Li et al. *Journal of Inequalities and Applications* (2016) 2016:142
DOI 10.1186/s13660-016-1087-z

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

Uniform convergence of estimator for nonparametric regression with dependent data

Xiaoqin Li, Wenzhi Yang* and Shuhe Hu

*Correspondence:
wzyang827@163.com
School of Mathematical Science,
Anhui University, Hefei, 230601,
China

Abstract

In this paper, the authors investigate the internal estimator of nonparametric regression with dependent data such as α-mixing. Under suitable conditions such as the arithmetically α-mixing and $E|Y_1|^s < \infty$ ($s > 2$), the convergence rate $|\widehat{m}_n(x) - m(x)| = O_P(a_n) + O(h^2)$ and uniform convergence rate $\sup_{x \in S_f'} |\widehat{m}_n(x) - m(x)| = O_p(a_n) + O(h^2)$ are presented, if $a_n = \sqrt{\frac{\ln n}{nh^d}} \to 0$. We generalize some results in Shen and Xie (Stat. Probab. Lett. 83:1915-1925, 2013).

MSC: 62G05; 62G08

Keywords: nonparametric regression; internal estimator; arithmetically α-mixing; uniform convergence rate

1 Introduction

Kernel-type estimators of the regression function are widely various situations because of their flexibility and efficiency, in the dependent cases as well as in the independent data case. This paper is concerned with the nonparametric regression model

$$Y_i = m(X_i) + U_i, \quad 1 \le i \le n, n \ge 1, \tag{1.1}$$

where $(X_i, Y_i) \in R^d \times R$, $1 \le i \le n$, U_i is random variable such that $E(U_i|X_i) = 0$, $1 \le i \le n$. Then one has

$$E(Y_i|X_i = x) = m(x), \quad 1 \le i \le n, n \ge 1.$$

The most popular nonparametric estimators of the unknown function $m(x)$ is the Nadaraya-Watson estimator $\widehat{m}_{NW}(x)$ given below and the local polynomials fitting. Let $K(x)$ be a kernel function. Define $K_h(x) = h^{-d}K(x/h)$, where $h = h_n$ is a sequence of positive bandwidths tending to zero as $n \to \infty$. Kernel-type estimators of the regression function are widely various situations because of their flexibility and efficiency, in the dependent data case as well as the independent data case. For the independent data, Nadaraya [2] and Watson [3] gave the most popular nonparametric estimators of the unknown function $m(x)$, named the Nadaraya-Watson estimator $\widehat{m}_{NW}(x)$, *i.e.*

$$\widehat{m}_{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)}. \tag{1.2}$$

Jones *et al.* [4] considered various versions of kernel-type regression estimators such as the Nadaraya-Watson estimator (1.2) and the local linear estimator. They also investigated the following internal estimator:

$$\widehat{m}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i K_h(x - X_i)}{f(X_i)} \tag{1.3}$$

for known density $f(\cdot)$. The term 'internal' stands for the fact that the factor $\frac{1}{f(X_i)}$ is internal to the summation, while the estimator $\widehat{m}_{NW}(x)$ has the factor $\frac{1}{\hat{f}(x)} = \frac{1}{n^{-1}\sum_{i=1}^{n} K_h(x-X_i)}$ externally to the summation.

The internal estimator was first proposed by Mack and Müller [5]. Jones *et al.* [4] studied various kernel-type regression estimators, including introduced the internal estimator (1.3). Linton and Nielsen [6] introduced 'integration method', based on direct integration of initial pilot estimator (1.3). Linton and Jacho-Chávez [7] studied the two internal nonparametric estimators with the estimator similar to estimator (1.3) but in place of an unknown density $f(\cdot)$, a classical kernel estimator $\widehat{f}(x) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i)$ is used. Much work has been done for the kernel density estimation. For example, Masry [8] gave the recursive probability density estimation for a mixing dependent sample, Roussas *et al.* [9] and Tran *et al.* [10] investigated the fixed design regression for dependent data, Liebscher [11] studied the strong convergence of sums of α-mixing random variables and gave its application to density estimation, Hansen [12] obtained the uniform convergence rates for kernel estimation with dependent data, and so on. For more work as regards kernel estimation, we can also refer to [13–30] and the references therein.

Let $(\Omega, \mathcal{F}, P)$ be a fixed probability space. Denote $N = \{1, 2, \ldots, n, \ldots\}$. Let $\mathcal{F}_m^n = \sigma(X_i, m \le i \le n, i \in N)$ be the σ-field generated by random variables $X_m, X_{m+1}, \ldots, X_n$, $1 \le m \le n$. For $n \ge 1$, we define

$$\alpha(n) = \sup_{m \in N} \sup_{A \in \mathcal{F}_1^m, B \in \mathcal{F}_{m+n}^\infty} \left| P(AB) - P(A)P(B) \right|.$$

Definition 1.1 If $\alpha(n) \downarrow 0$ as $n \to \infty$, then $\{X_n, n \ge 1\}$ is called a strong mixing or α-mixing sequence.

