Chin. Ann. of Math.
20B: 2(1999),173-184.

NONPARAMETRIC IDENTIFICATION FOR NONLINEAR AUTOREGRESSIVE TIME SERIES MODELS: CONVERGENCE RATES**

LU ZUDI* CHENG PING*

Abstract

In this paper, the optimal convergence rates of estimators based on kernel approach for nonlinear AR model are investigated in the sense of Stone[17,18]. By combining the α–mixing property of the stationary solution with the characteristics of the model itself, the restrictive conditions in the literature which are not easy to be satisfied by the nonlinear AR model are removed, and the mild conditions are obtained to guarantee the optimal rates of the estimator of autoregression function. In addition, the strongly consistent estimator of the variance of white noise is also constructed.

Keywords Nonlinear AR model, Optimal convergence rates, Kernel approach, Autoregression function, Variance of white noise, Consistency

1991 MR Subject Classification 62G07

Chinese Library Classification O212.7

§1. Introduction

Consider a nonlinear autoregressive (AR) model in the form

$$X_t = f(X_{t-1}, \cdots, X_{t-p}) + \varepsilon_t, \tag{1.1}$$

where $f : R^p \to R^1$ is an unknown Borel function on R^p and $\{\varepsilon_t\}$ is an i.i.d. white noise with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma^2 < \infty$ and ε_t independent of $\{X_s,\ s < t\}$. Concerning nonparametric approaches for identification of this kind of model, it has received increasing attention in the literature. The reader is referred to Tjøstheim[20], Härdle and Chen[9] and Tong[22] for surveys. When p is unknown, it needs to be estimated; see, for example, Cheng and Tong[6] and Tjϕstheim and Auestad[21] for its estimation. Here we assume for simplicity that p is known, and will focus on considering the estimates of the autoregression function f and the innovation (white noise) varience σ^2. In this paper, the pointwise optimal convergence rates of weak consistency and uniform optimal rates of strong consistency in sense of Stone[17,18] will be investigated for the estimator of the autoregression function, and by the way, the estimator of the variance of white noise will also be constructed and proved to be strongly consistent. The asymptotic normality of these estimators will be explored in another paper.

Manuscript received December 15, 1997. Revised September 3, 1998.

*Institute of Systems Science, Academia Sinica, Beijing 100080, China.

**Project supported by the National Natural Science Foundation of China.

Before prodeeding, we remark that, to our knowledge, Truong and Stone[23] and Masry and Tjøstheim[15] investigated the optimal convergence rates of nonparametric kernel estimators of the regression function for the stationary and dependent samples under α-mixing. Indeed, under some mild conditions, a lot of nonlinear time series models have a stationary solution which is α-mixing[1,13], so the α-mixing assumption on the dependent samples is acceptable. However, in the publications just cited, the assumptions imposed on the distribution of the stationary sequence and some others are very restrictive. For example, Condition 3 of Truong and Stone[23] is not satisfied by the model (1.1) when the order p is greater than 1 (see Remark 2.0 below), and also, in Lu and Cheng[14], we listed an example showing that the basic Assumption 3.1 in [15] is not satisfied even when ε_t is normally distributed and f is Lipschitz continuous. Just due to those restrictive assumptions, the results in the publications cannot be applied well to a lot of nonlinear time series models. The idea of this paper is, by combining the α–mixing property of the stationary solution with the characteristics of the model itself, to investigate the large sample properties of the estimators. Thus, we will remove a lot of the restrictive conditions in references, and get the asymptotic theory under mild conditions which can be applied well to the nonparametric identification of the nonlinear autoregressive models.

§2. Main Results

From [1, 2, 4, 19], it follows that the (geometrically) ergodic and hence α–mixing stationary solution to model (1.1) exists. For generality, we assume the following assumption:

A1. The α–mixing stationary solution to model (1.1) exists. Denote the mixing coefficients by $\alpha(\cdot)$.

Let $X_1, X_2, \cdots, X_n$ be a realization of size n from the stationary solution of the model (1.1). Set $Y_t = (X_{t-1}, X_{t-2}, \cdots, X_{t-p})'$, and write π for its distribution. The kernel estimator of $f(y)$ is defined by

$$\hat{f}_n(y) = \sum_{i=p+1}^{n} X_i K\Big(\frac{y-Y_i}{h_n}\Big) \Big/ \sum_{j=p+1}^{n} K\Big(\frac{y-Y_j}{h_n}\Big), \tag{2.1}$$

where the "n" over the sum sign in the denominator can be replaced by "$n+1$" in application (it is noted by the definition of Y_t that this is suitable), but the asymptotics is not affected and hence we adopt (2.1) for simplicity. In order to remove the effects of the extrem values (see [20]), we adopt the weighted estimator of σ^2 defined as follows

$$\widehat{\sigma^2} = \sum_{t=p+1}^{n} (X_t - \hat{f}_n(Y_t))^2 w(Y_t) \Big/ \sum_{t=p+1}^{n} w(Y_t), \tag{2.2}$$

where $w(\cdot) : R^p \to R^1$ is a non-negative and bounded Borel measurable weight function with its compact support denoted by $S(w)$. Throughout the paper, it is defined that $0/0 = 0$.

