NONPARAMETRIC ESTIMATION FOR STATIONARY PROCESSES

Wei Biao Wu
The University of Chicago

TECHNICAL REPORT NO. 536

Departments of Statistics
The University of Chicago
Chicago, Illinois 60637

April 11, 2003

NONPARAMETRIC ESTIMATION FOR STATIONARY PROCESSES

BY WEI BIAO WU

April 11, 2003

Department of Statistics, The University of Chicago

Abstract. We consider the kernel density and regression estimation problem for a wide class of causal processes. Asymptotic normality of the kernel estimators is established under minimal regularity conditions on bandwidths. Optimal uniform error bounds are obtained without imposing strong mixing conditions. The proposed method is based on martingale approximations and provides a unified framework for nonparametric time series analysis, and enables one to launch a systematic study for dependent observations.

Keywords: Kernel estimation; Nonlinear time series; Regression; Central limit theorem; Martingale; Markov chains; Linear processes.

1 Introduction

Let $\{\varepsilon_n\}_{n\in\mathbb{Z}}$ be a sequence of independent and identically distributed (iid) random elements. Consider the process

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

where F is a measurable function. Clearly $\{X_n\}_{n\in\mathbb{Z}}$ is a stationary and causal process and represents a huge class of time series models. For example, if F has the linear form $F(\ldots, \varepsilon_{n-1}, \varepsilon_n) = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$, where $\{a_i\}_{i=0}^{\infty}$ is a square summable sequence and ε_n has mean 0 and finite variance, then X_n is well-defined and corresponds to the widely used linear process which includes as special cases the important in practice ARMA and fractional ARIMA models. As another example, consider the class of nonlinear time series defined by the recursion

$$X_n = R_{\varepsilon_n}(X_{n-1}), \tag{2}$$

where R is a bivariate measurable function. Under suitable conditions on R, the recursion (2) has a unique and stationary initial distribution (Barnsley and Elton (1988) and Diaconis and Freedman (1999)). By iterating (2), X_n is also of the form (1). For different forms of R in (2), one can get threshold autoregressive models (TAR, Tong (1990)), AR models with conditionally heteroscedastic errors (ARCH, Engle (1982)), random coefficient models (Nicholls and Quinn, 1982) and exponential autoregressive models (EAR, Haggan and Ozaki (1981)) among others.

In this paper, we consider the kernel density and regression estimation problem of the class defined by (1). For $p \in \mathbb{N}$ let π_p be the joint distribution function of $(X_{n-1}, \ldots, X_{n-p})$. The density of π_p usually does not have a closed form and hence it is preferably estimated by nonparametric methods. Let K be a probability density function on $\mathbb{R}$. Following Rosenblatt (1956), we have that given the data sequence $X_{1-p}, \ldots, X_n$, the kernel density estimator of the joint density of $(X_{n-1}, \ldots, X_{n-p})$ at $\mathbf{x} = (x_{-1}, \ldots, x_{-p})$ is given by

$$f_n(\mathbf{x}) = \frac{1}{n} \sum_{t=1}^{n} \prod_{i=1}^{p} K_{b_n}(x_{-i} - X_{t-i}), \tag{3}$$

where $K_b(x) = K(x/b)/b$ and bandwidths $\{b_n\}$ satisfy the natural condition

$$b_n \to 0 \quad \text{and } nb_n^p \to \infty. \tag{4}$$

Under mild conditions on f, we show that (4) suffices to guarantee the asymptotic normality of $\sqrt{nb_n^p}[f_n(\mathbf{x}) - \mathbb{E}f_n(\mathbf{x})]$. In addition, we obtain an optimal error bound of $\sup_{\mathbf{x}} |f_n(\mathbf{x}) - f(\mathbf{x})|$. The optimality here is in the sense that the bound is as sharp as the one in the iid case, which further supports Hart's (1996) whitening by windowing principle.

To formulate the regression problem, we consider the model

$$Y_n = G(X_{n-1}, \ldots, X_{n-p}, \theta_n), \tag{5}$$

where $\{\theta_n\}_{n \in \mathbb{Z}}$ are also iid error terms and θ_n is independent of $(\ldots, \varepsilon_{n-2}, \varepsilon_{n-1})$. Then for $\mathbf{x} = (x_{-1}, \ldots, x_{-p}) \in \mathbb{R}^p$, the Nadaraya-Watson estimator of the regression function

$$g(\mathbf{x}) = \mathbb{E}[Y_n | (X_{n-1}, \ldots, X_{n-p}) = \mathbf{x}] = \mathbb{E}G(\mathbf{x}, \theta_n) \tag{6}$$

has the form

$$g_n(\mathbf{x}) = \frac{S_n(G; \mathbf{x})}{f_n(\mathbf{x})}, \text{ where } S_n(G; \mathbf{x}) = \frac{1}{n} \sum_{t=1}^{n} Y_t \prod_{i=1}^{p} K_{b_n}(x_{-i} - X_{t-i}). \tag{7}$$

Asymptotic normality of $g_n(\mathbf{x})$ is established under (4) and some regularity conditions on G and f.

There is an extensive literature concerning limiting properties of (3), (7) and other related issues such as optimal bandwidth selection for the case in which $\{X_t\}$ are iid (see for example, Silverman (1986) and Devroye and Györfi (1984)). For dependent random variables, Rosenblatt (1970) considered Markov sequences with geometric ergodicity and showed asymptotic normality of the kernel density estimators. Asymptotic issues for strongly mixing processes have been discussed by Robinson (1983), Singh and Ullah (1985), Castellana and Leadbetter (1986), Györfi, Härdle, Sarda and Vieu (1989) and Bosq (1996) among others. The recent work by Yu (1993), Neumann (1998) and Kreiss and Neumann (1998) deal with β-mixing processes. Further references are given in the excellent review by Härdle, Lütkepohl and Chen (1997) and Tjostheim (1994). However, in many practical situations the required strong mixing conditions are usually unverifiable and might also be too restrictive. The strong mixing properties of linear processes have been discussed by many people including Gorodetskii (1977), Withers (1981), Pham and Tran (1985) and Doukhan (1994) among others, where it is shown that fast decay rates of a_n are needed.

One of the main contributions at the technical level of this paper is the introduction of a martingale-based technique that enables us to study large sample properties in nonparametric time series analysis and more specifically derive central limit theorems and obtain estimates of the uniform error bound. As an alternative to strong mixing conditions, our assumption appears sufficiently mild, and more importantly, is easily verifiable in practice. In addition, the proposed approach enables one to obtain optimal results which are as sharp as those in the iid setting. We believe that our approach can be extended to the testing, model selection, minimax theory and other problems in the statistical inference for linear and nonlinear processes.

The rest of the paper is structured as follows. Main results are presented in Section 2 and their proofs are given in Section 4. Section 3 contains applications to nonlinear autoregressive time series and linear processes.

2 Main Results

Throughout the paper it is assumed that the kernel K is a nonnegative function on $\mathbb{R}$, $\int_{\mathbb{R}} K(u)du = 1$, $\sup_{u\in\mathbb{R}} K(u) \leq K_0 < \infty$ and K has a bounded support; namely, there exists an $M < \infty$ such that $K(x) = 0$ if $|x| \geq M$. Write $\kappa = \int_{\mathbb{R}} K^2(u)du$ and

$K(\mathbf{u}) = K(u_1)\dots K(u_p)$ for the vector $\mathbf{u} = (u_1, \dots, u_p)$. Let $\mathbf{X}_n = (\dots, \varepsilon_{n-1}, \varepsilon_n)$ be the shift process. For a p dimensional vector $\mathbf{x} = (x_{-1}, \dots, x_{-p})$, denote by $f(\mathbf{x}|\mathbf{X}_{n-p-1})$ the conditional density of $(X_{n-1}, \dots, X_{n-p})$ at $\mathbf{x}$ given $\mathbf{X}_{n-p-1} = (\dots, \varepsilon_{n-p-2}, \varepsilon_{n-p-1})$. Assume that there exists a $f_* < \infty$ such that

$$\sup_{\mathbf{y}\in\mathbb{R}^p} f(\mathbf{y}|\mathbf{X}_0) \le f_* \tag{8}$$

holds with probability 1. Define the projection operator $\mathcal{P}_k\xi = \mathbb{E}(\xi|\mathbf{X}_k) - \mathbb{E}(\xi|\mathbf{X}_{k-1})$.

