le a}|W_n(x)|=O_{a.s.}(\chi_n).\tag{79}$$

Recall (65) for $\Lambda_n^\circ(x,v)$ and (66) for $L_n(x)$. Similarly as Lemmas 1 or 2, under (i) or (ii), we have $\|\sup_x|\Lambda_n^\circ(x,v)|\|^2=\mathcal{O}(B_{n,2})$ or $\mathcal{O}(\Xi_n)$, respectively. Hence, by (79) and (74), we have (30) and (31), respectively. $\diamondsuit$

Proof of Corollary 1. Under the differentiability conditions on f,g, elementary calculations show that $g_n(x):=\mathbb{E}J_n(x)/\mathbb{E}\tilde{f}_n(x)=g(x)+\mathcal{O}(B_{n,3}/B_n)$. Write

$$\tilde{f}_n(x)[\tilde{g}_n(x)-g_n(x)]=H_n(x)+Q_n(x),\tag{80}$$

where $H_n(x)=J_n(x)-\mathbb{E}J_n(x)-(\tilde{f}_n(x)-\mathbb{E}\tilde{f}_n(x))g(x)$ and $Q_n(x)=(\tilde{f}_n(x)-\mathbb{E}\tilde{f}_n(x))[g(x)-g_n(x)]$, and apply Theorem 7 to $H_n(x)$ and $Q_n(x)$, we have (32). $\diamondsuit$

Proof of Proposition 3. We shall show that (34) holds with $\sup_{x\in[0,1]}$ therein replaced by $\max_{x_1,\ldots,x_N}$, where $x_i=in^{-\chi}$, $i=1,\ldots,N=\lfloor n^\chi\rfloor$, $0<\chi<\beta$. Let Z be a $N(0,1)$ random variable. For $\varrho\in(0,1)$, let $a_{N,\varrho}$ be such that $P(|Z|\ge a_{N,\varrho})=\varrho/N$. Let

$$L_n(x)=\frac{1}{\sqrt{f(x)\kappa}}\sum_{i=1}^n e_i\mathbf{1}_{|e_i|\le\log n}K_{b_i}(x,X_i).\tag{81}$$

Define event $A_i=\{|L_n(x_i)|\ge a_{N,\varrho}B_n^{1/2}\}$. Let $S_n=(L_n(x),L_n(x'))^\top$, where $|x-x'|\ge n^{-\chi}$. Note that $\tau_n:=\mathbb{E}(e_i^2\mathbf{1}_{|e_i|\le\log n})=1+\mathcal{O}(n^{-A})$ for any $A>2$. Under (**K**), we obtain $\mathbb{E}(L_n^2(x))=\tau_nf(x)\kappa(B_n+\mathcal{O}(B_{n,3}))$ and, for some constant $C>0$,

$$|\mathbb{E}[L_n(x)L_n(x')]|$$

$$
\begin{aligned}
=&\frac{\tau_n}{\sqrt{f(x)f(x')}\kappa}\left|\sum_{i=1}^{n}\mathbb{E}(K_{b_i}(x,X_i)K_{b_i}(x',X_i))\right| \\
\leq&\frac{\tau_n}{\sqrt{f(x)f(x')}\kappa}\sum_{i=1}^{n}\mathcal{O}(b_i\mathbf{1}_{|x-x'|\leq Mb_i})=\mathcal{O}(n^{\chi/\beta-\chi}) \quad (82)
\end{aligned}
$$

Then we can apply Lemma 4 with $c_n = n^{(\beta-1)/2}\log n$ and $R_n = B_n$. Elementary calculations show that $\gamma_n \asymp n^{(\beta-1)/2}$ and $a_{N,\varrho} = \sqrt{2\log N} + o(1) = o(\min(\gamma_n^{-1/3}, c_n^{-1}))$. Hence by (87) of Lemma 4, we obtain $P(A_i) = (1+o(1))\varrho/N$ and $P(A_i \cap A_j) = (1+o(1))\varrho^2/N^2$ uniformly in i and j. By the inclusion-exclusion formula, we have

$$
\begin{aligned}
&P\left(\cup_{1\leq i\leq N}A_i\right) \geq \sum_{i=1}^{N}P(A_i) - \sum_{1\leq i<j\leq N}P(A_i\cap A_j) \\
\to&\varrho - \frac{\varrho^2}{2}. \quad (83)
\end{aligned}
$$