We first state some basic assumptions:

A2. The non-negative Borel measurable kernel function $K(\cdot)$ satisfies

(1) $$a_1 I_{\{\|u\| \le R\}} \le K(u) \le a_2 I_{\{\|u\| \le R\}}, \tag{2.3}$$

where a_1, a_2 are two positive numbers with $a_1 < a_2$, and R is a positive number large enough;

(2) $K(x)$ is Lipschitz continuous of order 1 on its support.

A3. The autoregression function f is Lipschitz continuous of order r $(0 < r \leq 1)$ on R^p.

Let U be a non-empty open subset of R^p containing the origin point with $U \subset S(K)$, and $D \subset U$ a nonempty compact subset of R^p.

A4. The stationary sequence $\{Y_t\}$ satisfies that:

(1) the distribution of Y_t is absolutely continuous with respect to the Lebesgue measure, and its density function $\psi_Y(\cdot)$ is continuous over U such that, for some $M_1 > 0$,

$$0 < \psi_Y(y) \leq M_1, \quad \forall y \in U; \tag{2.4}$$

(2) there exists a constant $M_2 > 0$ such that

$$\psi_Y(y) \geq M_2 \text{ for all } y \in D; \tag{2.5}$$

(3) for any $j \geq 1$, given $Y_0 = y$, the conditional density $\psi_j(\cdot|y)$ of Y_j exists and satisfies that for some positive number M_3,

$$\psi_j(y'|y) \leq M_3, \quad \forall y,\, y' \in U, \quad j \geq 1. \tag{2.6}$$

The preceding assumptions will be used in this paper. The requirement of A2(1) on the kernel function is adopted by a lot of publications (e.g., [24, 7]). Although its support $S(K)$ is bounded, K may still be taken as an approximation to the kernel with unbounded support when R is large enough. The smoothness conditions of A2(2) and A3 on the kernel and the autoregression function are usually needed to study the convergence rates (e.g., [5]). The assumption A4 on the distribution of $\{Y_t\}$ is similar to Conditions 2 and 3 in [23], but weaker than their Condition 3 (which also requires that $\psi_j(y'|y) \geq M_3^{-1}, \quad \forall y,\, y' \in U, \quad j \geq 1$). It follows from the lemma below that the assumption A4 stated above is easily satisfied. However, Condition 3 of [23] is not easy to hold for model (1.1).

Lemma 2.1. *Suppose that f is a continuous function on R^p and that the density function $\psi_\varepsilon(t)$ of ε_t exists and is positive, bounded and continuous over R^1. Then $Y_t = (X_{t-1}, X_{t-2}, \cdots, X_{t-p})'$ defined by the stationary solution of the model* (1.1) *satisfies the assumption* A4.

Proof. It easily follows from (1.1) that

$$\psi_j(x|y) = \prod_{i=1}^{j} q_{ji}(x,y) \prod_{i=1}^{p-j} I_{(y_i = x_{i+j})},\ j = 1, \quad \cdots, p, \tag{2.7}$$

where $q_{ji}(x,y) = \psi_\varepsilon(x_i - f(x_{i+1}, \cdots, x_j, y_1, \ldots, y_{p-j+i}))$ for $i = 1, \cdots, j$. Then A4(1) can be deduced from the boundedness and the positiveness of $\psi_\varepsilon(t)$ and the equality

$$\psi_Y(y) = \int \psi_p(y|u)\pi(du). \tag{2.8}$$

Let B be a compact subset of R^p such that $\pi(B) > 0$. Since $\psi_\varepsilon(t)$ is positive and continuous over R^1 and f is continuous on R^p, it is easily known that for some $M_4 > 0$, $q_{ji}(y, u) > M_4$ for any $y \in D$, $u \in B$ and any $1 \leq j \leq p$ and $i = 1, \cdots, j$, and hence there exists an $M_5 > 0$ such that $\psi_p(y|u) > M_5$ for any $y \in D$, $u \in B$. Thus A4(2) easily follows from (2.8).

For $1 \leq j \leq p$, A4(3) is obvious by (2.7); if $j \geq p$, A4(3) can be obtained recursively.

Remark 2.1. The conditions on ε_t in Lemma 2.1 are easily satisfied (e.g., ε_t is Gaussian), and are usually imposed to guarantee that the model is (geometrically) ergodic (see [11]), the PL assumption in [11]. Also, by (2.7), it is easily known that if $p > 1$, Condition 3 of [23] is not satisfied.

