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 Krzyžak and Pawlak [24], Greblicki and Pawlak [11], Krzyžak [23], Walk [52], Gyorfi 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
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] present the recursive kernel estimates and the functional and Hilbert space under Markov sampling. Chen and White [22]; Duflo [8]; Kushner and Yin [25]. Smale and Wolfowitz [22]; Duflo [8]; Kushner and Yin [25]. Smale and Yao [44] and Yao [64] study 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 estimators 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 $c \leq s_n/t_n \leq c$ 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 $|g(x) - g(x')| \leq C|x - x'|$ 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 $X \in L^p, p \geq 1$, if $||X||_p := \mathbb{E}[|X|^p]^{1/p} < \infty$ and write $||X|| := ||X||_2$. The symbol $\Rightarrow$ means convergence in distribution. Denote by $C^k$ 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

\[
g_n(x) = \sum_{i=1}^{n} Y_i K_{b_n}(x, X_i) / \sum_{i=1}^{n} K_{b_n}(x, X_i),
\]

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 $b_n \neq b_{n+1}$, then one needs to re-calculate all $K_{b_n}(x, X_i)$ for $i = 1, \ldots, n$, thus having computational complexity $O(n)$. The latter problem becomes more severe if one wants to evaluate $g$ at multiple points $x_1, \ldots, x_i$ since in this case the computational complexity becomes $O(nl)$. The memory requirement is also $O(n)$ 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:

\[
\hat{g}_n(x) = \frac{\sum_{i=1}^{n} Y_i K_{b_n}(x, X_i)}{\sum_{i=1}^{n} K_{b_n}(x, X_i)} = \frac{\tilde{f}_n(x)}{f_n(x),
\]

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 $B_n = \sum_{i=1}^{n} b_i$,

\[
\tilde{f}_n(x) = \tilde{f}_{n-1}(x) + B_{n-1}^{-1} b_n [K_{b_n}(x, X_n)/b_n - \tilde{f}_{n-1}(x)],
\]

\[
\hat{g}_n(x) = \hat{g}_{n-1}(x) + \gamma_n [Y_n - \hat{g}_{n-1}(x)],
\]

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 $O(1)$ step and the memory complexity for the algorithm is also $O(1)$. When the quantity $\tilde{f}_n(x) B_n$ in (3) is replaced by $nb_n$ then the estimator $\hat{g}_n$ is very similar to the recursive estimator obtained by the method of stochastic approximation (Révész [35]). In (3), $\hat{e}_n := Y_n - \hat{g}_n-1(x)$ can be regarded as the one-step-ahead prediction error. The new information at time $n$ comes in the form of $\hat{e}_n$ which is used to update $g$ together with the previous estimator $\hat{g}_{n-1}$. The residuals $\hat{e}_n$ 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

\[
X_i = R(\ldots, \varepsilon_{i-1}, \varepsilon_i),
\]

where $\varepsilon_i, i \in \mathbb{Z}$, 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

\[
Y_i = G(X_i, \varepsilon_i),
\]

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

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

\[
X_{i+1} = G(X_i, \varepsilon_{i+1}),
\]
where $\epsilon_i$ are i.i.d., by letting $Y_t = X_{t+1}$ and $\epsilon_i = \epsilon_{i+1}$. 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 $l \geq 1$ let $F_l(x;\mathcal{F}_k) = P(X_{k+l} \leq x|\mathcal{F}_k)$ 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 $f(x) = E f_l(x;\mathcal{F}_k)$ Assume throughout the paper that there exists a constant $C > 0$ such that

$$\text{ess sup}_{x} \sup_{\mathcal{F}_k} f_l(x;\mathcal{F}_k) \leq C.$$  \hspace{1cm} (7)

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

$$\theta_k = \sup_{x \in \mathbb{R}} \|P_0 f_l(x;\mathcal{F}_{k-1})\|.$$  \hspace{1cm} (8)

Let $\epsilon_{i_1}, \epsilon_{i_2}, \ldots, \epsilon_{i_k} \in \mathbb{Z}$ be i.i.d. If $l \leq k$, let $\mathcal{F}_{k,l}(1) = (\ldots, \epsilon_{1-i_1}, \epsilon_{i_1}, \epsilon_{i_1+1}, \ldots, \epsilon_k)$ be a coupled version of $\mathcal{F}_k$ with $\epsilon_i$ in the latter replaced by $\epsilon_{i_j}$. If $l > k$ let $\mathcal{F}_{k,l}(1) = \mathcal{F}_k$. Note that for $k \geq 1$, $\mathbb{E}(f_l(x;\mathcal{F}_{k-1})) = f_l(x;\mathcal{F}_0)$ and $\mathbb{E}(f_l(x;\mathcal{F}_{k-1})|\mathcal{F}_{l-1}) = f_{l+1}(x|\mathcal{F}_{l-1}) = E(f_k(x;\mathcal{F}_{0,0})|\mathcal{F}_0)$. We can then interpret $\theta_k$ as the contribution of $\epsilon_0$ in predicting $X_k$. If $f_l(x;\mathcal{F}_{k-1})$ does not functionally depend on $\epsilon_0$, then $\theta_k = 0$. A simple upper bound for $\theta_k$ is the following functional dependence measure

$$\delta_k = \sup_{x \in \mathbb{R}} \|f_l(x;\mathcal{F}_{k-1}) - f_l(x;\mathcal{F}_{k-1,0})\|.$$  \hspace{1cm} (9)

Note that $\mathbb{E}(f_l(x;\mathcal{F}_{k-1,0})|\mathcal{F}_0) = \mathbb{E}(f_l(x;\mathcal{F}_{k-1})|\mathcal{F}_{l-1})$, we have $\theta_k \leq \delta_k$. 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 $\delta_k$. Proposition 1 provides bounds for $\delta_k$ for linear processes, which can have infinite variances, for example, $\alpha$-stable processes with $0 < \alpha < 2$. Proposition 2 concerns nonlinear time series that can be expressed as iterated random functions.