2.1 Asymptotic normality

To state the central limit theorem concerning the Nadaraya-Watson estimator (7), we need the following regularity condition.

Condition 1. *Let $V_2(\mathbf{y}) := \mathbb{E}G^2(\mathbf{y}, \theta_n)$ and $g(\mathbf{y})$ be continuous at $\mathbf{y} = \mathbf{x}$. There exists a $\delta > 0$ such that $V_{2+\delta}(\mathbf{y}) = \mathbb{E}[|G(\mathbf{y}, \theta_n)|^{2+\delta}]$ is bounded at a neighborhood of $\mathbf{x}$.*

Theorem 1. *Assume Condition 1, (4), (8) and*

$$\sup_{\mathbf{y}} \sum_{t=1}^{\infty} \|\mathcal{P}_0 f(\mathbf{y}|\mathbf{X}_t)\| < \infty. \tag{9}$$

Then

$$\sqrt{nb_n^p}\{S_n(G;\mathbf{x}) - \mathbb{E}[S_n(G;\mathbf{x})]\} \Rightarrow N[0,\ V_2(\mathbf{x})f(\mathbf{x})\kappa^p]. \tag{10}$$

Condition (9) is used instead of the strong mixing conditions. It holds for linear as well as many nonlinear processes; see Section 3 for some examples.

By letting $G \equiv 1$ in Theorem 1, we have the following central limit theorem for the joint density estimator (3).

Corollary 1. *Assume (4), (8) and (9). Then for all $\mathbf{x} \in \mathbb{R}^p$,*

$$\sqrt{nb_n^p}\{f_n(\mathbf{x}) - \mathbb{E}[f_n(\mathbf{x})]\} \Rightarrow N[0,\ f(\mathbf{x})\kappa^p]. \tag{11}$$

Corollary 2. *If $f(\mathbf{x}) > 0$ at a given $\mathbf{x} \in \mathbb{R}^p$, then under the conditions of Theorem 1, we have*

$$\sqrt{nb_n^p}\left\{\frac{S_n(G;\mathbf{x})}{S_n(\mathbf{x})} - \frac{\mathbb{E}S_n(G;\mathbf{x})}{\mathbb{E}S_n(\mathbf{x})}\right\} \Rightarrow N\{0, [V_2(\mathbf{x}) - g^2(\mathbf{x})]\kappa^p/f(\mathbf{x})\}. \tag{12}$$

In kernel estimation theory it is routine to compute the bias $\mathbb{E}S_n(G;\mathbf{x})/\mathbb{E}S_n(\mathbf{x}) - g(\mathbf{x})$. If g is twice differentiable, K is symmetric and f is differentiable at $\mathbf{x}$, then it is easily seen that the bias is of order $\mathcal{O}(b_n^2)$.

Proof of Corollary 2. Let $\nu_n(\mathbf{x}) = \mathbb{E}S_n(G;\mathbf{x})$ and $\mu_n(\mathbf{x}) = \mathbb{E}S_n(\mathbf{x})$. Since $f(\mathbf{x}) > 0$, K has bounded support and g is continuous at $\mathbf{x}$, we have $\nu_n(\mathbf{x})/\mu_n(\mathbf{x}) \to g(\mathbf{x})$. Observe that

$$\begin{aligned} S_n(G;\mathbf{x}) - S_n(\mathbf{x})\frac{\nu_n(\mathbf{x})}{\mu_n(\mathbf{x})} &= \{S_n(G;\mathbf{x}) - S_n(\mathbf{x})g(\mathbf{x}) - n[\nu_n(\mathbf{x}) - \mu_n(\mathbf{x})g(\mathbf{x})]\} \\ &\quad + [S_n(\mathbf{x}) - n\mu_n(\mathbf{x})][g(\mathbf{x}) - \nu_n(\mathbf{x})/\mu_n(\mathbf{x})] =: A_n + B_n. \end{aligned}$$

By Corollary 1, $B_n = o_{\mathbb{P}}(1/\sqrt{nb_n^d})$ and $S_n(\mathbf{x}) \xrightarrow{\mathcal{P}} f(\mathbf{x})$. Hence by Theorem 1, $\sqrt{nb_n^p}A_n \Rightarrow N\{0, [V_2(\mathbf{x}) - g^2(\mathbf{x})]\kappa^p f(\mathbf{x})\}$, which by the Slutsky theorem yields (12). ♢

2.2 An optimal uniform bound

Corollary 1 indicates that $f_n(\mathbf{x}) - \mathbb{E}[f_n(\mathbf{x})]$ has a magnitude of the order $1/\sqrt{nb_n^p}$. Theorem 2 below provides a uniform error bound. For $\alpha = (\alpha_1, \ldots, \alpha_p)$, where α_i are nonnegative integers, let the partial derivative $g^{(\alpha)}(\mathbf{u}) = \partial^{\alpha_1+\ldots+\alpha_p} g(\mathbf{u})/\partial u_1^{\alpha_1} \ldots \partial u_p^{\alpha_p}$. For $n \geq 0$ define

$$\Delta_n = \sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_p=0}^{1} \int_{\mathbb{R}^p} \|\mathcal{P}_0 f^{(\alpha)}(\mathbf{u}|\mathbf{X}_n)\|^2 d\mathbf{u}.$$

A function h is said to be Lipschitz continuous with index $\eta > 0$ if there exists $L_h < \infty$ such that for all x and y, $|h(x) - h(y)| \leq L_h|x - y|^\eta$.

Theorem 2. *Let K be Lipschitz continuous with index $\eta > 0$ and $\mathbb{E}(|X_1|^\alpha) < \infty$ for some $\alpha > 0$. Assume (4), (8) and*

$$\sum_{t=0}^{\infty} \sqrt{\Delta_t} < \infty. \tag{13}$$

Then

$$\sqrt{nb_n^p} \sup_{\mathbf{y} \in \mathbb{R}^p} |f_n(\mathbf{y}) - \mathbb{E}[f_n(\mathbf{y})]| = \mathcal{O}(\log n) \text{ almost surely.} \tag{14}$$

Furthermore, the bound in (14) is reduced to $\mathcal{O}(\sqrt{\log n})$ if $\log^2 n = o(nb_n^p)$.

Corollary 3. *Let conditions of Theorem 2 be satisfied; let f be differentiable and for some $C > 0$, $|f(\mathbf{y} + \mathbf{z}) - f(\mathbf{y}) - \mathbf{z}f'(\mathbf{y})| \leq C|\mathbf{z}|^2$ for all $\mathbf{y}$ and $\mathbf{z}$ in $\mathbb{R}^p$. In addition assume*

$\int_{\mathbb{R}} sK(s)ds = 0$ *and* $b_n = (n^{-1}\log n)^{1/(p+4)}$. *Then*

$$\sup_{\mathbf{y}\in\mathbb{R}^p} |f_n(\mathbf{y}) - f(\mathbf{y})| = \mathcal{O}[(n^{-1}\log n)^{2/(p+4)}] \text{ almost surely.} \tag{15}$$

Let $d = 1$. For iid random variables $\{Y_t\}$ with density function f_Y, let $f_{Y,n}(x) = n^{-1}\sum_{i=1}^n K_{b_n}(x - Y_i)$ be the kernel density estimator. Bickel and Rosenblatt (1973) obtained a distributional limit of the uniform error bound. Their result suggests that the uniform error bound $\sup_x |f_{Y,n}(x) - f_Y(x)|$ has magnitude $\mathcal{O}_P[\sqrt{(\log n)/(nb_n)}]$ if $b_n = n^{-\delta}$ for $0 < \delta < 1/2$. Stute (1982) showed that

$$\left(\frac{n}{\log n}\right)^{2/5} \sup_{|x|\le\lambda} \left|\frac{f_{Y,n}(x) - f_Y(x)}{f_Y(x)}\right|$$

converges almost surely to a non-zero constant for all $\lambda > 0$ such that $f_Y(x) > 0$ if $|x| \le \lambda$. Therefore, the optimal uniform error bound for $\sup_{|x|\le\lambda} |f_{Y,n}(x) - f_Y(x)|$ is of the order $\mathcal{O}[(n^{-1}\log n)^{2/5}]$. In this sense, Corollary 3 seems interesting in that the optimal uniform bound is obtained for dependent data without any α (strong) mixing assumptions. For strong mixing processes with exponentially decaying strong mixing coefficients, one can obtain uniform bounds which are close to optimal; see Theorem 2.2 in Bosq (1998) where a bound of the form $(n^{-1}\log n)^{2/(p+4)} \log\ldots\log n$ is obtained. If the process is β-mixing (or absolutely regular) with suitable mixing rate, Yu (1993) obtains optimal manimax rates; see also Neumann (1998).