Note that, for any $A > 2$, $P(|e_i| \geq \log n) = \mathcal{O}(n^{-A})$. Then we have

$$
\begin{aligned}
&\limsup_{n\to\infty} P\left(\sup_{x\in[0,1]}\frac{|\sum_{i=1}^{n}e_iK_{b_i}(x,X_i)|}{\sqrt{f(x)\kappa}} \geq a_{N,\varrho}B_n^{1/2}\right) \\
\geq&\varrho - \frac{\varrho^2}{2}. \quad (84)
\end{aligned}
$$

As in the proof of Corollary 1, we shall now deal with $Q_n(x) = \sum_{i=1}^{n}(g(X_i) - g(x))K_{b_i}(x,X_i)$. By Freedman's [10] inequality, for some constants $C_1, C_2 > 0$,

$$
P(|Q_n(x) - \mathbb{E}Q_n(x)| \geq \theta) \leq 2\exp\left(\frac{-\theta^2}{C_1\theta + C_2B_{n,3}}\right). \quad (85)
$$

Let $\Upsilon_n := \log n + (B_{n,3}\log n)^{1/2}$. Using the argument in (78), for any $A > 2$, there exists $\varphi > 0$ such that

$$
P\left(\sup_{x\in[0,1]}|Q_n(x) - \mathbb{E}Q_n(x)| \geq \varphi\Upsilon_n\right) = \mathcal{O}(n^{-A}). \quad (86)
$$

Then we have (34) in view of (86), (84) and the argument in (80). $\diamondsuit$

For completeness, we include the modified version of Lemma 4.2 in Lin and Liu [26] here. For a vector $z \in \mathbf{R}^d$, define $|z|_\star = \min\{|z_i| : 1 \leq i \leq d\}$.

LEMMA **4**. *Let $\xi_{n,1}, \cdots, \xi_{n,k_n}$ be independent mean zero $\mathbf{R}^d$ random vectors, and $S_n = \sum_{i=1}^{k_n}\xi_{n,i}$. Assume that $|\xi_{n,i}| \leq c_nR_n^{1/2}, 1 \leq i \leq k_n$ for some $c_n \to 0, R_n \to \infty$ and $|R_n^{-1}\mathrm{cov}(S_n) - I_d| \leq C_0c_n^2$, where I_d is a $d\times d$ identity matrix and C_0 is a positive constant. Suppose that $\gamma_n := R_n^{-3/2}\sum_{i=1}^{k_n}E|\xi_{n,i}|^3 \to 0$. Then we have*

$$
\begin{aligned}
&|P(|S_n|_\star \geq \lambda) - P(|N|_\star \geq \lambda R_n^{-1/2})| \\
&\qquad = o(P(|N|_\star \geq \lambda R_n^{-1/2})) \quad (87)
\end{aligned}
$$

uniformly over $\lambda \in [\sqrt{R_n}, \delta_n\min(c_n^{-1}, \gamma_n^{-1/3})\sqrt{R_n}]$ with any $\delta_n \to 0$ with $\delta_n\min(c_n^{-1}, \gamma_n^{-1/3}) \to \infty$. Here N is a centered normal random vector with covariance matrix I_d.