Theorem 2.1. *Assume that* A1, A2, A3, A4(1), A4(3) *hold. If* $E\varepsilon_t^2 < \infty$, *and the mixing coefficient* $\alpha(\cdot)$ *and the bandwidth* h_n *also satisfy the conditions of Theorem* 1 *or Theorem* 2 *of* [12], *then for any* $y \in U$,

$$|\hat{f}_n(y) - f(y)| = O_P(h_n^r) + O_P\Big(\Big(\frac{1}{nh_n^p}\Big)^{1/2}\Big). \tag{2.9}$$

Furthermore, if $h_n = (\frac{1}{n})^{1/(2r+p)}$, *and the conditions above hold, then*

$$|\hat{f}_n(y) - f(y)| = O_P\Big(\Big(\frac{1}{n}\Big)^{r/(2r+p)}\Big). \tag{2.10}$$

Theorem 2.2. *Assume that* A1, A2, A3, A4 *hold. If* $E|\varepsilon_t|^m < \infty$ *for some* $m > 2$, *and the mixing coefficient* $\alpha(\cdot)$ *and the bandwidth* h_n *also satisfy*

$$\sum_{j=1}^{\infty} j^{a-1}\alpha(j) < \infty, \quad h_n = \Big(\frac{\ln n}{n}\Big)^{\theta},$$

where $0 < \theta < \frac{m-2}{mp}$, $a > \frac{(1+p\theta)[2(p+2)(m-1)+m]}{2m(1-p\theta)-4}$, *then as* $n \to \infty$,

$$\sup_{y\in D} |\hat{f}_n(y) - f(y)| = O(h_n^r) + O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \text{ a.s.} \tag{2.11}$$

Furthermore, if $m > 2 + p/r$, *let* $h_n = (\frac{\ln n}{n})^{1/(2r+p)}$, *then*

$$\sup_{y\in D} |\hat{f}_n(y) - f(y)| = O\Big(\Big(\frac{\ln n}{n}\Big)^{r/(2r+p)}\Big), \text{ a.s.} \tag{2.12}$$

Theorem 2.3. *Under the conditions of Theorem* 2.2, *if* $S(w) \subset D$, *then*

$$\widehat{\sigma^2} \overset{\text{a.s.}}{\to} \sigma^2. \tag{2.13}$$

Remark 2.2. Theorem 2.1 can not be obtained by Theorem 1 of [23] (in Theorem 1 of [23] there, it is needed that $n\sum\limits_{j=n}^{\infty}(\alpha(j))^{1-2/\nu} = O(1)$ for some $\nu > 2$; while, in Theorem 2.1 here, the mixing coefficient $\alpha(\cdot)$ admits to be unsummable[12]. Due to the facts pointed out in Section 1, Theorem 2.2 does not follow from [15]. Also, it must be noted that in Theorem 3 of [23], only the uniform weak convergence rate was obtained, but their conditions are very strong ($\alpha(j) = O(\rho^j)$, $0 < \rho < 1$ and $P(|Z_0| \le M|Y_0 = y) = 1, \quad y \in U$ for some $M > 0$).

Remark 2.3. According to Stone[17,18], the convergence rates in (2.10) and (2.12) under i.i.d. case are optimal.

§3. Proofs of Theorems

Set $K_i = K_{ni} = K_{ni}(y) = K(\frac{Y_i - y}{h_n})$. Let c be a generic positive constant which may differ at different places in the following.

Lemma 3.1. *Suppose that* A2(1) *and* A4 *hold. Then*

$$EK_iK_{j+i} = \begin{cases} O(h_n^{2p}) & \text{for } j > 0, \\ O(h_n^p) & \text{for } j = 0. \end{cases}$$

Proof. If $j = 0$, by A2(1), A4(1), we have

$$EK_i^2 = O(h_n^p).$$

If $j > 0$, then it follows from A2(1), A4(1) and A4(3) that

$$EK_iK_{j+i} = O(h_n^{2p}).$$

Lemma 3.2. *Let $K(u)$ and $g(x)$ be two Borel functions on R^p such that* (a) *K is bounded on R^p;* (b) *$\int_{R^p} |K(u)|du < \infty$;* (c) *$\lim\limits_{\|u\|\to\infty} \|u\|^p K(u) = 0$;* (d) *$\int |g(x)|dx < \infty$. Set $g_n(x) = h_n^{-p} \int K(\frac{u-x}{h_n})g(u)du$, where h_n is a sequence of positive constants with $h_n \to 0$ $(n \to \infty)$. If g is continuous at x, then*

$$\lim_{n\to\infty} g_n(x) = g(x) \int K(u)du.$$

Proof. See the proof of Theorem 1A of [16].

Lemma 3.3. *Let $X_1, \cdots, X_n$ be independent random variables satisfying $|X_i| \le M$, $EX_i = 0$ and $\mathrm{Var}(X_i) \le \sigma^2$ for all i and for some positive constants M and σ^2. Then, for $0 \le t \le 2/M$,*

$$E\Big[\exp\Big(t\sum_{i=1}^n X_i\Big)\Big] \le \exp\Big[nt^2\sigma^2\frac{(1+tM)}{2}\Big].$$

Proof. This is Lemma 6 of [8].