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

$$X_t = \sum_{j=0}^\infty a_j \epsilon_{t-j}$$  \hspace{1cm} (10)

is well-defined, and, if $\sup_{x\in \mathbb{R}} E(|f_{\epsilon}(x)| + |f'_{\epsilon}(x)|) < \infty$, then $\delta_k = O(|a_k|^{\alpha/2})$.

**Proposition 2.** Consider the nonlinear time process $X_{t+1} = G(X_t, \epsilon_{t+1})$ in (6), where $\epsilon_j$ are i.i.d. and $G(\cdot, \cdot)$ satisfies $G(x, \epsilon_0) \in C^\alpha$, $\alpha > 1$ and

$$\|G(x, \epsilon_0) - G(x', \epsilon_0)\|_\tau < 1.$$  \hspace{1cm} (11)

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

$$\sup_{x \neq u} \frac{\sup_{u \neq u'} |f(x|u) - f(x|u')|}{|u - u'|^\alpha} \leq H, \alpha > 0.$$  \hspace{1cm} (12)

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

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

$$X_t = a_1 X_{t-1}^+ + a_2 X_{t-1}^- + \epsilon_t,$$  \hspace{1cm} (13)

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

**Example 2.** (ARCH Processes). Assume that $\epsilon_t$ are i.i.d. with density $f_{\epsilon}$ and have a finite $\tau$th moment, $\tau > 0$. Define

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

where $a_1$ and $a_2$ are real parameters such that $|a_1|^\tau < 1$, $|a_2|^\tau < 1$. 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 $f(x|u) = f_x(x/g(u))/g(u)$. Assume that $\sup_u f'_x(u) < \infty$. Elementary calculations show that $\sup_{u,x} |\partial f(x|u)/\partial u| < \infty$, implying (12) with $\alpha = 1$. As in Example 1 the moment condition here on $\epsilon_t$ 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 $\Theta_m := \sum_{k=m}^\infty \theta_k < \Delta_m < \infty$.

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

Condition **D1** implies that the cumulative contribution of the input $\epsilon_0$ in predicting future values $\{X_k\}_{k \geq 1}$ 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 $\epsilon_0$ can have a long-range predictive capability, due to the slow decay of $\delta_k$. 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 $(X_t)$ given in (4). As in (2), we consider the recursive kernel density estimator

$$\hat{f}_n(x) = \frac{1}{B_n} \sum_{i=1}^n K_{h_i}(x, X_i),$$  \hspace{1cm} (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 \(\sup_{u} |K(u)| + \int_{R} u^2 |K(u)| du \leq C_K\). Let \(\epsilon := \int_{R} K^2(s) ds < \infty\).

Condition (K) is satisfied by many popular choices of kernels including the rectangle kernel \(K(v) = 1_{|v|<1/2}\), the Epanechnikov kernel \(K(v) = 3(1 - v^2)1_{|v|<1}/4\), the triangular kernel \(K(v) = (1 - |v|)1_{|v|<1}\), the biweight kernel \(K(v) = (15(1 - v^2)31_{|v|<1}/16\) and the triweight kernel \(K(v) = 35(1 - v^2)^21_{|v|<1}/32\) among others, where \(1\) is the indicator function.

A. Asymptotic Properties

Let \(B_{n,l} = \sum_{i=1}^{n} b_i\), \(l \in Z\), 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), \(B_n \to \infty\), \(\sum_{k=1}^{\infty} B_{2^{k+1}, 2^{k}} < \infty\), and \(B_{n,2} = o(B_n)\);

(ii) Condition (D2), \(b_i \asymp \epsilon^{-\beta} f_0(i)\) where \(2(1 - \gamma) < \beta < 1\) and \(f_0(\cdot)\) is a slowly varying function.

Then we have the central limit theorem

\[
\sqrt{B_n}[\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)] = N(0, \kappa f(x)).
\]

If in addition \(B_{2,3}^2 = o(B_n)\), then (16) holds with \(\mathbb{E}\hat{f}_n(x)\) replaced by \(f(x)\).

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 \(b_i \asymp \epsilon^{-\beta}\), where \(0 < \beta < 1\). Then under (D1), \(B_n = \sum_{i=1}^{n} b_i \asymp n^{-\beta}\), and \(B_{n,2} = \sum_{i=1}^{n} b_i^2 \asymp n^{-2\beta}\) (resp. \(\log(n)\) or \(O(1)\)) 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 \(\lim_{n \to \infty} W_n = c\) if \(\limsup_{n} W_n = c\) and \(\liminf_{n} W_n = c\).

**Theorem 2.** Assume condition (K), \(\sup_{u} |f'(u)| < \infty\), and either one of the following:

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

(ii) Condition (D2), \(b_i \asymp \epsilon^{-\beta} f_0(i)\) where \(2(1 - \gamma) < \beta < 1\) and \(f_0(\cdot)\) is a slowly varying function.