Proof of Corollary 3. The argument to establish the result is standard. Under the conditions of the corollary, it is easily seen that the bias is

$$|\mathbb{E}f_n(\mathbf{y}) - f(\mathbf{y})| = \left|\int_{\mathbb{R}^p} K(\mathbf{v})[f(\mathbf{y} - b_n\mathbf{v}) - f(\mathbf{y}) + b_n\mathbf{v}f'(\mathbf{y})]d\mathbf{v}\right| = \mathcal{O}(b_n^2).$$

Hence for $b_n = (n^{-1}\log n)^{1/(p+4)}$, we have

$$\begin{aligned}\sup_{\mathbf{y}\in\mathbb{R}^p} |f_n(\mathbf{y}) - f(\mathbf{y})| &\le \sup_{\mathbf{y}\in\mathbb{R}} |f_n(\mathbf{y}) - \mathbb{E}[f_n(\mathbf{y})]| + \mathcal{O}(b_n^2) \\ &= \frac{\mathcal{O}(\sqrt{\log n})}{\sqrt{nb_n^p}} + \mathcal{O}(b_n^2) = \mathcal{O}(b_n^2)\end{aligned}$$

almost surely by Theorem 2. ♢

3 Applications

In this section we present two important applications and show that the crucial conditions (9) in Theorem 1 and (13) in Theorem 2 are satisfied. A very general nonlinear autoregressive model is considered in Section 3.1 and the commonly used linear process is discussed in Section 3.2.

3.1 Nonlinear time series

Let $q \geq 1$ be a fixed integer. Consider the nonlinear AR(q) model

$$X_n = R_{\varepsilon_n}(X_{n-1}, \ldots, X_{n-q}), \tag{16}$$

where R is a measurable function such that $\{X_n\}$ adopts a stationary distribution. We shall mention the important special case, the nonlinear autoregressive conditional heteroscedastic model which assumes the form

$$X_n = m(X_{n-1}, \ldots, X_{n-q}) + \sigma(X_{n-1}, \ldots, X_{n-q})\varepsilon_n,$$

where m and σ^2 are the conditional mean and variance functions, respectively. It is easily seen that X_n can be expressed in terms of (1) by iterating R in (16). Let $\{\varepsilon'_j\}$ be an iid copy of $\{\varepsilon_j\}$ and $X'_n = F(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)$ be the coupled version of X_n. Assume that there exists $C, \alpha > 0$ and $0 < r(\alpha) < 1$ such that

$$\mathbb{E}\{|F(\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n) - F(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)|^\alpha\} \leq Cr^n(\alpha) \tag{17}$$

holds for all $n \in \mathbf{N}$. Without loss of generality let $\alpha < 1$; otherwise, Hölder's inequality could be employed. Condition (17) is actually very mild. In the special case in which $q = 1$ (namely (2)), (2) admits a unique stationary distribution if

$$\mathbb{E}(\log L_\varepsilon) < 0,\ \mathbb{E}(L_\varepsilon^\alpha) + \mathbb{E}[|x_0 - R_\varepsilon(x_0)|^\alpha] < \infty,\ \text{where } L_\varepsilon = \sup_{x' \neq x} \frac{|R_\varepsilon(x) - R_\varepsilon(x')|}{|x - x'|} \tag{18}$$

holds for some $\alpha > 0$ and x_0 (Diaconis and Freedman (1999)). It is easily verified that these conditions are satisfied for many popular nonlinear time series models such as TAR, RCA, ARCH and EAR under suitable conditions on model parameters. Condition (18) actually also implies (17) (cf Lemma 3 in Wu and Woodroofe (2000)).

An important issue of (16) is to estimate the conditional mean

$$Qh(\mathbf{x}) = \mathbb{E}[h(X_n)|(X_{n-1}, \ldots, X_{n-q}) = \mathbf{x}],$$

where h satisfies $\mathbb{E}[h^2(X_n)] < \infty$ and $\mathbf{x} = (x_{-1}, \ldots, x_{-q}) \in \mathbb{R}^q$. In the case $q = 1$, Qh corresponds to the transition kernel of the Markov chain X_n, which plays the central role in the Markov chain theory. Let $G(\mathbf{x}, \varepsilon_n) = h(R_{\varepsilon_n}(\mathbf{x}))$ and as in (5), $Y_n = G(X_{n-1}, \ldots, X_{n-q}, \varepsilon_n)$. Then $g(\mathbf{x}) = Qh(\mathbf{x})$ and then asymptotic normality, as developed in Section 2.1, holds.

Let $\mathbf{u}$ be a p-dimensional vector. By the structure of process X_n defined in (16), $f(\mathbf{u}|\mathbf{X}_n)$ is equal to the conditional density of $(X_{n+p}, \ldots, X_{n+1})$ at $\mathbf{u}$ given $(X_n, \ldots, X_{n-q+1})$. Thus we can also write $f(\mathbf{u}|X_n, \ldots, X_{n-q+1})$ for $f(\mathbf{u}|\mathbf{X}_n)$. Theorem 3 states that a Lipschitz continuity condition on $f(\cdot|\mathbf{z})$ ($z \in \mathbb{R}^q$) suffices to ensure (9). Since such a condition is directly related to the data-generating mechanism of the process X_n, it seems tractable; see Example 1 for an illustration.

Theorem 3. *Assume (8), (17), and that there exists C and $\beta > 0$ such that for all $\mathbf{z}$ and $\mathbf{z}'$ in $\mathbb{R}^q$,*

$$\sup_{\mathbf{u}\in\mathbb{R}^p} |f(\mathbf{u}|\mathbf{z}) - f(\mathbf{u}|\mathbf{z}')| \le C|\mathbf{z} - \mathbf{z}'|^\beta. \tag{19}$$

Then, $\sup_{\mathbf{u}\in\mathbb{R}^p} \|\mathcal{P}_0 f(\mathbf{u}|\mathbf{X}_n)\| = \mathcal{O}(\rho^n)$ *for some* $\rho \in (0,1)$ *and hence (9) holds.*

Proof. Let $n > q$. Since $(X'_n, \ldots, X'_{n-q+1})$ is independent of $\mathbf{X}_0$ and has the same distribution as $(X_n, \ldots, X_{n-q+1})$,

$$\begin{aligned} \mathbb{E}[f(\mathbf{y}|X'_n, \ldots, X'_{n-q+1})|X_0] &= \mathbb{E}[f(\mathbf{y}|X'_n, \ldots, X'_{n-q+1})] \\ &= \mathbb{E}[f(\mathbf{y}|X_n, \ldots, X_{n-q+1})]. \end{aligned}$$