Lemma 3.4. *Under the conditions of Theorem* 1 *(or Theorem* 2*) of* [12]*, if $\psi_Y(y)$ is continuous, then*

$$\frac{1}{nh_n^p}\sum_{i=p+1}^n K\Big(\frac{Y_i - y}{h_n}\Big) \xrightarrow{P} \psi_Y(y)\int K(u)du \quad (n\to\infty).$$

Proof. The arguments are completely similar to those of Theorems 1 and 2 of [12] .

Proof of Theorem 2.1. Set

$$A_n(y) = \hat{f}_n(y) - f(y),$$

$$A_{n1}(y) = \frac{\sum\limits_{i=p+1}^n K_{ni}(f(Y_i) - f(y))}{\sum\limits_{i=p+1}^n K_{ni}},$$

$$A_{n2}(y) = \frac{\sum\limits_{i=p+1}^n K_{ni}\varepsilon_i}{\sum\limits_{i=p+1}^n K_{ni}}.$$

Then it is obvious that

$$A_n(y) = A_{n1}(y) + A_{n2}(y). \tag{3.1}$$

Now using A3 and then A2(1), we obtain the first term

$$|A_{n1}(y)| \le c\frac{\sum\limits_{i=p+1}^n \|Y_i - y\|^r I_{(\|Y_i-y\|\le Rh_n)}}{\sum\limits_{i=p+1}^n I_{(\|Y_i-y\|\le Rh_n)}} = O(h_n^r). \tag{3.2}$$

For the second term $A_{n2}(y)$, we first observe

$$P\Big(\Big|(nh_n^p)^{-1}\sum_{i=p+1}^n K_{ni}\varepsilon_i\Big|>\epsilon\Big)\le\frac{E\Big|\sum\limits_{i=p+1}^n K_{ni}\varepsilon_i\Big|^2}{(nh_n^p\epsilon)^2}=\frac{\sum\limits_{i=p+1}^n EK_{ni}^2\varepsilon_i^2}{(nh_n^p\epsilon)^2}$$
$$\le\frac{\sigma^2EK_{ni}^2}{n(h_n^p\epsilon)^2}=O\Big(\frac{1}{nh_n^p\epsilon^2}\Big),$$

from which together with Lemma 3.1 and A4(1) it follows that

$$\Big|(nh_n^p)^{-1}\sum_{i=p+1}^n K_{ni}\varepsilon_i\Big|=O_P((nh_n^p)^{-1/2}) \tag{3.3}$$

for any $y\in U$. Thus together with Lemma 3.4 and A4(1), we have

$$A_{n2}(y)=O_P((nh_n^p)^{-1/2}), \tag{3.4}$$

for any $y\in U$. Finally, by (3.1), (3.2), (3.4), the desired results can be obtained.

Proof of Theorem 2.2. In (3.2), in fact, it can be obtained that

$$\sup_{y\in D}|A_{n1}(y)|=\sup_{y\in D}\frac{\sum\limits_{i=p+1}^n\|Y_i-y\|^rI_{(\|Y_i-y\|\le Rh_n)}}{\sum\limits_{i=p+1}^n I_{(\|Y_i-y\|\le Rh_n)}}=O(h_n^r). \tag{3.5}$$

Set $B_{n1}(y)=((n-p)h_n^p)^{-1}\sum\limits_{i=p+1}^n K_{ni}\varepsilon_i$, $B_{n2}(y)=((n-p)h_n^p)^{-1}\sum\limits_{i=p+1}^n K_{ni}$. Then

$$A_{n2}(y)=B_{n1}(y)/B_{n2}(y). \tag{3.6}$$

Let $b_n=\big(\frac{n}{h_n^p\ln n}\big)^{1/(2(m-1))}$,

$$\varepsilon_i'\hat{=}\varepsilon_iI_{(|\varepsilon_i|\le b_n)},\quad \varepsilon_i''\hat{=}\varepsilon_iI_{(|\varepsilon_i|>b_n)}.$$

Obviously, $EB_{n1}(y)=0$.

$$\begin{aligned}B_{n1}(y)&=\frac{1}{(n-p)h_n^p}\sum_{i=p+1}^n(\varepsilon_i'-E\varepsilon_i')K_{ni}-\frac{1}{(n-p)h_n^p}\sum_{i=p+1}^n(E\varepsilon_i')(K_{ni}-EK_{ni})\\&\quad+\frac{1}{(n-p)h_n^p}\sum_{i=p+1}^n\varepsilon_i''K_{ni}-\frac{1}{(n-p)h_n^p}\sum_{i=p+1}^nE\varepsilon_i''K_{ni}\\&\hat{=}B_{n1}^{(1)}(y)+B_{n1}^{(2)}(y)+B_{n1}^{(3)}(y)+B_{n1}^{(4)}(y).\end{aligned} \tag{3.7}$$

Now we begin to treat $B_{n1}^{(i)}(y)$ respectively. First we deal with the last two terms of (3.7).