Then we have the law of the iterated logarithm

\[
\lim_{n \to \infty} \frac{\sqrt{B_n}[\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)]}{\sqrt{\log \log B_n}} = \pm \sqrt{2\kappa f(x)}.
\]
detail space \( W_j \) be given by \( V_{j+1} = V_j \bigoplus W_j \), which has the orthonormal basis \( \{ \psi_{j,k}(x) = 2^j/2\psi(2^jx - k), k \in \mathbb{Z} \} \) for some function \( \psi \). Then \( V_{m} \bigoplus W_{m} \bigoplus V_{m+1} \bigoplus W_{m+2} \ldots \) is dense in \( C_2(\mathbb{R}) \). Let

\[
f(x) = \sum_{k \in \mathbb{Z}} c_{m,k} \phi_{m,k}(x) + \sum_{j \geq m, k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x),
\]

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

\[
\hat{f}_n^w(x) = \sum_{k \in \mathbb{Z}} \hat{c}_{m,k} \phi_{m,k}(x) = \frac{1}{n} \sum_{k \in \mathbb{Z}} \phi_{m,k}(x) \sum_{i=1}^{n} \phi_{m,k}(X_i)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{k \in \mathbb{Z}} 2^{\frac{m}{2}} \phi(2^m x - k) 2^{\frac{m}{2}} \phi(2^m X_i - k)
\]

\[
= (nb_n)^{-1} \sum_{i=1}^{n} K \left( \frac{x - X_i}{b_n} \right),
\]

where \( K(u,v) = \sum_{k \in \mathbb{Z}} \phi(u - k) \phi(v - k) \) 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

\[
\hat{f}_n^w(x) = \frac{\sum_{i=1}^{n} K(x/b_i, X_i/b_i)}{\sum_{i=1}^{n} b_i}
\]  

**Theorem 4.** Assume that \( K(\cdot, \cdot) \) is a symmetric bivariate kernel satisfying \( \sup_{u,v} |K(u,v)| < \infty \), \( \sup_{u} \int_{\mathbb{R}} |K(u,v)| dv < \infty \), \( K(u,v) \to 0 \) as \( |u| = |v| \to \infty \), and for \( \sigma^2(u) := \int_{\mathbb{R}} K^2(u,v) dv \), \( \inf_u \sigma^2(u) > 0 \). 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

\[
\frac{B_n[\hat{f}_n^w(x_n) - \mathbb{E}[\hat{f}_n^w(x_0)]]}{\sqrt{B_n(x_0)}} \to N(0, f(x_0)).
\]  

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

\[
\lim_{n \to \infty} B_n[\hat{f}_n^w(x_n) - \mathbb{E}[\hat{f}_n^w(x_0)]] = \pm \sqrt{2f(x_0)}.
\]  

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

**D. Bandwidth Selection**

For the traditional kernel density estimate

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

we have the bias \( \mathbb{E}[\hat{f}_n(x) - f(x)] \sim b_n^2 \nu_2 f''(x)/2 \), where \( \nu_2 = \int u^2 |K(u)du| \), and the variance \( \text{var}[\hat{f}_n(x)] \sim (nb_n)^{-1} f(x) \kappa \) (Silverman [43]). The optimal bandwidth \( b_n \) minimizing the asymptotic mean squared error (AMSE) has the form \( b_n = \tilde{c} n^{-1/5} \) with \( \tilde{c} = (f(x)\kappa)^{1/5}(\nu_2 f''(x))^{-2/5} \). 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 \( 2 \log(n)b_n^2 f''(x)/3 \) if \( \beta = 1/3 \), or \( O((nb_n)^{-1}) \) if \( \beta > 1/3 \). And \( \text{var}[\tilde{f}_n(x)] \sim (1 - \beta)(nb_n)^{-1} f(x) \kappa \). Let \( \beta = 1/5 \). Then the AMSE-optimal \( c \) satisfies

\[
c = \tilde{c} \left\{ \left( (1 - 3/5)^2 \right)^{1/5} \right\} \quad = 5^{-1/5} \tilde{c}.
\]

Hence we can first conduct a pilot study and, based on part of the sample \( X_1, \ldots, X_N \) with a relatively large \( N \), we can apply the classical bandwidth selection procedure (for example, Jones et al [21]) and obtain an estimate of \( \tilde{c} \) for the non-recursive estimate. Then we use \( b_i = ci^{-\beta} \) with \( c = 5^{-1/5} \tilde{c} \) for our recursive estimates. With this choice, we have \( \text{AMSE}(\tilde{f}_n(x))/\text{AMSE}(\hat{f}_n(x)) = (2 \times 5^{-2/5}) \approx 1.1 \), 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 \( V_p(x) = \mathbb{E}[|Y_i|^p|X_i = x] \), 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 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

\[
\sqrt{B_n}[J_n(x_0) - \mathbb{E}[J_n(x_0)]] \to N(0, V_2(x_0) f(x_0) \kappa).
\]

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

\[
\sqrt{B_n} \left( \tilde{g}_n(x_0) - \frac{\mathbb{E}[J_n(x_0)]}{\mathbb{E}[\hat{f}_n(x_0)]} \right) \to N(0, \sigma^2(x_0) \kappa \sqrt{f(x_0)}).
\]