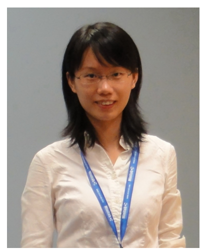
REFERENCES

[1] J. Andel, I. Netuka and K. Svara, "On threshold autoregressive processes", *Kybernetika*, vol. 20, pp. 89–106, 1984.
[2] I. Ahmad and P. Lin, "Nonparametric sequential estimation of a multiple regression function", *Bull. Math. Statist.*, vol. 17, 63–75, 1976.
[3] A. Amiri, "Recursive regression estimators with application to nonparametric prediction", *J. Nonparam. Statist.*, vol. 24, no. 1, pp. 169–186, 2012.
[4] X. Chen and H. White, "Asymptotic Properties of Some Projection-based Robbins-Monro Procedures in a Hilbert Space", *Studies in Nonlinear Dynamics & Econometrics*, vol. 6, no. 1, pp. 1–55, 2002.
[5] I. Daubechies, *Ten lectures on wavelets*, Number 61, CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia, 1992.
[6] P. Deheuvels, "Sur l'estimation sequentielle de la densite", *C. R. Acad. Sci. Paris Ser. A*, vol. 276, pp. 1119–1121, 1973.
[7] L. Devroye and T. J. Wagner, "On the L_1 convergence of kernel estimators of regression functions with applications in discrimination", *Z. Wahrsch. Verw. Gebiete.*, vol. 51, pp. 15–25, 1980.
[8] M. Duflo, *Algorithmes Stochastiques*, volume 23 of *Mathématiques & Applications*, Berlin: Springer-Verlag, 1996.
[9] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag, 2003.
[10] D. Freedman, "On tail probabilities for martingales", *Ann. Probab.*, vol. 3, pp. 100–118, 1975.
[11] W. Greblicki and M. Pawlak, "Necessary and sufficient pointwise consistency conditions for recursive kernel regression estimate", *J. Multivariate Analysis*, vol. 23, pp. 67–76, 1987.
[12] H. Gu and J. Lafferty, "Sequential Nonparametric Regression", *Proceddings of the 29th International Conference on Machine Learning*, 2012.
[13] L. Györfi, M. Kohler, A. Krzyżak and B. Walk, *A Distribution-Free Theory of Nonparametric Regression*, ser. Statistics, New York: Springer, 2002.
[14] L. Györfi, M. Kohler and H. Walk, "Weak and strong universal consistency of semi-recursive kernel and partitioning regression estimates", *Statist. Decisions*, vol. 16, pp. 1–18, 1998.
[15] L. Györfi and E. Masry, "The L_1 and L_2 strong consistency of recursive kernel density estimation from dependent samples", *IEEE Trans. Inform. Theory*, vol. 36, pp. 531–539, 1990.
[16] P. Hall, "Laws of the iterated logarithm for nonparametric density estimators", *Z. Wahrsch. Verw. Gebiete.*, vol. 56, pp. 47–61, 1981.
[17] P. Hall and C. Heyde, *Martingale Limit Theorem and its Application*, New York: Academic Press, 1980.
[18] P. Hall and P. Patil, "On the efficiency of on-line density estimators", *IEEE Trans. Inform. Theory*, vol. 40, pp. 1504–1512, 1994.
[19] W. Härdle, "A law of the iterated logarithm for nonparametric regression function estimators", *Ann. Statist.*, vol. 12, pp. 624–635, 1984.
[20] J. R. Hosking, "Fractional differencing", *Biometrika*, vol. 68, pp. 165–176, 1981.
[21] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation", *J. Amer. Statist. Assoc.*, vol. 91, pp. 401–407, 1996.
[22] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function", *Ann. Math. Statist.* vol. 23, pp. 462–466, 1952.
[23] A. Krzyżak, "Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation", *IEEE Trans. Inform. Theory*, vol. 38, pp. 1323–1338, 1992.
[24] A. Krzyżak and M. Pawlak, "Almost everywhere convergence of recursive regression function estimate and classification", *IEEE Trans. Inform. Theory*, vol. 30, pp. 91–93, 1984.
[25] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Berlin and New York : Springer-Verlag, 2003.
[26] Z. Lin and W. Liu, "On maxima of periodograms of stationary processes", *Ann. Statist.* vol. 37, no. 5, pp. 2676–2695, 2009.
[27] E. Masry, "Recursive probability density estimation for weakly dependent stationary processes", *IEEE Trans. Inform. Theory*, vol. 32, pp. 254–267, 1986.
[28] E. Masry, "Almost sure convergence of recursive density estimators for stationary mixing processes", *Statist. Probab. Lett.*, vol. 5, pp. 249–254, 1987.
[29] E. Masry and L. Gÿorfi, "Strong consistency and rates for recursive probability density estimators of stationary processes", *J. Multivariate Anal.*, vol. 22, pp. 79–93, 1987.
[30] A. Mokkadem, M. Pelletier and Y. Slaoui, "The stochastic approximation method for the estimation of a multivariate probability density", *J. Statist. Plann. Inference*, vol. 139, pp. 2459–2478, 2009.