$$\begin{aligned}\sup_{y\in D}|B_{n1}^{(3)}(y)|&\le\frac{c}{(n-p)h_n^p}\sum_{i=p+1}^n|\varepsilon_i''|\\&\le\frac{c}{(n-p)h_n^p}\sum_{i=p+1}^n|\varepsilon_i|^mI_{(|\varepsilon_i|>b_n)}b_n^{1-m}\\&=O\Big(\frac{1}{b_n^{m-1}h_n^p}\Big)=O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big),\quad\text{a.s.}\end{aligned} \tag{3.8}$$

$$\begin{aligned}\sup_{y\in D}|B_{n1}^{(4)}(y)| &= \sup_{y\in D}|EB_{n1}^{(3)}(y)| \\ &\le \frac{c}{h_n^p}E|\varepsilon_i|^m I_{(|\varepsilon_i|>b_k)}b_n^{1-m} \\ &= O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big).\end{aligned} \tag{3.9}$$

Since D is a compact subset of R^p, it can be covered by a finite number N_n of cubes I_k with centers y_k whose sides are of length L_n: $L_n = c/N_n^{1/p} = c\big(\frac{\ln nh_n^{p+2}}{n}\big)^{1/2}$.

$$\begin{aligned}\sup_{y\in D}|B_{n1}^{(i)}(y)| \le& \max_{1\le k\le N_n}\sup_{y\in D\cap I_k}|B_{n1}^{(i)}(y)-B_{n1}^{(i)}(y_k)| \\ &+\max_{1\le k\le N_n}|B_{n1}^{(i)}(y_k)| \\ &\hat{=} B_{n1}^{(i')}+B_{n1}^{(i'')}, \quad i=1,\,2.\end{aligned} \tag{3.10}$$

Using Assumption A2(2), we have

$$\begin{aligned}B_{n1}^{(1')} &= \max_{1\le k\le N_n}\sup_{y\in D\cap I_k}\Big|\frac{1}{(n-p)h_n^p}\sum_{i=1}^n(\varepsilon_i^{'}-E\varepsilon_i^{'})(K_{ni}(y)-K_{ni}(y_k))\Big| \\ &\le \max_{1\le k\le N_n}\sup_{y\in D\cap I_k}\frac{c}{(n-p)h_n^p}\sum_{i=1}^n|\varepsilon_i^{'}-E\varepsilon_i^{'}|\|\frac{y-y_k}{h_n}\| \\ &\le O\Big(\frac{L_n}{h_n^{p+1}}\Big)=O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.}\end{aligned} \tag{3.11}$$

Similarly, it is easy to get

$$B_{n1}^{(2')}=O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.} \tag{3.12}$$

Now, we proceed to prove

$$B_{n1}^{(i'')}=O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s. } i=1,\,2. \tag{3.13}$$

Since the proofs are similar, we mainly treat the case $i=1$ in the following.

To treat $B_{n1}^{(1'')}$, the Bernstein's block technique is essential. For this purpose, set $n = 2s(n)r(n)+v(n)$, where $s(n)$, $r(n)$ and $v(n)$ are integer numbers satisfying $s(n)\to\infty$, $r(n)\to\infty$ as $n\to\infty$ and $0\le v(n)<2r(n)$. Let

$$\begin{aligned}V_n(j,k) &= \frac{1}{(n-p)h_n^p}\sum_{i=(j-1)r(n)+1}^{jr(n)}(\varepsilon_i^{'}-E\varepsilon_i^{'})K_{ni}(y_k), \\ V_{n1}^{(1)}(k) &= \sum_{j=1}^{s(n)}V_n(2j-1,k), \\ V_{n1}^{(2)}(k) &= \sum_{j=1}^{s(n)}V_n(2j,k), \\ R_n(k) &= \frac{1}{(n-p)h_n^p}\sum_{i=2s(n)r(n)+1}^{n}(\varepsilon_i^{'}-E\varepsilon_i^{'})K_{ni}(y_k).\end{aligned}$$

Then

$$B_{n1}^{(1'')}=V_{n1}^{(1)}(k)+V_{n1}^{(2)}(k)+R_n(k). \tag{3.14}$$

The following is to prove

$$\max_{1\le k\le N_n} |V_{n1}^{(1)}(k)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.}, \tag{3.15a}$$

$$\max_{1\le k\le N_n} |V_{n1}^{(2)}(k)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.}, \tag{3.15b}$$

$$\max_{1\le k\le N_n} |R_n(k)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.} \tag{3.15c}$$

For (3.15c), it is obvious that

$$\begin{aligned}\max_{1\le k\le N_n} |R_n(k)| &\le \frac{cv(n)}{(n-p)h_n^p}\frac{1}{v(n)}\sum_{i=2s(n)r(n)+1}^{n} |\varepsilon_i| \\ &= O\Big(\frac{r(n)}{nh_n^p}\Big) = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.,}\end{aligned} \tag{3.16}$$

if $r(n) = O((nh_n^p \ln n)^{1/2})$. The most difficult step is how to treat $V_{n1}^{(1)}$ ($V_{n1}^{(2)}$ follows the same argument). We utilize the independence approximation for α–mixing due to Bradley[3,Theorem 3]. By this, we can construct $\{V_n^\star(2j-1,k)\}_{j=1}^{s(n)}$ such that