**B. A Law of the Iterated Logarithm**

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

(i) Condition (D1), \( (\log n)^{2p/(p-2)} + B_{n,2} \log n = O(B_n) \)

and \( \sum_{k=1}^{\infty} B_{2^{k+2}}/B_{2^k} < \infty \); or

(ii) Condition (ii) of Theorem 2.
Then for any \( x \) with \( f(x) > 0 \), we have the LIL

\[
\lim_{n \to \infty} \frac{B_n^{1/2} (g_n(x) - g(x))}{\sqrt{\log \log B_n}} = \pm \frac{\sqrt{2 \kappa}}{\sqrt{f(x)}} \sigma(x). \tag{28}
\]

C. Uniform Almost Sure Convergence

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

\[
\delta_k^* = \left\| \sup_{x \in \mathbb{R}} |f_1(x|F_{k-1}) - f_1(x|F_{k-1}(0))| \right\|.
\]

(29)

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

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

\[
\sup_{x \in [-\gamma, a]} B_n |g_n(x) - \mathbb{E}g_n(x)| = O_{a.s.}(\epsilon_n), \tag{30}
\]

where \( \epsilon_n = n^{1/\gamma} \log n + \sqrt{B_n \log n} \).

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

\[
\sup_{x \in [-\gamma, a]} B_n |g_n(x) - \mathbb{E}g_n(x)| = O_{a.s.}(\eta_n), \tag{31}
\]

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

**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 \( n^{3/2-\gamma} \ell_2(n) \) which accounts for dependence. Note that if \( 3/2 - \beta - \gamma < 1/\rho \) or \( 2(1-\gamma) < \beta \), then the second term in \( \eta_n \) is absorbed into the first term \( \epsilon_n \).

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

\[
\max_{x \in [-\gamma, a]} |g_n(x) - g(x)| = O_{a.s.}(\rho_n/B_n + B_{n,3}/B_n). \tag{32}
\]

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

\[
Y_i = g(X_i) + e_i, \tag{33}
\]

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

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

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

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 \( X_i = \theta |X_{i-1}| + \epsilon_i \), where \( \epsilon_i \) 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 \( \phi(.) \) and \( \Phi(.) \) are the standard normal density and distribution functions (cf. Anděl et al. [1] and Tong [48]);

(ii) Fractionally integrated process \( (1-B)^d X_i = \epsilon_i \) where \( \epsilon_i \) 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 \( X_i = \sum_{j=0}^{\infty} \psi_j \epsilon_{i-j} \) 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.
APPENDIX

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

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:

\[
\|X_k - X_{k,(0)}\|_\tau = O(L^k), \quad \text{where} \quad X_{k,(0)} = R(F_{k,(0)}).
\]

Let \( \delta_k \) be such that \( \sup_{x \in \mathbb{R}} \| f_1(x|F_{k-1}) - f_1(x|F_{k-1,(0)}) \|_p \). Then by (12),

\[
|f_1(x|F_{k-1}) - f_1(x|F_{k-1,(0)})| \leq \min(2C_0, H|X_k - X_{k,(0)}|\).
\]

Hence, if \( \alpha \) is short-range dependent satis-

\[
\sum_{k=2}^\infty \mathbb{E}W_k \leq \int_{|R|} |K(u)| ||\max_{n \leq 2k} |\Lambda_n(x,u)| ||^2 \, du
\]

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

\[
\sum_{k=2}^\infty \mathbb{E}W_k \leq \int_{|R|} |K(u)| ||\max_{n \leq 2k} |\Lambda_n(x,u)| ||^2 \, du
\]

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

Lemma 1 concerns the short-range dependent case with \( \Theta_1 < \infty \). Under long-range dependence, \( \Theta_1 \) 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

\[
\left| \max_{k \leq n} |\Lambda_k(x,u)| \right|^2 \leq \sum_{i=1}^n \left( \sum_{i=1}^n b_{i,j} \right)^2 =: \Xi_n.
\]

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

Proof of Lemma 2. Write \( \Lambda_n = \Lambda_n(x,u) \). Let \( \Lambda_n(t) \) be a coupled version of \( \Lambda_n \) with \( \varepsilon_t \) in the latter replaced by the i.i.d. copy \( \varepsilon_t \). By Jensen’s and the triangle inequalities,

\[
\left| \max_{k \leq n} |\Lambda_k(x,u)| \right|^2 \leq \left( \sum_{i=1}^n \left( \sum_{j=1}^n b_{i,j} \right)^2 \right)^2 =: \Xi_n.
\]

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

We now prove the second assertion. Let \( b^n_i = i^{-\beta} \ell_0(i) \), \( \delta^n_i = i^{-\gamma} \ell_0(i) \), \( i \geq 1 \), \( b^n_0 = \delta^n_0 = 1 \) and \( b^n_{i+1} = \delta^n_{i+1} = 0 \) if \( i < 0 \), and \( \Xi^n_n = \sum_{i=-\infty}^{n} \sum_{j=-\infty}^{n} b^n_i \delta^n_{i-j} \). By Karamata’s Theorem, \( \sum_{i=1}^n b^n_i \approx n b^n_n \) and

\[
\sum_{i=-\infty}^{n} \sum_{j=-\infty}^{n} b^n_i \delta^n_{i-j} \approx \sum_{i=1}^{n} \theta(\delta^n_i)^2 \left( \sum_{i=1}^{n} b^n_i \right)^2
\]