By (19) and (8),

$$\begin{aligned} \|\mathcal{P}_0 f(\mathbf{y}|\mathbf{X}_n)\| &\le \|\mathbb{E}[f(\mathbf{y}|X_n, \ldots, X_{n-p+1})|X_0] - \mathbb{E}[f(\mathbf{y}|X_n, \ldots, X_{n-p+1})]\| \\ &= \|\mathbb{E}[f(\mathbf{y}|X_n, \ldots, X_{n-p+1})|X_0] - \mathbb{E}[f(\mathbf{y}|X'_n, \ldots, X'_{n-p+1})|X_0]\| \\ &\le \|f(\mathbf{y}|X_n, \ldots, X_{n-p+1}) - f(\mathbf{y}|X'_n, \ldots, X'_{n-p+1})\| \\ &\le C \sum_{j=n-q+1}^{n} \|\min(1, |X_j - X'_j|^\beta)\|. \end{aligned}$$

By (17) and Hölder's inequality we obtain

$$\begin{aligned}\mathbb{E}[\min(1, |X_n - X_n'|^{2\beta})] &\leq \mathbb{E}[|X_n - X_n'|^{\min(\alpha, 2\beta)}] \\ &\leq [\mathbb{E}(|X_n - X_n'|^{\alpha})]^{\min(1,\, 2\beta/\alpha)} = \mathcal{O}(\rho^{2n}),\end{aligned}$$

where $\rho = r(\alpha)^{\min(1/2,\, \beta/\alpha)}$. Thus, $\sup_{\mathbf{y}\in\mathbb{R}^p} \|\mathcal{P}_0 f(\mathbf{y}|\mathbf{X}_n)\| = \mathcal{O}(\rho^n)$. $\diamondsuit$

Example 1. Let $p = q = 1$ and consider the $AR(1)$ model with $ARCH$ errors

$$X_n = R_{\varepsilon_n}(X_{n-1}) = \theta_1 X_{n-1} + \varepsilon_n \sqrt{\theta_2^2 + \theta_3^2 X_{n-1}^2}, \tag{20}$$

where $\theta_1, \theta_2, \theta_3$ are real-valued parameters and $\varepsilon, \varepsilon_n$ are iid random variables. Observe that $L_\varepsilon = \sup_x |\partial R_\varepsilon(x)/\partial x| \leq |\theta_1| + |\theta_3 \varepsilon|$. By (18), a simple sufficient condition for the existence of a stationary distribution is $\mathbb{E}[\log(|\theta_1| + |\theta_3\varepsilon|)] < 0$ and $\mathbb{E}(|\varepsilon|^\alpha) < \infty$ for some $\alpha > 0$. Here we allow the case in which ε does not have a mean, namely $\mathbb{E}(|\varepsilon|) = \infty$. By Doukhan (1994, p. 106), the process is geometrical β-mixing if $\mathbb{E}(|\varepsilon|) < \infty$, ε has a nowhere vanishing density and $\lim_{n\to\infty} \mathbb{E}|R_\varepsilon(x)|/|x| < 1$. Notice that none of the above conditions imply another condition, and they have different applicability ranges. On the other hand, our conditions ensure (17), which is the basis of our approach. The classical treatment of dependent data usually imposes strong mixing conditions as the underlying assumptions.

Let F_ε be the distribution function of ε; let f_ε and f_ε' be the density function and its derivative. Then the conditional density $f(z|x) = f_\varepsilon[(z - \theta_1 x)/\sqrt{\theta_2^2 + \theta_3^2 x^2}]/\sqrt{\theta_2^2 + \theta_3^2 x^2}$. Assume that

$$\sup_{u\in\mathbb{R}} [|u f_\varepsilon'(u)| + f_\varepsilon(u)] < \infty. \tag{21}$$

Now we claim that (21) entails (19). Let $u = (z - \theta_1 x)/\sqrt{\theta_2^2 + \theta_3^2 x^2}$. Then after some elementary manipulations,

$$\frac{\partial f(z|x)}{\partial x} = -f_\varepsilon'(u) \frac{\theta_1 \sqrt{\theta_2^2 + \theta_3^2 x^2} + u\theta_3^2 x}{(\theta_2^2 + \theta_3^2 x^2)^{3/2}} - f_\varepsilon(u) \frac{\theta_3^2 x}{(\theta_2^2 + \theta_3^2 x^2)^{3/2}},$$

which entails that $C = \sup_z |\partial f(z|x)/\partial x| < \infty$ in view of (21). So (19) holds with this C and $\beta = 1$.

3.2 Linear process

Let $X_n = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$, where $\sum_{i=0}^{\infty} a_i^2 < \infty$ and ε_i are iid with mean 0 and finite variance; let f_ε be the density function of ε_i. Denote by $\mathcal{C}^p(\mathbb{R})$ the class of functions having up to pth order derivatives.

Theorem 4. *Assume that $f_\varepsilon \in \mathcal{C}^{p+1}(\mathbb{R})$ and*

$$\sum_{i=0}^{p+1} \int_{\mathbb{R}} |f_\varepsilon^{(i)}(x)|^2 dx < \infty. \tag{22}$$

Then

$$\Delta_n = \mathcal{O}\left(\sum_{i=1}^{p} a_{n+i+1}^2\right). \tag{23}$$

Proof. Without loss of generality we consider $p = 2$ and $a_0 = 1$. Let $f_{2,1}$ be the joint density of $(\varepsilon_{n+2} + a_1\varepsilon_{n+1}, \varepsilon_{n+1})$. Then $f_{2,1}(u, v) = f_\varepsilon(u - a_1 v) f_\varepsilon(v)$ and the conditional density of (X_{n+2}, X_{n+1}) given $\mathbf{X}_n$ is $f_{2,1}(u - Y_n, v - Z_n)$, where $Y_n = \sum_{i=2}^{\infty} a_i\varepsilon_{n+2-i}$ and $Z_n = \sum_{i=1}^{\infty} a_i \varepsilon_{n+1-i}$. Now we show that

$$\Delta_n^{(00)} := \int_{\mathbb{R}^2} \|\mathcal{P}_0 f_{2,1}(u - Y_n, v - Z_n)\|^2 du dv = \mathcal{O}(a_{n+2}^2 + a_{n+3}^2). \tag{24}$$

Recall that $\{\varepsilon_i'\}_{i\in\mathbb{Z}}$ is an iid copy of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. Let $\delta_1 = a_{n+3}(\varepsilon_{-1}' - \varepsilon_{-1})$, $\delta_2 = a_{n+2}(\varepsilon_{-1}' - \varepsilon_{-1})$, $Y_n^* = Y_n + \delta_1$ and $Z_n^* = Z_n + \delta_2$. Then

$$\mathbb{E}[f_{2,1}(u - Y_n, v - Z_n)|\mathbf{X}_{-1}] = \mathbb{E}[f_{2,1}(u - Y_n^*, v - Z_n^*)|\mathbf{X}_0]$$

almost surely. By Cauchy's inequality,

$$\begin{aligned} \|\mathcal{P}_0 f_{2,1}(u - Y_n, v - Z_n)\| &= \|\mathbb{E}[f_{2,1}(u - Y_n, v - Z_n) - f_{2,1}(u - Y_n^*, v - Z_n^*)|\mathbf{X}_0]\| \\ &\le \|f_{2,1}(u - Y_n, v - Z_n) - f_{2,1}(u - Y_n^*, v - Z_n^*)\|. \end{aligned}$$

Observe that for a differential function h, we have

$$\begin{aligned} \int_{\mathbb{R}} |h(x+\delta) - h(x)|^2 dx &\le \int_{\mathbb{R}} [\int_0^\delta |h'(x+t)|dt]^2 dx \\ &\le \int_{\mathbb{R}} \delta \int_0^\delta |h'(x+t)|^2 dt dx = \delta^2 \int_{\mathbb{R}} |h'(t)|^2 dt. \end{aligned} \tag{25}$$