Recently, Shen and Xie [1] obtained the strong consistency of the internal estimator (1.3) under α-mixing data. In their paper, the process is assumed to be geometrically α-mixing sequence, *i.e.* the mixing coefficients $\alpha(n)$ satisfy $\alpha(n) \le \beta_0 e^{-\beta_1 n}$, where $\beta_0 > 0$ and $\beta_1 > 0$. They also supposed that the sequence $\{Y_n, n \ge 1\}$ is bounded, as well as the density $f(x)$ of X_1. Inspired by Hansen [12], Shen and Xie [1] and other papers above, we also investigate the convergence of the internal estimator (1.3) under α-mixing data. The process is supposed to be an arithmetically α-mixing sequence, *i.e.* the mixing coefficients $\alpha(n)$ satisfy $\alpha(n) \le Cn^{-\beta}$, $C > 0$, and $\beta > 0$. Without the bounded conditions of $\{Y_n, n \ge 1\}$ and the density $f(x)$ of X_1, we establish the convergence rate and uniform convergence rate for the internal estimator (1.3). For the details, please see our results in Section 2. The conclusion and the lemmas and proofs of the main results are presented in Section 3 and Section 4, respectively.

Regarding notation, for $x = (x_1, \ldots, x_d) \in R^d$, set $\|x\| = \max(|x_1|, \ldots, |x_d|)$. Throughout the paper, $C, C_1, C_2, C_3, \ldots$, denote some positive constants not depending on n, which may be different in various places. $\lfloor x \rfloor$ denotes the largest integer not exceeding x. $\to$ means to take the limit as $n \to \infty$, $\xrightarrow{P}$ means to convergence in probability. $X \overset{d}{=} Y$ means that the random variables X and Y have the same distribution. A sequence $\{X_n, n \geq 1\}$ is said to be of second-order stationarity if $(X_1, X_{1+k}) \overset{d}{=} (X_i, X_{i+k})$, $i \geq 1$, $k \geq 1$.

2 Results and discussion

2.1 Some assumptions

Assumption 2.1 We assume the data observed $\{(X_n, Y_n), n \geq 1\}$ valued in $R^d \times R$ comes from a second-order stationary stochastic sequence. The sequence $\{(X_n, Y_n), n \geq 1\}$ is also assumed to be arithmetically α-mixing with mixing coefficients $\alpha(n)$ such that

$$\alpha(n) \leq An^{-\beta}, \tag{2.1}$$

where $A < \infty$ and for some $s > 2$

$$E|Y_1|^s < \infty \tag{2.2}$$

and

$$\beta \geq \frac{2s-2}{s-2}. \tag{2.3}$$

The known density $f(\cdot)$ of X_1 is upon its compact support S_f and it is also assumed that $\inf_{x \in S_f} f(x) > 0$. Let B_0 be a positive constant such as

$$\sup_{x \in S_f} E\bigl(|Y_1|^s | X_1 = x\bigr) f(x) \leq B_0. \tag{2.4}$$

Also, there is a $j^* < \infty$ such that for all $j \geq j^*$

$$\sup_{x_1 \in S_f, x_{j+1} \in S_f} E\bigl(|Y_1 Y_{j+1}| | X_1 = x_1, X_{j+1} = x_{j+1}\bigr) f_j(x_1, x_{j+1}) \leq B_1, \tag{2.5}$$

where B_1 is a positive constant and $f_j(x_1, x_{j+1})$ denotes the joint density of (X_1, X_{j+1}).

Assumption 2.2 There exist two positive constants $\bar{K} > 0$ and $\mu > 0$ such that

$$\sup_{u \in R^d} \bigl|K(u)\bigr| \leq \bar{K} \quad \text{and} \quad \int_{R^d} \bigl|K(u)\bigr| \, du = \mu. \tag{2.6}$$

Assumption 2.3 Denote by S_f^0 the interior of S_f. For $x \in S_f^0$, the function $m(x)$ is twice differentiable and there exists a positive constant b such that

$$\left| \frac{\partial^2 m(x)}{\partial x_i \, \partial x_j} \right| \leq b, \quad \forall i, j = 1, 2, \ldots, d.$$