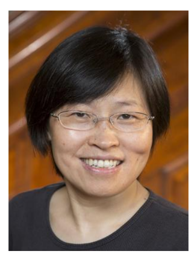
[31] A. Mokkadem, M. Pelletier and Y. Slaoui, “Revisiting Révész’s stochastic approximation method for the estimation of a regression function”, *Alea*, vol. 6, pp. 63–114, 2009.
[32] E. A. Nadaraya, “On estimating regression”, *Theor. Probab. Appl.*, vol. 9, no. 1, pp. 141–142, 1964.
[33] H. T. Nguyen, “Asymptotic normality of recursive density estimators in Markov processes”, *Publ. Inst. Statist. Univ. Paris*, vol. 26, pp. 73–93, 1981.
[34] H. T. Nguyen, “Recursive nonparametric estimation in stationary Markov processes”, *Publ. Inst. Statist. Univ. Paris*, vol. 29, pp. 65–84, 1984.
[35] P. Révész, “Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I, II.” *Studia Sci. Math. Hungar*, vol. 8, pp. 391–398, 469–472, 1973.
[36] P. Révész, “How to apply the method of stochastic approximation in the nonparametric estimation of a regression function”, *Math. Aperationsforsch. Statist. Ser. Statistics*, vol. 8, no. 1, pp. 119–126, 1977.
[37] P. M. Robinson, “Nonparametric estimators for time series”, *J. Time Ser. Anal.* vol. 4, no. 3, pp. 185–207, 1983.
[38] P. M. Robinson, “On the consistency and finite-sample properties of nonparametric kernel time series regression, autoregression and density estimators”, *Ann. Inst. Statist. Math.*, vol. 38, no. 3, pp. 539–549, 1986.
[39] H. Robbins and S. Monro, “A Stochastic Approximation Method”, *Ann. Math. Statist*, vol. 22, no. 3, pp. 400–407, 1951.
[40] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function”, *Ann. Math. Statist.*, vol. 27, no. 3, pp. 832–827, 1956.
[41] G. G. Roussas, “Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function: application to regression and hazard rate estimation”, *J. Nonparam. Statist.*, vol. 1, no. 3, 171–195, 1992.
[42] G. G. Roussas and L. T. Tran, “Asymptotic normality of the recursive kernel regression estimate under dependence conditions”, *Ann. Stat.*, vol. 20, no. 1, pp. 98–120, 1992.
[43] B. W. Silverman, *Density Estimation For Statistics And Data Analysis,* Monographs on Statistics and Applied Probability, London: Chapman & Hall, 1986.
[44] S. Smale and Y. Yao, “Online Learning Algorithms”, *Found. Comput. Math.*, vol. 6, no. 2, pp. 145–170, 2006.
[45] S. Smale and D. X. Zhou, “Online learning with Markov sampling”, *Anal. Appl.*, vol. 7, no. 1, pp. 87–113, 2009.
[46] V. Strassen, “Almost sure behaviour of sums of independent random variables and martingales”, *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, vol. 2, pp. 315–343, 1967.
[47] H. Takahata, “Almost sure convergence of density estimators for weakly dependent stationary processes”, *Bull. Tokyo. Gakugei Univ. Ser. IV*, vol. 32, pp. 11–32, 1980.
[48] H. Tong, *Non-Linear Time Series: A Dynamic System Approach*, Oxford: Oxford University Press, 1990.
[49] L. T. Tran, “Recursive density estimation under dependence”, *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 1103–1108, 1989.
[50] L. T. Tran, “Recursive kernel density estimators under a weak dependence condition”, *Ann. Inst. Statist. Math.*, vol. 42, no. 2, pp. 305–329, 1990.