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

**Proof of Theorem 1.** By Lemmas 1 and 2, it follows that $R_n = o_{a.s.}\left(B_{n/2}\right)$ since almost sure convergence implies convergence in probability, hence it suffices to show that

$$\sum_{i=1}^{n} D_i / B_{n/2} \implies N(0, 2k\sigma f(x_0)),$$

(47)

where $D_i = K_{b_i}(x_0, X_i) - E[K_{b_i}(x_0, X_i)] / F_{i-1}$. To this end, we shall apply the martingale central limit theorem; see Hall and Heyde [17]. Since $|D_i|$ is bounded by $2C_{K_i}$ and $B_n \to \infty$, for any $\epsilon > 0$, we have $D_i^2 / |D_i| \geq B_{n/2} / \epsilon = 0$ for all large $n$. It remains to show that

$$B_n^{-1} \sum_{i=1}^{n} E[D_i^2 / F_{i-1}] \to k\sigma f(x_0) \text{ in probability}.$$  

(48)

Since

$$E[D_i^2 / F_{i-1}] = E[K_{b_i}(x_0, X_i)] / F_{i-1} - E^2[K_{b_i}(x_0, X_i)] / F_{i-1},$$

and elementary calculation readily shows that $E[K_{b_i}^2(x_0, X_i)] / F_{i-1} \Rightarrow k\sigma f(x_0) b_i$ as $i \to \infty$; and $E[K_{b_i}(x_0, X_i)] / F_{i-1} = O(b_i)$. Notice that $B_n \to \infty$, it follows by Toeplitz lemma that,

$$B_n^{-1} \sum_{i=1}^{n} E[D_i^2 / F_{i-1}] = B_n^{-1} \sum_{i=1}^{n} \left[ E[K_{b_i}^2(x_0, X_i)] / F_{i-1} - E^2[K_{b_i}(x_0, X_i)] / F_{i-1} \right] \to \sum_{i=1}^{n} (k\sigma f(x_0) b_i + O(b_i^2))$$

$$\Rightarrow k\sigma f(x_0)$$

(49)

Hence (47), and then (16) follows. ◇

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

$$B_n \left( \hat{f}_n(x) - E \hat{f}_n(x) \right) = M_n(x) + R_n(x),$$

(50)

where

$$M_n(x) = \sum_{i=1}^{n} \left[ K_{b_i}(x, X_i) - E[K_{b_i}(x, X_i)] / F_{i-1} \right]$$

is a martingale relative to the filtration $F_n$. Then the Theorem follows from Lemmas 1, 2 above and 3 below. ◇

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

$$\lim_{n \to \infty} \frac{M_n(x_0)}{\sqrt{B_n \log \log B_n}} = \pm \sqrt{2k\sigma f(x_0)}$$

(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 $B$ and a sequence of nonnegative random variables $\tau_1, \tau_2, \ldots$ such that

$$\{M_k, k \geq 1\} = \mathbb{P}(B(\tau_k), k \geq 1),$$

where $\tau_k = \sum_{i=1}^{k} \tau_i$, for $k \geq 1$,

$$\mathbb{E}(\tau_k | F_{k-1}) = \mathbb{E}(D_k^2 | F_{k-1}) \text{ a.s.},$$

(53)

and for $C_q = 2(8/\pi^2)^{q/2}$ for $q > 2$.

$$\mathbb{E}(\tau_k^2 | F_{k-1}) \leq C_q \mathbb{E}(D_k^4 | F_{k-1}) \text{ a.s. for } q > 2.$$  

(54)

Let $t_n = \mathbb{E}(T_n) = \sum_{i=1}^{n} \mathbb{E}(D_i^2)$. By (7) and condition (K), $|E[K_{b_i}(x_0 - X_i)] / F_{i-1}| \leq C_b$, which by sup$_n |f'(u)| < \infty$ implies $\mathbb{E}(D_i^2) = \kappa f(x_0) b_i(1 + O(b_i^2))$. Hence

$$t_n = \kappa f(x_0) B_n + O(B_{n/2}).$$

(55)

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

$$T_n - t_n = \sum_{i=1}^{n} (\tau_i - \mathbb{E}(\tau_i | F_{i-1})) + \sum_{i=1}^{n} (\mathbb{E}(\tau_i | F_{i-1}) - \mathbb{E}(\tau_i))$$

$$= t_n + Q_n.$$  

(56)

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

$$\left\| \max_{1 \leq n} |P_i| \right\|^2 \leq 4\|P_n\|^2 \leq 4 \sum_{i=1}^{n} \mathbb{E}(\tau_i^2)$$

$$\leq 4C_1 \sum_{i=1}^{n} \mathbb{E}(D_i^4) = O(B_n).$$

(57)

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

$$Q_n = \sum_{i=1}^{n} \left( \mathbb{E}(D_i^2 | F_{i-1}) - \mathbb{E}(D_i^2) \right)$$

$$= \sum_{i=1}^{n} \left( \mathbb{E}(K_{b_i}^2(x_0, X_i) | F_{i-1}) - \mathbb{E}(K_{b_i}^2(x_0, X_i)) + O(b_i^2) \right)$$

$$= \int_{\mathbb{R}} K^2(x) \Lambda_n(x, 0, v)dv + O(B_{n/2}).$$

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

$$T_n - t_n = O_{a.s.}(\Delta_n),$$

where $\Delta_n = B_{n/2} \log n + B_{n/2}$. By Lévy’s modulus of continuity for Brownian motions, we have almost surely that

$$\sup_{k \leq n} |B(T_k) - B(T_k)| \leq \sup_{|x-y| \leq C_{\Delta_n}, 0 \leq y \leq t_n} |B(x) - B(y)|$$

$$\leq C(\Delta_n \log(B_n / \Delta_n))^{1/2}.$$  

(58)

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

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

(59)
Since $t_n \to \infty$, by the law of the iterated logarithm for Brownian motions, we have
\[ \lim_{n \to \infty} \frac{\mathbb{E}(t_n)}{\sqrt{2t_n \log \log t_n}} = \pm 1. \quad (60) \]
Hence by (53), (58) and (59), we have (52).