By a change of variables we get

$$\begin{aligned}
\Delta_n^{(00)} &\leq \mathbb{E}\int_{\mathbb{R}^2} |f_{2,1}(u-Y_n, v-Z_n) - f_{2,1}(u-Y_n^*, v-Z_n^*)|^2 dudv \\
&= \mathbb{E}\int_{\mathbb{R}^2} |f_{2,1}(u+\delta_1, v+\delta_2) - f_{2,1}(u,v)|^2 dudv \\
&\leq 2\mathbb{E}\int_{\mathbb{R}^2} |f_\varepsilon(u+\delta_1 - a_1(v+\delta_2)) - f_\varepsilon(u-a_1 v)|^2 f_\varepsilon^2(v+\delta_2) dudv \\
&\quad +2\mathbb{E}\int_{\mathbb{R}^2} f_\varepsilon^2(u-a_1 v)|f_\varepsilon(v+\delta_2) - f_\varepsilon(v)|^2 dudv \\
&\leq 2\mathbb{E}[(\delta_1 - a_1\delta_2)^2 + \delta_2^2]\int_{\mathbb{R}} f_\varepsilon^2(u)du \int_{\mathbb{R}} |f_\varepsilon'(v)|^2 dv = \mathcal{O}(a_{n+2}^2 + a_{n+3}^2).
\end{aligned}$$

By a similar argument, (22) implies that

$$\Delta_n^{(\alpha)} := \int_{\mathbb{R}^2} \|\mathcal{P}_0 f_{2,1}^{(\alpha)}(u-Y_n, v-Z_n)\|^2 dudv = \mathcal{O}(a_{n+2}^2 + a_{n+3}^2)$$

holds for $\alpha = (0,1)$, $(1,0)$ and $(1,1)$. Thus $\Delta_n = \mathcal{O}(a_{n+2}^2 + a_{n+3}^2)$. ◇

Corollary 3 and Theorem 4 immediately yield

Corollary 4. *Assume (13) and* $\sum_{i=0}^\infty |a_i| < \infty$. *Then (15) holds.*

If $a_n = n^{-\beta}L(n)$, where $\beta \in (1/2, 1)$ and L is a slowly varying function, then a_n is not summable and the process X_n is long-range dependent. Wu and Mielniczuk (2002) derived limiting distributions of $f_n(x) - \mathbb{E}f_n(x)$ and obtained the interesting dichotomous and trichotomous phenomena for different choices of b_n.

If f_ε is Lipshitz continuous, Wu and Mielniczuk (2002) show that $\sup_y \|\mathcal{P}_0 f_\varepsilon(y|\mathbf{X}_n)\| = \mathcal{O}(|a_n|)$ and hence (9) holds if $\sum_{i=0}^\infty |a_i| < \infty$.

4 Proofs

Let

$$\xi_{n,t} = \sqrt{\frac{b_n^p}{n}} Y_t \prod_{i=1}^p K_{b_n}(x_{-i} - X_{t-i}) \text{ and } \zeta_{n,t} = \sqrt{\frac{b_n^p}{n}} \prod_{i=1}^p K_{b_n}(x_{-i} - X_{t-i}).$$

Lemma 1. *Let* $\{X_{n,i}, i \in \mathbb{Z}\}$, $n \in \mathbb{N}$ *be a triangular array stationary process and* $\{\mathcal{G}_i, i \in \mathbb{Z}\}$ *be an increasing sequence of sigma-algebras such that* $X_{n,i}$ *is* $\mathcal{G}_i$ *measurable. Assume that* $\sup_n \mathbb{E}|X_{n,1}| < \infty$ *and that there exists a* $\delta > 0$ *for which* $\mathbb{E}(|X_{n,1}|^{1+\delta}) = o(n^\delta)$. *Then*

$$\frac{1}{n}\sum_{i=1}^{n}[X_{n,i} - \mathbb{E}(X_{n,i}|\mathcal{G}_{i-p})] \xrightarrow{\mathcal{P}} 0 \tag{26}$$

for all $p \in \mathbb{N}$.

Proof of Lemma 1. Observe that $\mathbb{E}\{|\mathbb{E}(X_{n,i}|\mathcal{G}_{i-1})|^r\} \le \mathbb{E}(|X_{n,1}|^r)$ for $r = 1$ or $r = \delta + 1$. Hence the general case in which $p > 1$ easily follows from (26) with $p = 1$ by considering the sequence $\mathbb{E}(X_{n,i}|\mathcal{G}_{i-1})$ in view of

$$\frac{1}{n}\sum_{i=1}^{n}[X_{n,i} - \mathbb{E}(X_{n,i}|\mathcal{G}_{i-p})] = \frac{1}{n}\sum_{k=0}^{p-1}\sum_{i=1}^{n}[\mathbb{E}(X_{n,i}|\mathcal{G}_{i-k}) - \mathbb{E}(X_{n,i}|\mathcal{G}_{i-k-1})].$$

For any $\eta > 0$ and $\epsilon > 0$ let $a_n = n\epsilon^2\eta$ and $X'_{n,i} = X_{n,i}\mathbf{1}(|X_{n,i}| \le a_n)$. Write $S_n = \sum_{i=1}^{n}[X_{n,i} - \mathbb{E}(X_{n,i}|\mathcal{G}_{i-1})]$ and $S'_n = \sum_{i=1}^{n}[X'_{n,i} - \mathbb{E}(X'_{n,i}|\mathcal{G}_{i-1})]$. Then under the proposed conditions of the lemma,

$$\begin{aligned}
\limsup_{n\to\infty}\mathbb{P}(|S_n| \ge 2n\epsilon) &\le \limsup_{n\to\infty}\mathbb{P}(|S'_n| \ge n\epsilon) + \limsup_{n\to\infty}\mathbb{P}(|S_n - S'_n| \ge n\epsilon)\\
&\le \limsup_{n\to\infty}\frac{1}{n^2\epsilon^2}\mathbb{E}(|S'_n|^2) + \limsup_{n\to\infty}\frac{1}{n\epsilon}\mathbb{E}(|S_n - S'_n|)\\
&\le \limsup_{n\to\infty}\frac{1}{n\epsilon^2}\mathbb{E}(|X'_{n,1}|^2) + \limsup_{n\to\infty}\frac{2}{\epsilon}\mathbb{E}[|X_{n,i}|\mathbf{1}(|X_{n,i}| \ge a_n)]\\
&\le \limsup_{n\to\infty}\frac{a_n}{n\epsilon^2}\mathbb{E}|X'_{n,1}| + \limsup_{n\to\infty}\frac{2}{a_n^\delta\epsilon}\mathbb{E}(|X_{n,i}|^{1+\delta}) \le \eta\sup_n\mathbb{E}|X_{n,1}|,
\end{aligned}$$

which implies the lemma since η is arbitrarily chosen. ♢

Lemma 2. *Under conditions of Theorem 1, we have* $n\|\mathbb{E}(\xi_{n,1}|\mathbf{X}_0)\|^2 = \mathcal{O}(b_n)$

Proof of Lemma 2. Let the support of K be contained in the finite interval $[-M, M]$. Since g is countinous at $\mathbf{x}$, there exists a $\delta_0 > 0$ such that $C_g := \sup\{|g(\mathbf{y})| : |\mathbf{y} - \mathbf{x}| \le \delta_0\} < \infty$. Observe that

$$\mathbb{E}[K_{b_n}(x_{-1} - X_0)|\mathbf{X}_{-1}] = \int_{\mathbb{R}} K(u)f(x_{-1} - b_n u|\mathbf{X}_{-1})du.$$

By (8), $\mathbb{E}[K_{b_n}(x_{-1}-X_0)|\mathbf{X}_{-1}] \le f_* < \infty$ with probability 1. For $b_n \le \delta_0/M$,

$$\begin{aligned} n\|\mathbb{E}(\xi_{n,1}|\mathcal{G}_0)\|^2 &= b_n^p \left\| \mathbb{E}\left[g(X_0,\ldots,X_{1-p}) \prod_{i=1}^p K_{b_n}(x_{-i}-X_{1-i}) \middle| \mathbf{X}_{-1} \right] \right\|^2 \\ &\le C_g^2 b_n^p \left\| \mathbb{E}[K_{b_n}(x_{-1}-X_0)|\mathbf{X}_{-1}] \prod_{i=2}^p K_{b_n}(x_{-i}-X_{1-i}) \right\|^2 \\ &= \mathcal{O}(b_n^p) \int_{\mathbb{R}^{p-1}} \prod_{i=2}^p K_{b_n}^2(x_{-i}-u_i) f(u_2,\ldots,u_p) du_2 \ldots du_p = \mathcal{O}(b_n), \end{aligned}$$

which ensures the lemma by (8) and the change of variable $v_i = (x_{-i}-u_i)/b_n$. ◊