[51] J. A. Vilar-Fernández and J. M. Vilar-Fernández, “Recursive local polynomial regression under dependence conditions”, *Test*, vol. 9, no. 1, pp. 209–232, 2000.
[52] H. Walk, “Strong universal pointwise consistency of recursive regression estimates”, *Ann. Inst. Statist. Math.* vol. 53, no. 4, pp. 691–707, 2001.
[53] H. Walk and S. Yakowitz, “Iterative nonparametric estimation of a log-optimal portfolio selection function”, *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 324–333, 2002.
[54] L. Wang and H. Liang, “Strong uniform convergence of the recursive regression estimator under ϕ-mixing conditions”, *Metrika* vol. 59, no. 3, pp. 245–261, 2004.
[55] G. S. Watson, “Smooth regression analysis”, *Sankhya, Ser. A*, vol. 26, no. 24, pp. 359–372, 1964.
[56] E. J. Wegman and H. I. Davies, “Remarks on some recursive estimators of a probability density”, *Ann. Statist.*, vol. 7, no. 2, pp. 316–327, 1979.
[57] W. Wertz, “Sequential and recursive estimators of the probability density”, *Statistics*, vol. 16, no. 2, pp. 277–295, 1985.
[58] C. T. Wolverton and T. J. Wagner, “Asymptotically optimal discriminant functions for a pattern classification”, *IEEE Trans. Inform. Theory*, vol. IT-15, no. 2, pp. 258–265, 1969.
[59] W. B. Wu, “Nonlinear system theory: another look at dependence”, *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 40, pp. 14150–14151, 2005.
[60] W. B. Wu, “Asymptotic theory for stationary processes”, *Stat. Interface.*, vol. 4, no. 2, pp. 207–226, 2011.
[61] W. B. Wu, Y. Huang and Y. Huang, “Kernel estimation of time series, an asymptotic theory”, *Stoc. Proc. Appl.*, vol. 120, no. 12, pp. 2412–2431, 2010.
[62] W. B. Wu and X. Shao, “Limit theorems for iterated random functions”, *J. Appl. Probab.*, vol. 41, no. 2, pp. 425–436, 2004.
[63] H. Yamato, “Sequential estimation of a continous probability density funciton and mode”, *Bull. Math. Statist.*, vol. 14, pp. 1–12, 1971.
[64] Y. Yao, “On complexity issue of online learning algorithms”, *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6470–6481, 2010.

ACKNOWLEDGMENT

The authors thank the Associate Editor and two anonymous referees for their many constructive comments.



Yinxiao Huang received the Ph.D. degree in statistics in 2011 from The University of Chicago, Chicago, IL.
She is currently a visiting Assistant Professor of Statistics at University of Illinois, Urbana-Champaign. Her research interests include nonparametric estimation, long memory time series, locally stationary processes and survival data analysis.



Xiaohong Chen received the Ph.D. degree in economics in 1993 from University of California, San Diego, CA.
She is currently a Professor of Economics at Yale University, New Haven, CT. Her research interests include econometric theory, semi/nonparametric estimation and inference methods, and nonlinear time series.

Wei Biao Wu received the Ph.D. degree in statistics in 2001 from The University of Michigan, Ann Arbor, MI.
He is currently a Professor of Statistics at The University of Chicago, Chicago, IL. His research interests include probability theory, statistics and econometrics. He is currently interested in estimating covariance matrices of temporally observed series.