**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
\[ B_n(f_n^w(x) - \mathbb{E}f_n^w(x)) = M_n + R_n, \quad (61) \]
where, noting that
\[ \mathbb{E}(K(x/b_i, X_i/b_i)|\mathcal{F}_{i-1}) = b_i \int_{\mathbb{R}} K(x/b_i, u)f_1(b_i,u)|\mathcal{F}_{i-1})du, \]
the term
\[ R_n = \sum_{k=1}^{\infty} R_{n,k}, \quad (62) \]
where $R_{n,k} = \sum_{i=1}^{n} b_i \int_{\mathbb{R}} K(x/b_i, u)\mathcal{P}_{i-k}f_1(b_i,u)|\mathcal{F}_{i-1})du$.
By the Cauchy-Schwarz inequality and conditions on $K(\cdot, \cdot)$, we have
\[ \|R_{n,k}\|^2 = \sum_{i=1}^{n} b_i^2 \mathbb{E} \left( \int_{\mathbb{R}} K(x/b_i, u)\mathcal{P}_{i-k}f_1(b_i,u)|\mathcal{F}_{i-1})du \right)^2 \leq \sum_{i=1}^{n} b_i^2 \int_{\mathbb{R}} |K(x/b_i, u)|du	imes \mathbb{E} \int_{\mathbb{R}} |K(x/b_i, u)||\mathcal{P}_{i-k}f_1(b_i,u)|\mathcal{F}_{i-1})|^2du = \sum_{i=1}^{n} b_i^2 \mathbb{O}(\theta_1^2 b_i^2). \quad (63) \]
Hence under the short-range dependent case (D1), $\|R_{n}\| = \mathbb{O}(\Theta_1 b_i^2)$, 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
\[ \mathbb{E}K^2(x/b, X_i/b) = \int_{\mathbb{R}} K^2(x/b, u)f(bu)du = \left( \int_{|bu-x| \geq b^{1/2}} + \int_{|bu-x| < b^{1/2}} \right) K^2(x/b, u)f(bu)du = \sigma^2(x/b)f(x) + o(1) \quad (64) \]
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.

**Proof of Theorem 5.** To simplify notation we let $x = x_0$.
Let $G_i = (\cdots, e_{i-1}, e_i, F_i)$, $\xi_i = Y_iK_{b_i}(x, X_i)$ and
\[ \Lambda_n^\circ(x,v) := \sum_{i=1}^{n} g(x-b_iv)b_i(f_1(x-b_iv|F_{i-1}) - f(x-b_iv)) \quad (65) \]
Note that $\xi_i$ is $G_i$-measurable and $\mathbb{E}(Y_i|X_i = x) = g(x)$. As in (39), we can write
\[ L_n := \sum_{i=1}^{n} (\mathbb{E}(\xi_i|G_{i-1}) - \mathbb{E}\xi_i) = \int K(v)\Lambda_n^\circ(x,v)dv. \quad (66) \]
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 $d_i = \xi_i - \mathbb{E}(\xi_i|G_{i-1})$. Then (26) follows from
\[ \frac{1}{\sqrt{B_n}} \sum_{i=1}^{n} d_i \Rightarrow N[0, V_2(x)f(x)\kappa]. \quad (67) \]
To this end, we shall apply the martingale central limit theorem. Note that $|\mathbb{E}(\xi_i|G_{i-1})| = \mathbb{O}(b_i)$. Hence $|\mathbb{E}(d_n^2|G_{i-1}) - \mathbb{E}(\xi_i^2|G_{i-1})| = \mathbb{O}(b_i^2)$. By Lemmas 1 and 2, we have
\[ \frac{1}{B_n} \sum_{i=1}^{n} (\mathbb{E}(\xi_i^2|G_{i-1}) - \mathbb{E}(\xi_i^2)) = V_2(x)K^2\left( \frac{x-v}{b_i} \right) \mathbb{E}(f_1(x|G_{i-1}) - f(v))dv = \mathcal{O}(B_n^{-1}) \int K^2(u)\mathcal{O}_n(x, u)du = o_P(1) \]
since $V_2(x)$ is Lipschitz continuous. Note that $K$ has bounded support,
\[ \frac{1}{B_n} \sum_{i=1}^{n} \mathbb{E}(d_n^2|G_{i-1}) = \frac{1}{B_n} \sum_{i=1}^{n} \mathbb{E}(\xi_i^2) + o_P(1) \]
\[ = \frac{1}{B_n} \sum_{i=1}^{n} \int V_2(x-b_iv)K^2(x-b_iv)du + o_P(1) = V_2(x)f(x) \int K^2(v)dv + o_P(1). \]
The proof is now complete since $\sum_{i=1}^{n} ||B_n^{-1/2}d_i||^p \leq 2pB_n^{-p/2} \sum_{i=1}^{n} ||\xi_i||^p = \mathcal{O}(B_n^{-p/2}) \to 0$, Lindeberg’s condition is satisfied. By Slutsky’s Lemma, we have (27).