Lemma 3. *For* $\mathbf{y} \in \mathbb{R}^p$ *let* $H_n(\mathbf{y}) = \sum_{t=1}^n f(\mathbf{y}|\mathbf{X}_t) - nf(\mathbf{y})$. *(a) Relation (9) implies*

$$\sup_{\mathbf{y}} \|H_n(\mathbf{y})\|^2 = \mathcal{O}(n). \tag{27}$$

(b). Relation (13) implies

$$\mathbb{E}\left[\sup_{\mathbf{y}} H_n^2(\mathbf{y})\right] = \mathcal{O}(n) \tag{28}$$

Proof of Lemma 3. (a) The argument in Wu and Mielniczuk (2002) is applicable here. Let $C = \sup_{\mathbf{y}} \sum_{t=1}^\infty \|\mathcal{P}_0 f(\mathbf{y}|\mathbf{X}_t)\| < \infty$. Notice that $\|\mathcal{P}_k H_n(\mathbf{y})\| \le \sum_{t=1}^n \|\mathcal{P}_k f(\mathbf{y}|\mathbf{X}_t)\|$ and $\|\mathcal{P}_k f(\mathbf{y}|\mathbf{X}_t)\| = 0$ if $k > t$,

$$\begin{aligned} \|H_n(\mathbf{y})\|^2 &= \sum_{k=-\infty}^n \|\mathcal{P}_k H_n(\mathbf{y})\|^2 \\ &\le \sum_{k=-\infty}^0 \left[\sum_{t=1}^n \|\mathcal{P}_k f(\mathbf{y}|\mathbf{X}_t)\|\right]^2 + \sum_{k=1}^n \left[\sum_{t=k}^n \|\mathcal{P}_k f(\mathbf{y}|\mathbf{X}_t)\|\right]^2 \\ &\le \sum_{k=-\infty}^0 C \sum_{t=1}^n \|\mathcal{P}_k f(\mathbf{y}|\mathbf{X}_t)\| + nC^2 \le 2nC^2 \end{aligned} \tag{29}$$

For a univariate, differentiable function $L(\cdot)$, by Lemma 4 in Wu (2003), we have

$$\sup_{x\in\mathbb{R}} L^2(x) \le 2\int_{\mathbb{R}} [L^2(x) + |L'(x)|^2] dx. \tag{30}$$

Iterating (30), one obtains the multivariate version

$$\sup_{\mathbf{u}\in\mathbb{R}^p} L^2(\mathbf{u}) \le 2^p \sum_{\alpha_1=0}^1 \ldots \sum_{\alpha_p=0}^1 \int_{\mathbb{R}^p} |L^{(\alpha)}(\mathbf{u})|^2 d\mathbf{u}. \tag{31}$$

To see this, without loss of generality let $p = 2$. Then (30) implies that

$$\begin{aligned}
\sup_{u_1,u_2\in\mathbb{R}} L^2(u_1,u_2) &\le 2\sup_{u_1\in\mathbb{R}}\int_{\mathbb{R}}[L^2(u_1,u_2)+|L^{(0,1)}(u_1,u_2)|^2]du_2 \\
&\le 2\int_{\mathbb{R}}\left\{\sup_{u_1\in\mathbb{R}} L^2(u_1,u_2)+\sup_{u_1\in\mathbb{R}}|L^{(0,1)}(u_1,u_2)|^2\right\}du_2 \\
&\le 2\int_{\mathbb{R}}\left\{2\int_{\mathbb{R}}[L^2(u_1,u_2)+|L^{(1,0)}(u_1,u_2)|^2]du_1\right. \\
&\quad \left. +2\int_{\mathbb{R}}[|L^{(0,1)}(u_1,u_2)|^2+L^{(1,1)}(u_1,u_2)|^2]du_1\right\}du_2,
\end{aligned}$$

which is equal to the right hand side of (31) with $p = 2$. To obtain (28), we shall apply (31) with $L(\cdot) = H_n(\cdot)$. For $t \ge 0$ let $\lambda_{t,\alpha} = \int_{\mathbb{R}^p}\|\mathcal{P}_0 f^{(\alpha)}(\mathbf{u}|\mathbf{X}_t)\|^2 d\mathbf{u}$. As (29), by Cauchy's inequality,

$$\begin{aligned}
\int_{\mathbb{R}^p}\|H_n^{(\alpha)}(\mathbf{u})\|^2 d\mathbf{u} &= \sum_{k=-\infty}^{n}\int_{\mathbb{R}^p}\|\mathcal{P}_k H_n^{(\alpha)}(\mathbf{u})\|^2 d\mathbf{u} \\
&\le \sum_{k=-\infty}^{n}\int_{\mathbb{R}^p}\left[\sum_{j=\max(1,k)}^{n}\|\mathcal{P}_k f^{(\alpha)}(\mathbf{u}|\mathbf{X}_j)\|\right]^2 d\mathbf{u} \\
&\le \sum_{k=-\infty}^{n}\int_{\mathbb{R}^p}\left[\sum_{j=\max(1,k)}^{n}\frac{\|\mathcal{P}_k f^{(\alpha)}(\mathbf{u}|\mathbf{X}_j)\|^2}{\lambda_{j-k,\alpha}^{1/2}}\right]\left[\sum_{j=\max(1,k)}^{n}\lambda_{j-k,\alpha}^{1/2}\right] d\mathbf{u} \\
&\le \sum_{k=-\infty}^{n}\left[\sum_{j=\max(1,k)}^{n}\lambda_{j-k,\alpha}^{1/2}\right]^2 = \mathcal{O}(n)
\end{aligned}$$

in view of (13). Thus (28) follows from (31). ♢

Lemma 4. *(a) Let $V_2(\cdot)$ and $f(\cdot|\mathbf{X}_0)$ be continuous at $\mathbf{x}$. Assume (8). Then*

$$\sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-p}) \xrightarrow{\mathcal{P}} V(\mathbf{x})f(\mathbf{x})\kappa_2^p. \tag{32}$$

(b) Assume (9). Then

$$\left|\sum_{t=1}^{n}[\mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-p-1}) - \mathbb{E}(\xi_{n,t})]\right| = \mathcal{O}(\sqrt{b_n^p}). \tag{33}$$

Proof of Lemma 4 (a) Observe that

$$\sum_{t=1}^{n} \mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-p}) = \frac{1}{n}\sum_{t=1}^{n}\int_{\mathbb{R}^p} V(\mathbf{x}-b_n\mathbf{u})K^2(\mathbf{u})f(\mathbf{x}-b_n\mathbf{u}|\mathbf{X}_{t-p-1})d\mathbf{u},$$

and by the ergodic theorem, $n^{-1}\sum_{t=1}^{n} f(\mathbf{x}|\mathbf{X}_{t-p-1}) \to f(\mathbf{x})$ almost surely. Since $b_n \to 0$, by another application of the ergodic theorem,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\int_{\mathbb{R}^p} K^2(\mathbf{u})|V(\mathbf{x}-b_n\mathbf{u})f(\mathbf{x}-b_n\mathbf{u}|\mathbf{X}_{t-p-1}) - V(\mathbf{x})f(\mathbf{x}|\mathbf{X}_{t-p-1})|d\mathbf{u}\\
&\le \frac{\kappa}{n}\sum_{t=1}^{n}\sup_{|\mathbf{y}-\mathbf{x}|\le\delta}|V(\mathbf{y})f(\mathbf{y}|\mathbf{X}_{t-p-1}) - V(\mathbf{x})f(\mathbf{x}|\mathbf{X}_{t-p-1})|\\
&\to \kappa\mathbb{E}\left[\sup_{|\mathbf{y}-\mathbf{x}|\le\delta}|V(\mathbf{y})f(\mathbf{y}|\mathbf{X}_0) - V(\mathbf{x})f(\mathbf{x}|\mathbf{X}_0)|\right]
\end{aligned}$$