(i) $\{V_n^\star(2j-1,k)\}_{j=1}^{s(n)}$ are independent;

(ii) $V_n^\star(2j-1,k)$ has the same distribution as $V_n(2j-1,k)$, $j=1,\ 2,\ \cdots,\ s(n)$;

(iii)
$$\begin{aligned}&P[|V_n^\star(2j-1,k) - V_n(2j-1,k)| > \epsilon] \\ &\le 18(\|V_n\|_\infty/\epsilon)^{1/2} \quad \sup|P(AB) - P(A)P(B)|\end{aligned} \tag{3.17}$$

for any $0 < \epsilon \le \|V_n(2j-1,k)\|_\infty \hat{=} \text{esssup}|V_n(j,k)| \le 2cr(n)b_n/(nh_n^p)$, where the supremum is taken over all sets A, B with A, B in the σ-algebras of events generated by $\{V_n(1,k),\ V_n(3,k),\ \cdots, V_n(2j-3,k)\}$ and $V_n(2j-1,k)$, respectively.

Set

$$V_{n1}^{'}(k) = \sum_{j=1}^{s(n)} V_n^\star(2j-1,k),$$

$$V_{n1}^{''}(k) = \sum_{j=1}^{s(n)} (V_n(2j-1,k) - V_n^\star(2j-1,k)).$$

Then

$$V_{n1}^{(1)}(k) = V_{n1}^{'}(k) + V_{n1}^{''}(k). \tag{3.18}$$

By Markov's inequality,

$$\begin{aligned}P_{n1} &= P\big(\max_{1\le k\le N_n} |V_{n1}^{'}(k)| > \epsilon_n\big) \\ &\le 2N_n \exp[-\lambda_n\epsilon_n] E\Big[\exp\Big(\lambda_n \sum_{j=1}^{s(n)} V_n^\star(2j-1,k)\Big)\Big].\end{aligned} \tag{3.19}$$

Take

$$\epsilon_n = c_0\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}, \quad \lambda_n = c[nh_n^p \ln n]^{1/2},$$

$$|V_n(j,k)| \le c\frac{b_n r(n)}{nh_n^p} = M_n, \quad r^a(n) = nN_n\Big(\frac{b_n}{h_n^p\epsilon_n}\Big)^{1/2}.$$

Then, recalling $h_n = (\frac{\ln n}{n})^\theta$ and

$$0 < \theta < \frac{m-2}{mp}, \quad a > \frac{(1+p\theta)[2(p+2)(m-1)+m]}{2m(1-p\theta)-4},$$

we have

$$\begin{aligned}
r^a(n) &= O\Big\{\Big(\frac{n}{h_n^p}\Big)^{1+\frac{p}{2}+\frac{m}{4(m-1)}}\Big/(\ln n)^{\frac{p}{2}+\frac{m}{4(m-1)}}\Big\}\\
&= O\Big\{\Big(\frac{n^{1+p\theta}}{\ln^{p\theta} n}\Big)^{1+\frac{p}{2}+\frac{m}{4(m-1)}}\Big/(\ln n)^{\frac{p}{2}+\frac{m}{4(m-1)}}\Big\}\\
&= o\Big(\Big(\frac{n}{\ln n}\Big)^{\frac{a(m(1-p\theta)-2)}{2(m-1)}}\Big)\\
&= o\Big(\Big(\frac{nh_n^p}{\lambda_n b_n}\Big)^a\Big),
\end{aligned}$$

and hence $\lambda_n M_n < 2$ for n large enough.

Since

$$\begin{aligned}
E|V_n^{\star 2}(2j-1,k)| &= E|V_n^2(2j-1,k)|\\
&\le \Big(\frac{1}{(n-p)h_n^p}\Big)^2 E\Big(\sum_{i=(2j-1)r(n)+1}^{2jr(n)}(\varepsilon_i^{'}-E\varepsilon_i^{'})K_{ni}(y_k)\Big)^2\\
&\le \frac{r(n)}{((n-p)h_n^p)^2}E(\varepsilon_i^{'}K_{ni}(y_k))^2\\
&= O\Big(\frac{r(n)}{n^2h_n^p}\Big) \quad \text{as} \quad n\to\infty,
\end{aligned} \tag{3.20}$$

it can be deduced by Lemma 3.3 that

$$\begin{aligned}
P_{n1} &\le 2N_n\exp\Big\{-\lambda_n\epsilon_n + s(n)\lambda_n^2 c\frac{r(n)}{n^2h_n^p}\Big\}\\
&= 2N_n\exp\{-(c_0-1)c\ln n\}.
\end{aligned}$$

(Here we may take the first c equal to the second, because the first c may be taken any positive constant.)