**Proof of Theorem 6.** As in the proof of Theorem 5, let $G_i = (\cdots, e_{i-1}, e_i, F_i)$, $\zeta_i = (Y_i - g(x))K_{b_i}(x, X_i)$, $D_i^0 = \zeta_i - \mathbb{E}(\zeta_i|G_{i-1})$ and $M_n = \sum_{i=1}^{n} D_i$.
We shall show that
\[ \lim_{n \to \infty} \frac{\sum_{i=1}^{n} D_i^0}{\sqrt{B_n \log \log B_n}} = \pm 2\sqrt{2\kappa f(x)\sigma(x)} \quad (68) \]
almost surely. To this end, we shall use the arguments in the proof of Lemma 3. There exists a standard Brownian motion $\mathbb{B} \mathbb{R}$ and a sequence of nonnegative random variables $\tau_1, \tau_2, \cdots$, with $T_k = \sum_{i=1}^{k} \tau_i$ such that $\{M_k\}_{k \geq 1} = \mathcal{D} \{\mathbb{E}(T_k^2)\}_{k \geq 1}$, and
\[ \mathbb{E}(\tau_k^2|G_{k-1}) = \mathbb{E}((D_k^0)^2|G_{k-1}) \text{ a.s.} \]
\[ \mathbb{E}((\tau_k^2)^{p/2}|G_{k-1}) \leq C_p \mathbb{E}((D_k^0)^{p/2}|G_{k-1}) \text{ a.s.} \quad (69) \]
where $C_p = 2(8/\pi^2)^{p-1}(p+1)$. Since $g$ is Lipschitz continuous and $\mathbb{E}(Y_i - g(x)|F_i) = g(X_i) - g(x)$, by (7), we have
\[ |\mathbb{E}(\zeta_i|G_{i-1})| = |\mathbb{E}((g(X_i) - g(x))K_{b_i}(x, X_i)|F_{i-1})| \leq \int |g(u) - g(x)|K\left( \frac{x-u}{b_i} \right)f_1(u|F_{i-1})du \leq b_iC_1 \int |g(x-b_iv) - g(x)|K(v)dv \leq b_iC \int C|b_i|K(v)dv = \mathcal{O}(b_i^2) \quad (70) \]
since $K$ has bounded support. Similarly, $\mathbb{E}\{g(X_i) - g(x)\}K_{b_i}^2(x, X_i) = \mathcal{O}(b_i^2)$, and since $\sigma^2(\cdot)$ is also Lipschitz...
continuous, $\mathbb{E}[\sigma^2(X_i) - \sigma^2(x)]K_b^2(x, X_i) = O(b_i^2)$. By the condition $\sup_n |f'(u)| < \infty$, we have $\mathbb{E}K^2_n(x, X_i) = b_i \kappa f(x) + O(b_i^2)$. Note that $\mathbb{E}((Y_i - g(x))^2|F_i) = \sigma^2(X_i) + (g(X_i) - g(x))^2$. Then

$$
E \sum_{i=1}^k \tau_i^g = \mathbb{E} \sum_{i=1}^k (\zeta_i - \mathbb{E}(\zeta_i|G_{i-1}))^2
$$

$$
= \sum_{i=1}^k (E \zeta_i^2 + O(b_i^4))
$$

$$
= \sum_{i=1}^k \mathbb{E} \left\{ (\sigma^2(x_i) + |g(x_i) - g(x)|^2)K^2_b(x, X_i) + O(B_{k,4}) \right\}
$$

$$
= \sigma^2(x) b_i \kappa f(x) + O(b_i^2) + O(B_{k,2}).
$$

(71)

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

$$
\|\max_{k \leq n} \mathbb{E} \left[ \sum_{i=1}^k \tau_i^g - \mathbb{E}(\tau_i^g|G_{i-1}) \right] \|_{p/2}^p
\leq c_p \| \sum_{i=1}^n \tau_i^g - \mathbb{E}(\tau_i^g|G_{i-1}) \|_{p/2}^p
\leq c_p \| \sum_{i=1}^n \tau_i^g - \mathbb{E}(\tau_i^g|G_{i-1}) \|_{p/2}^p
\leq c_p \| \sum_{i=1}^n \tau_i^g - \mathbb{E}(\tau_i^g|G_{i-1}) \|_{p/2}^p
\leq c_p \| \sum_{i=1}^n \tau_i^g - \mathbb{E}(\tau_i^g|G_{i-1}) \|_{p/2}^p
$$

(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), (68) in view of

$$
T_n^g - \sigma^2(x) B_n \kappa f(x)
$$

$$
= \mathbb{E} \left[ \sum_{i=1}^k \tau_i^g - \mathbb{E}(\tau_i^g|G_{i-1}) \right] \|_{p/2}^p
$$

(73)

By the representation (70) and the arguments in Lemmas 1 and 2, we have $\sum_{i=1}^n (E \zeta_i|G_{i-1}) - E \zeta_i = o_{a.s.}(B_{1/2})$. So we have a LIL for $\sum_{i=1}^n (\zeta_i - E \zeta_i)$ in the form of (68) with $D_g$ therein replaced by $\zeta_i - E \zeta_i$. With the latter, we have (28) by Slutsky’s theorem.