holds for any $\delta > 0$. Note that $V(\cdot)$ and $f(\cdot|\mathbf{X}_0)$ are continuous at $\mathbf{x}$ and $f(\cdot|\mathbf{X}_0)$ is bounded. By the Lebesgue dominated convergence theorem we get,

$$\begin{aligned}
&\lim_{\delta\downarrow 0}\mathbb{E}\left[\sup_{|\mathbf{y}-\mathbf{x}|\le\delta}|V(\mathbf{y})f(\mathbf{y}|\mathbf{X}_0) - V(\mathbf{x})f(\mathbf{x}|\mathbf{X}_0)|\right]\\
&\le \lim_{\delta\downarrow 0}\mathbb{E}\left[\sup_{|\mathbf{y}-\mathbf{x}|\le\delta}|V(\mathbf{y}) - V(\mathbf{x})|f(\mathbf{x}|\mathbf{X}_0)\right] + \lim_{\delta\downarrow 0}\mathbb{E}\left[\sup_{|\mathbf{y}-\mathbf{x}|\le\delta}V(\mathbf{x})|f(\mathbf{y}|\mathbf{X}_0) - f(\mathbf{x}|\mathbf{X}_0)|\right]\\
&\le V(\mathbf{x})\lim_{\delta\downarrow 0}\mathbb{E}\left[\sup_{|\mathbf{y}-\mathbf{x}|\le\delta}|f(\mathbf{y}|\mathbf{X}_0) - f(\mathbf{x}|\mathbf{X}_0)|\right] = 0,
\end{aligned}$$

which guarantees (32).

(b) By (a) of Lemma 3 and Cauchy's inequality we have that

$$\begin{aligned}
\mathbb{E}\left|\sum_{t=1}^{n}[\mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-p-1}) - \mathbb{E}(\xi_{n,t})]\right| &\le \frac{\sqrt{b_n^p}}{\sqrt{n}}\int_{\mathbb{R}^p} K(\mathbf{u})|g(\mathbf{x}-b_n\mathbf{u})|\mathbb{E}|H_n(\mathbf{x}-b_n\mathbf{u})|d\mathbf{u}\\
&= \mathcal{O}(\sqrt{b_n^p}).
\end{aligned}$$

◇

Proof of Theorem 1. Let $\mathcal{G}_n = (\ldots,\theta_{n-1},\theta_n;\ldots,\varepsilon_{n-2},\varepsilon_{n-1})$. Then $\xi_{n,t}$ is $\mathcal{G}_n$-measurable. Clearly (10) follows from

$$\sum_{t=1}^{n}[\xi_{n,t} - \mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1})] \Rightarrow N[0,\, V(\mathbf{x})f(\mathbf{x})\kappa^p] \tag{34}$$

and

$$\sum_{t=1}^{n}[\mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1}) - \mathbb{E}\xi_{n,t}] \xrightarrow{\mathcal{P}} 0. \tag{35}$$

For (34), observe that the summands form (triangular) stationary martingale differences; therefore, we can apply the martingale central limit theorem, since

$$\sum_{t=1}^{n}\mathbb{E}\{[\xi_{n,t} - \mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1})]^2|\mathcal{G}_{t-1})\} = \sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-1}) - \sum_{t=1}^{n}\mathbb{E}^2(\xi_{n,t}|\mathcal{G}_{t-1}),$$

By Lemma 2, the latter sum tends to 0 in probability. For the former one, let $v_{n,t} = n\mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-1})$. Since V is continuous at $\mathbf{x}$ and K has bounded support, we have

$$\begin{aligned}
\mathbb{E}(v_{n,t}) &= \mathbb{E}\left[Y_t^2\prod_{i=1}^{p}K_{b_n}^2(x_{-i} - X_{t-i})\right] \\
&= \mathbb{E}\left[V(X_{t-1},\dots,X_{t-p})K_{b_n}^2(x_{-i} - X_{t-i})\right] \\
&= \int_{\mathbb{R}^p}V(\mathbf{x} - b_n\mathbf{u})K^2(\mathbf{u})f(\mathbf{x} - b_n\mathbf{u})d\mathbf{u} \to V(\mathbf{x})f(\mathbf{x})\kappa^p
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}(v_{n,t}^{1+\delta}) &= \mathbb{E}\left[|Y_t|^{2(1+\delta)}\prod_{i=1}^{p}K_{b_n}^{1+\delta}(x_{-i} - X_{t-i})\right] \\
&= b_n^{-\delta}\int_{\mathbb{R}^p}\mathbb{E}[|G(\mathbf{x} - b_n\mathbf{u},\theta_t)|^{2(1+\delta)}]K^{2(1+\delta)}(\mathbf{u})f(\mathbf{x} - b_n\mathbf{u})d\mathbf{u} = \mathcal{O}(n^\delta).
\end{aligned}$$

Hence Lemma 1 is applicable and by (a) of Lemma 4, the convergence of the conditional variance $\sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}^2|\mathcal{G}_{t-1}) \xrightarrow{\mathcal{P}} V(\mathbf{x})f(\mathbf{x})\kappa^p$ holds. The Lindeberg condition follows from

$$\begin{aligned}
n\mathbb{E}(|\xi_{n,1}|^{2+\delta}) &= n(b_n^p/n)^{(2+\delta)/2}\mathbb{E}\left[\prod_{i=1}^{p}K_{b_n}^{2+\delta}(x_{-i} - X_{t-i})\right] \\
&= n(b_n^p/n)^{(2+\delta)/2}b_n^p b_n^{-p(2+\delta)} = (nb_n^p)^{-\delta/2} \to 0.
\end{aligned}$$

Now we shall prove (35). Observe that by Lemma 2, for $i = 0,\dots,p-1$,

$$\begin{aligned}
\left\|\sum_{t=1}^{\lfloor n/p\rfloor}[\mathbb{E}(\xi_{n,tp+i}|\mathcal{G}_{tp-1+i}) - \mathbb{E}(\xi_{n,tp+i}|\mathcal{G}_{tp-p-1+i})]\right\|^2 &= \lfloor n/p\rfloor\|\mathbb{E}(\xi_{n,0}|\mathcal{G}_{-1}) - \mathbb{E}(\xi_{n,0}|\mathcal{G}_{-p-1})\|^2 \\
&\leq \lfloor n/p\rfloor\|\mathbb{E}(\xi_{n,0}|\mathcal{G}_{-1})\|^2 = \mathcal{O}(b_n),
\end{aligned}$$

which by summing over $i = 0, \dots, p-1$ implies that

$$\sum_{t=1}^{n}[\mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-1}) - \mathbb{E}(\xi_{n,t}|\mathcal{G}_{t-p-1})] \xrightarrow{\mathcal{P}} 0.$$

Thus (b) of Lemma 4 implies (35). $\diamondsuit$

Lemma 5. *Assume (8). Let* $M_n(\mathbf{y}) = \sum_{t=1}^{n}\{\zeta_{n,t}(\mathbf{y}) - \mathbb{E}[\zeta_{n,t}(\mathbf{y})|\mathbf{X}_{t-p}]\}$. *(a) For all* $\mathbf{y} \in \mathbb{R}^p$ *and* $a > 0$,

$$\mathbb{P}[|M_n(\mathbf{y})| > pa] \leq 2p\exp[-a^2/(2K_0 a/\sqrt{nb_n^p} + K_0 q_*)]. \tag{36}$$

(b). Let K *be Lipschitz continuous with index* $\eta > 0$ *and* $\mathbb{E}(|X_1|^\alpha) < \infty$ *for some* $\alpha > 0$. *Then for all* $\varsigma > 0$, *there exists* $C > 0$ *such that*

$$\mathbb{P}\left[\sup_{\mathbf{y}\in\mathbb{R}^p}|M_n(\mathbf{y})| > C\log n\right] = \mathcal{O}(n^{-\varsigma}). \tag{37}$$

In addition, $\log n = o(nb_n^p)$, *then (37) holds with* $\log n$ *replaced by* $\sqrt{\log n}$.