Hence if c_0 is large enough such that

$$(c_0-1)c > (p/2)(1+\theta(p+2))+1,$$

then

$$\sum_{n=1}^\infty P_{n1} < \infty,$$

from which it follows that

$$\max_{1\le k\le N_n}|V_{n1}'(k)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.} \tag{3.21}$$

Now we prove

$$\max_{1\le k\le N_n}|V_{n1}''(k)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.} \tag{3.22}$$

Observe

$$
\begin{aligned}
P_{n2} &= P(\max_{1\le k\le N_n} |V''_{n1}(k)| > \epsilon_n) \\
&\le \sum_{k=1}^{N_n} P(|V''_{n1}(k)| > \epsilon_n) \\
&\le \sum_{k=1}^{N_n}\sum_{j=1}^{s(n)} P\{|V^\star_n(2j-1,k) - V_n(2j-1,k)| > \epsilon/s(n)\} \\
&\le 18cN_n s(n)(\frac{r(n)b_n s(n)}{nh_n^p\epsilon_n})^{1/2}\alpha(r(n)) \\
&= 18c(r(n))^{a-1}\alpha(r(n)).
\end{aligned} \tag{3.23}
$$

Hence, by the conditions of the theorem, it easily follows that

$$\sum_{n=1}^{\infty} P_{n2} < \infty.$$

Thus (3.24) is obtained by Borel-Cantelli's Lemma.

By (3.24), (3.21) and (3.22), (3.15a) follows; and so does (3.15b) similarly. Together with (3.14) and (3.15), (3.13) with $i = 1$ is proved. For the case $i = 2$ in (3.13), if (3.20) used in the above arguments is replaced by

$$
\begin{aligned}
&E\Big(\sum_{i=(j-1)r(n)+1}^{jr(n)} \frac{1}{(n-p)h_n^p}(E\varepsilon'_i)(K_{ni}(y_k) - EK_{ni}(y_k))\Big)^2 \\
&= O\Big(\frac{r(n)}{n^2h_n^p}\Big) \quad \text{as} \quad n\to\infty,
\end{aligned} \tag{3.24}
$$

then (3.13) with $i = 2$ is also proved. (3.24) can be proved as follows. The left–hand side of (3.24) equals

$$
\begin{aligned}
&\Big(\frac{E\varepsilon'_i}{(n-p)h_n^p}\Big)^2 \sum_{i=(j-1)r(n)+1}^{jr(n)} E(K_{ni}(y_k) - EK_{ni}(y_k))^2 \\
&+ 2(\frac{E\varepsilon'_i}{(n-p)h_n^p})^2 \sum_{i_1=(j-1)r(n)+1}^{jr(n)-1}\ \sum_{i_2=i_1+1}^{jr(n)} \mathrm{Cov}(K_{ni_1}(y_k),\, K_{ni_2}(y_k)).
\end{aligned} \tag{3.25}
$$

By Lemma 3.1, the first term of (3.25) is bounded by

$$O\Big(\frac{r(n)}{n^2h_n^p}\Big) \quad \text{as} \quad n\to\infty$$

and the second term, by Lemma 3.1 and the appendix of [10], is bounded by

$$
\begin{aligned}
&c\frac{r(n)}{(nh_n^p)^2}\sum_{i=1}^{n}\min\{\alpha(i),\, h_n^{2p}\} \\
&\le c\frac{r(n)}{n^2h_n^p}h_n^{-p}\Big(\sum_{i=1}^{[h_n^{-p}]} h_n^{2p} + \sum_{i=[h_n^{-p}]+1}^{n}\alpha(i)\Big) \\
&= O\Big(\frac{r(n)}{n^2h_n^p}\Big).
\end{aligned}
$$

Thus (3.24) follows.

Now, by (3.7)–(3.13), we have

$$\sup_{y\in D}|B_{n1}(y)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.} \tag{3.26}$$

Since $\psi_Y(y)$ is uniformly continuous on D, it is easily proved that, as $n\to\infty$,

$$\frac{1}{nh_n^p}\sum_{i=p+1}^{n} K\Big(\frac{Y_i-y}{h_n}\Big) \overset{\text{a.s.}}{\to} \psi_Y(y)\int K(u)du \geq M_2\int K(u)du \hat{=} c_1 > 0,$$

uniformly for $y\in D$, hence

$$\inf_{y\in D}|B_{n2}(y)| \geq c_1/2 \quad \text{for } n \text{ large enough.} \tag{3.27}$$

Thus it follows from (3.6), (3.26), (3.27) that

$$\sup_{y\in D}|A_{n2}(y)| = O\Big(\Big(\frac{\ln n}{nh_n^p}\Big)^{1/2}\Big), \quad \text{a.s.},$$

which, together with (3.1), (3.5), deduces the desired results of this theorem.