**Proof of Theorem 7.** Let $\tilde{n} = 2[\log n / \log 2]$, $Y_i = Y_i 1_{|Y_i| \leq \sqrt{n}}$, $\xi'(x) = Y_i \kappa b_i K_b(x, X_i)$, $Y''_i = Y_i - Y_i', \xi''(x) = Y''_i \kappa b_i K_b(x, X_i)$ and $R_d = \sum_{i=1}^d |Y''_i| + \mathbb{E}(|Y''_i| |G_{i-1})$. Elementary calculations show that there exists a constant $C$ such that

$$
\sum_{d=1}^\infty \mathbb{E} R_d^2 \leq C \sum_{d=1}^\infty \mathbb{E} |Y_1^c 1_{|Y_1| \geq 2d^{1/2}}| \leq C \mathbb{E} |Y_1|^2.
$$

By the Borel-Cantelli lemma, $R_d = o_{a.s.}(2d^{1/2})$ as $d \to \infty$. Since $K$ is bounded, we have

$$
\sup_x \sum_{i=1}^n \left| \xi''(x) - \mathbb{E}(\xi''(x)|G_{i-1}) \right| \leq C_K R_d = o_{a.s.}(n^{1/p}),
$$

(74)

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

$$
W_n(x) = \sum_{i=1}^n \left( \xi''(x) - \mathbb{E}(\xi''(x)|G_{i-1}) \right).
$$

(75)

Let $N = n^5$ and $|u|_N = |uN|/N$. Since $K$ is Lipschitz continuous,

$$
E \left| \sum_{i=1}^n \left( \xi''(x) - \mathbb{E}(\xi''(x)|G_{i-1}) \right) \right| \leq \sum_{i=1}^n \mathbb{E} |Y_1^c 1_{|Y_1| \leq 2^d^{1/2}}(K_b(x, X_i) - K_{b_i}(|x|_N, X_i))| \leq \sum_{i=1}^n 2^{d^{1/2}} C \frac{1}{b_i} \leq C n^{1/p} B_{n-1}/N.
$$

Notice that the same bound holds for $\sup_x \sum_{i=1}^n \left( \mathbb{E}(\xi''(x)|G_{i-1}) - \mathbb{E}(\xi''(|x|_N)|G_{i-1}) \right)$. Therefore

$$
E \left| \sum_{i=1}^n \left( \xi''(x) - \mathbb{E}(\xi''(x)|G_{i-1}) \right) \right| \leq C n^{1/p} B_{n-1}/N.
$$

(76)

By (7), since $V_p(\cdot)$ is bounded on $[-a, a + \psi]$, there is a constant $c_0$ such that

$$
\mathbb{E}\left[ \sum_{i=1}^n \left( \xi''(x) - \mathbb{E}(\xi''(x)|G_{i-1}) \right) \right] \leq C_n B_{n-1}/N.
$$

(77)

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

$$
P \left[ \max_{|x| \leq a} |W_n(x)| \geq \lambda \chi_n \right] \leq a \chi_n \exp \left\{ - \frac{\lambda^2 \chi_n^2}{2 n \kappa C_n} \right\},
$$

(78)

which is of order $O(n^{-2})$ by choosing a sufficiently large $\lambda$. Therefore, by (76) and since $B_{n-1}/N = o(\chi_n)$, we have

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

(79)

Recall (65) for $A_{\infty}^g(x, v)$ and (66) for $L_n(x)$. Similarly as Lemmas 1 or 2, under (i) or (ii), we have $\| \mathbb{E}(\xi''(x, v)|G_{i-1}) \|^2 = O(B_{n,2})$ or $O(\chi_n)$, respectively. Hence, by (79) and (74), we have (30) and (31), respectively.

**Proof of Corollary 1.** Under the differentiability conditions on $f, g$, elementary calculations show that $g_n(x) = \mathbb{E}J_n(x)/\mathbb{E}f_n(x) = g(x) + O(B_{3/2}, B_n)$. Write

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

(80)

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

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

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

(81)

Define event $A_1 = \{L_n(x)| \geq a_{N, B_{1/2}}\}$. Let $S_n = (L_n(x), L_n(x)')$, where $|x - x'| \geq n^{-\chi}$. Note that $\tau_n := E(\xi''_n(1_{|e_i| \leq \log n}) = 1 + O(n^{-A})$ for any $A > 2$. Under (K), we obtain $E(L_n(x)') = \tau_n f(x)\kappa(B_n + O(B_{3,5}))$ and, for some constant $C > 0$,

$$
E[L_n(x) L_n(x)']
$$
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 \( a_{N,c} = \sqrt{2 \log N} + o(1) = o(\min(\gamma_n^{-1/3}, c_n^{-1})) \). Hence by (87) of Lemma 4, we obtain \( P(A_i) = (1 + o(1)) N / N \) and \( P(A_i \cap A_j) = (1 + o(1)) \rho^2 / N^2 \) uniformly in \( i \) and \( j \). By the inclusion-exclusion formula, we have

\[
P \left( \bigcup_{1 \leq i \leq N} A_i \right) \geq \sum_{i=1}^{N} P(A_i) - \sum_{1 \leq i < j \leq N} P(A_i \cap A_j) \rightarrow g - \frac{\rho^2}{2}.
\]

(83)

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

\[
\limsup_{n \to \infty} P \left( \sup_{x \in [0,1]} |Q_n(x) - \mathbb{E}Q_n(x)| \geq \theta \right) \leq 2 \exp \left( -\frac{\theta^2}{C_1 \theta + C_2 B_{n,3}} \right).
\]

(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

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

(86)

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

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

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

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

\[
P(|S_n|_* \geq \lambda) - P(|N|_* \geq \lambda R_n^{-1/2})
\]

\[= o(P(|N|_* \geq \lambda R_n^{-1/2})) \quad (87)
\]

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**


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.