Proof of Lemma 5. (a) Let $m = \lfloor n/p \rfloor$ and observe that for all $0 \leq i \leq p-1$,

$$\begin{aligned} T_n(\mathbf{y}) &:= \sum_{j=0}^{m-1}\{\mathbb{E}[\zeta^2_{n,i+jp}(\mathbf{y})|\mathbf{X}_{i+jp-p}] - \mathbb{E}^2[\zeta_{n,i+jp}(\mathbf{y})|\mathbf{X}_{i+jp-p}]\} \\ &\leq \frac{1}{n}\sum_{j=0}^{m-1}\int_{\mathbb{R}^p} K^2(\mathbf{u})f(\mathbf{y} - b_n\mathbf{u}|\mathbf{X}_{i+jp-p})d\mathbf{u} \leq K_0 q_*. \end{aligned}$$

Clearly, with probability 1, $|\zeta_{n,t}(\mathbf{y}) - \mathbb{E}[\zeta_{n,t}(\mathbf{y})|\mathbf{X}_{t-p}]| \leq K_0/\sqrt{nb_n^p}$. By Freedman's inequality (cf. Freedman 1975),

$$\mathbb{P}\left[\left|\sum_{j=0}^{m-1}\{\zeta_{n,i+jp}(\mathbf{y}) - \mathbb{E}[\zeta_{n,i+jp}(\mathbf{y})|\mathbf{X}_{i+jp-p}]\}\right| \geq a\right] \leq 2\exp[-a^2/(2K_0 a/\sqrt{nb_n^p} + K_0 q_*)],$$

which trivially implies (36) by summing over $i = 0, \dots, p-1$. Here we employ the martingale inequality to estimate tail probabilities. In comparison, exponential inequalities for strong mixing processes are used to obtain similar asymptotic properties; see for example Bosq (1998, p. 27).

(b) Without loss of generality, assume that $p = 1$, $\mathbb{E}(|X_1|^\alpha) < \infty$ for some $\alpha > 0$ and that K is Lipschitz continuous with index 1. Then by the Markov inequality,

$$\mathbb{E}\left[\sup_{|x|\geq n^A}|M_n(x)|\right] \leq n\mathbb{E}\left[\sup_{|x|\geq n^A}|\zeta_{n,1}(x) - \mathbb{E}[\zeta_{n,1}(x)|\mathbf{X}_0]|\right]$$

$$
\begin{aligned}
&\leq 2n\mathbb{E}\left[\sup_{|x|\geq n^A}|\zeta_{n,1}(x)|\right] \\
&\leq \frac{2n}{\sqrt{nb_n}}K_0\mathbb{P}(|X_1|\geq n^A - Mb_n) \\
&\leq \frac{2n}{\sqrt{nb_n}}K_0(n^A - Mb_n)^{-\alpha}\mathbb{E}(|X_1|^\alpha) = \mathcal{O}(n^{-\varsigma}).
\end{aligned}
$$

Next we consider the behavior of $M_n(x)$ when $|x| < n^A$. Let $a = 2C\log n$. By (a), for all sufficiently large $C > 0$,

$$\mathbb{P}\{|M_n(x)| \geq 2C\log n\} = \mathcal{O}(n^{-C}).$$

Choose $C > A + 3 + \varsigma$. Then

$$\mathbb{P}\left\{\sup_{i=0,1,\ldots,2n^{A+3}}|M_n(-n^A + in^{-3})| \geq a\right\} = n^{A+3}\mathcal{O}(n^{-C}) = \mathcal{O}(n^{-\varsigma}),$$

which implies $\mathbb{P}[\sup_{|x|<n^A}|M_n(x)| \geq a] = \mathcal{O}(n^{-\varsigma})$ in view of

$$|M_n(x) - M_n(y)| = \mathcal{O}[n|x-y|/(b_n\sqrt{nb_n})]$$

since K is Lipschitz continuous with index 1. Thus (37) holds. If $\log n = o(nb_n)$, we can choose $a = 2C\sqrt{\log n}$ in (a) and the upper bound $\mathcal{O}(\sqrt{\log n})$ similarly follows. $\diamondsuit$

Remark 1. ¿From the proof of Lemma 5, it is easily seen that the Lipschitz continuity of K can be relaxed to piecewise Lipschitz continuity with finitely many pieces.

Proof of Theorem 2. Let $N_n(\mathbf{y}) = \sum_{t=1}^n\{\mathbb{E}[\zeta_{n,t}(\mathbf{y})|\mathbf{X}_{t-p}] - \mathbb{E}\zeta_{n,t}(\mathbf{y})\}$. By (b) of Lemma 3,

$$\mathbb{E}\left[\sup_{\mathbf{y}}|N_n(\mathbf{y})|^2\right] = \mathbb{E}\left[\sup_{\mathbf{y}}\frac{b_n^p}{\sqrt{nb_n^p}}\int_{\mathbb{R}}K(\mathbf{u})|H_n(\mathbf{y}-b_n\mathbf{u})|d\mathbf{u}\right]^2 = \mathcal{O}(b_n^p).$$

Thus by (b) of Lemma 5, (14) holds since $\sqrt{nb_n^p}[f_n(\mathbf{y}) - \mathbb{E}[f_n(\mathbf{y})]] = M_n(\mathbf{y}) + N_n(\mathbf{y})$. $\diamondsuit$

Acknowledgments. The author thanks Jan Mielniczuk and George Michailidis for many helpful suggestions.

REFERENCES

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

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

Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation. *Stochastic Process. Appl.* **21** 179–193.

Chan, K. S. and Tong, H., 2001. Chaos: a statistical perspective. Springer, New York.

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

Devroye, L. and Györfi, L. (1984). *Nonparametric density estimation: the L^1 view.* New York: Wiley

Diaconis, P. and Freedman. D., 1999. Iterated random functions. SIAM Rev. 41, 41–76.

Doukhan, P. (1994). *Mixing: properties and examples.* Springer, New York.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 50, 987–1007.

Freedman, David A. (1975). On tail probabilities for martingales *Ann. Probab.* **3** 100–118

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

Gyorfi, L., Hardle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric curve estimation from time series.* Springer, Berlin.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68** 189–196.

Hardle W., Lutkepohl H. and Chen R. (1997). A review of nonparametric time series analysis *Int. Stat. Rev.* **65** 49–72

Hart, J. D. (1996). Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics* **6** 15-142.

Neumann, Michael H (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.* **26** 2014–2048.

Neumann, Michael H. and Kreiss, Jens-Peter. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26** 1570–1613.

Nicholls, D. F. and Quinn, B. G., 1982. Random coefficient autoregressive models: an introduction. Springer, New York.

Pham, T. D. and Tran, L. T. (1985). Some mixing properties of time series models.

Stochastic Process. Appl. **19** 297–303.

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

PRIESTLEY, M. B. (1988). *Non-linear and non-stationary time series analysis.* Academic Press, New York.

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

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

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

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

SINGH, R. S. and ULLAH, A. (1985). Nonparametric time-series estimation of joint DGP, conditional DGP, and vector autoregression *Econometric Theory* **1** 27–52

STUTE, W.. (1982). A Law of the Logarithm for Kernel Density Estimators. *Ann. of Probab.* **10** 414–422.

TONG, H., (1990). *Non-linear time series: a dynamical system approach.* Oxford University Press.

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

WU, W. B. (2003). Empirical processes of long-memory sequences. To appear, *Bernoulli.*

WU, W. B. and J. MIELNICZUK (2002). Kernel density estimation for linear processes. *Ann. of Statist.* **30**, 1441–1459.

WU, W. B. and WOODROOFE, M., (2000). A central limit theorem for iterated random functions. *J. Appl. Probab.* **37**, 748–755.

YU, B., (1993). Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Ann. Statist.* **21**, 711–735

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