Proof of Theorem 2.3. First, substituting (1.1) into (2.2) we get

$$\begin{aligned}
\widehat{\sigma^2} &= \sum_{t=p+1}^{n}(f(Y_t)+\varepsilon_t-\hat{f}_n(Y_t))^2 w(Y_t)\Big/\sum_{t=p+1}^{n} w(Y_t)\\
&= \sum_{t=p+1}^{n}(f(Y_t)-\hat{f}_n(Y_t))^2 w(Y_t)\Big/\sum_{t=p+1}^{n} w(Y_t)\\
&\quad + 2\sum_{t=p+1}^{n}(f(Y_t)-\hat{f}_n(Y_t))\varepsilon_t w(Y_t)\Big/\sum_{t=p+1}^{n} w(Y_t)\\
&\quad + \sum_{t=p+1}^{n}\varepsilon_t^2 w(Y_t)\Big/\sum_{t=p+1}^{n} w(Y_t).
\end{aligned} \tag{3.28}$$

Recall that $\{Y_t\}$ is α–mixing and hence is ergodic. Thus, by Theorem 2.2, the first and the last terms on the right-hand side of (3.28) converge to 0 and σ^2 almost surely, respectively, as $n\to\infty$, and hence the second term converges to 0 almost surely. The desired result is thus obtained.

Acknowledgements. The content of this paper is a part of the first author's Ph.D dissertation. Here he would like to express his sincere gratitude to Prof. Chen Xiru and Prof. An Hongzhi for their careful refereeing and to Prof. Li Guoying for her great helps and encouragements. Thanks are also due to an anonymous referee who pointed out an error in (2.7) of the original version of this paper.

References

[1] An, H. Z. & Huang, F. C., The geometric ergodicity of nolinear autoregressive model, Technical Report, Institute of Applied Mathematics, Academia Sinica, Beijing, 1994.

[2] Bhattacharya, R. N. and Lee, C., Ergodicity of nonlinear first order autoregressive models, *Journal of Theoretical Probability*, **8**(1995), 207–219.

[3] Bradley, R. C., Approximation theorem for strongly mixing random variables, *Michigan Mathematical Journal*, 30(1983), 69–81.

[4] Chan, K. S. & Tong, H., On the use of the deterministic Lyapunov fuction for the ergodicity of stochastic difference equation. *Advances of Applied Probability*, **17**(1985), 666–678.

[5] Chan, K. S. & Tong, H., A note on noisy chaos, *J. R. S. S.*, **56B:**2(1994), 301–311.
[6] Cheng, B. & Tong, H., On consistent non-parametric order determination and chaos, *J.R.S.S.*, **54B** (1992), 427–474.
[7] Cheng Ping, On strong consistency and convergence rate of improved kernel estimates for the regression function, *Journal of System Science and Mathematical Sciences*, 3(1983), 304–315.
[8] Cheng, P. E., Strong consistency of nearest neighbor regression function estimation, *Journal of Multivariate Analysis*, **15**(1984), 63–72.
[9] Hardle, W. & Chen, R., Nonparametric time series, a selective review with examples, *Invited Paper of 50th Session of ISI*, Beijing, 1995.
[10] Hall, P. & Heyde, C. C., Martingale limit theory and its application, Academic Press, New York, 1980.
[11] Lu Zudi, A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) models, *Statistics and Probability Letters*, **30**(1966), 305–311.
[12] Lu Zudi, Consistency of nonparametric kernel regression under α–mixing dependence, *Chinese Science Bulletin*, **41:**24(1996), 2219–2221 (in Chinese).
[13] Lu Zudi, Some topics for nonlinear time series, Ph.D Dissertation, Institute of Systems Science, Academia Sinica, Beijing, 1996.
[14] Lu Zudi & Cheng Ping, Distribution-free strong consistency of nonpaprametric kernel regression involving nonlinear time series, *Journal of Statistical Planning and Inference*, **65**(1997), 67–86.
[15] Masry, E. & Tjϕstheim, Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality, *Econometric Theory*, **11**(1995), 258–289.
[16] Parzen, E., On estimation of a probability density function and mode, *Annals of Mathematical Statistics*, 33(1962), 1065–1976.
[17] Stone, C. J., Optimal rates of convergence for nonparametric estimators, *Annal of Statistics*, **8**(1980), 1348–1360.
[18] Stone, C. J., Optimal global rates of convergence for nonparametric kernel regression, *Annal of Statistics*, **10**(1982), 1040–1053.
[19] Tjϕstheim, D., Nonlinear time series and Markov chains, *Advances in Applied Probability*, **22**(1990), 587–611.
[20] Tjϕstheim, D., Nonlinear time series: a selective review, *Scandinavian Journal of Statistics*, **21**(1994), 97–130.
[21] Tjϕstheim, D. & Auestad, B., Non-parametric identification of nonlinear time series: selecting significant lags, *Journal of American Statistical Association*, **89**(1994), 1410–1419.
[22] Tong, H., A personal overview of nonlinear time series analysis from a chaos perspective (with discussion and comments), *Scandinavian Journal of Statistics*, **22**(1995), 399–445.
[23] Truong, Y. M. & Stone, C. J., Nonparametric function estimation involving time series, *Ann. Statist.*, **20**(1992), 77–97.
[24] Zhao, L. C. & Fang, Z. B., Strong convergence of kernel estimates of nonparametric kernel regression functions, *Chinese Annals of Mathematics*, **6B:**2(1985),147–155.