The kernel density function $K(\cdot)$ is symmetrical and satisfies

$$\int_{R^d} |v_i||v_j|K(v)\,dv < \infty, \quad \forall i,j = 1,2,\ldots,d.$$

Assumption 2.4 The kernel function satisfies the Lipschitz condition, *i.e.*

$$\exists L > 0, \quad \bigl|K(u) - K(u')\bigr| \le L\|u - u'\|, \quad u,u' \in R^d.$$

Remark 2.1 Similar to Assumption 2 of Hansen [12], Assumption 2.1 specifies that the serial dependence in the data is of strong mixing type, and equations (2.1)-(2.3) specify a required decay rate. Condition (2.4) controls the tail behavior of the conditional expectation $E(|Y_1|^s|X_1 = x)$, condition (2.5) places a similar bound on the joint density and conditional expectation. Assumptions 2.2-2.4 are the conditions of kernel function $K(u)$, *i.e.*, Assumption 2.2 is a general condition, Assumption 2.3 is used to estimate the convergence rate of $|E\widehat{m}_n(x) - m(x)|$, and Assumption 2.4 is used to investigate the uniform convergence rate of the internal estimator $\widehat{m}_n(x)$.

2.2 Main results

First, we investigate the variance bound of estimator $\widehat{m}_n(x)$. For $1 \le r \le s$ and $s > 2$, denote $\bar{\mu}(r,s) := \frac{(B_0)^{r/s}\bar{K}^{r-1}\mu}{(\inf_{x\in S_f} f(x))^{r-1+r/s}}$, where B_0, $\inf_{x\in S_f} f(x)$, $\bar{K}$, and μ are defined in Assumptions 2.1 and 2.2.

Theorem 2.1 *Let Assumption* 2.1 *and Assumption* 2.2 *be fulfilled. Then there exists a* $\Theta < \infty$ *such that for n sufficiently large and* $x \in S_f$

$$\operatorname{Var}\bigl(\widehat{m}_n(x)\bigr) \le \frac{\Theta}{nh^d}, \tag{2.7}$$

where $\Theta := \bar{\mu}(2,s) + 2j^*\bar{\mu}(2,s) + 2(\bar{\mu}^2(1,s) + \frac{B_1\mu^2}{(\inf_{x\in S_f} f)^2}) + \frac{16A^{1-2/s}\bar{\mu}^{\frac{2}{s}}(s,s)}{(s-2)/s}$.

As an application to Theorem 2.1, we obtain the weak consistency of estimator $\widehat{m}_n(x)$.

Corollary 2.1 *Let Assumption* 2.1 *and Assumption* 2.2 *be fulfilled and* $K(\cdot)$ *be a density function. For* $x \in S_f$, $m(x)$ *is supposed to be continuous at x. If* $nh^d \to \infty$ *as* $n \to \infty$, *then*

$$\widehat{m}_n(x) \xrightarrow{P} m(x). \tag{2.8}$$

Next, the convergence rate of estimator $\widehat{m}_n(x)$ is presented.

Theorem 2.2 *For* $0 < \theta < 1$ *and* $s > 2$, *let Assumptions* 2.1-2.3 *hold, where the mixing exponent* β *satisfies*

$$\beta > \max\left\{\frac{2 - 4\theta + 2\theta s}{(1-\theta)(s-2)}, \frac{2s-2}{s-2}\right\}. \tag{2.9}$$

Denote $a_n = \sqrt{\frac{\ln n}{nh^d}}$ *and take* $h = n^{-\theta/d}$. *Then for* $x \in S_f^0$, *one has*

$$\bigl|\widehat{m}_n(x) - m(x)\bigr| = O_P(a_n) + O\bigl(h^2\bigr). \tag{2.10}$$

Third, we now investigate the uniform convergence rate of estimator $\widehat{m}_n(x)$ and its convergence over a compact set. Let S'_f be any compact set contained in S^0_f.

Theorem 2.3 *For* $0 < \theta < 1$ *and* $s > 2$, *let Assumptions* 2.1-2.3 *be fulfilled*, *where the mixing exponent* β *satisfies*

$$\beta > \max\left\{\frac{s\theta d + 3\theta s + sd + s - 2\theta d - 4\theta}{(1-\theta)(s-2)}, \frac{2s-2}{s-2}\right\}. \tag{2.11}$$

Suppose that Assumption 2.4 *is also fulfilled*. *Denote* $a_n = \sqrt{\frac{\ln n}{nh^d}}$ *and take* $h = n^{-\theta/d}$. *Then*

$$\sup_{x\in S'_f}\bigl|\widehat{m}_n(x) - m(x)\bigr| = O_p(a_n) + O\bigl(h^2\bigr). \tag{2.12}$$

2.3 Discussion

The parametric θ in Theorem 2.2 and Theorem 2.3 plays the role of a bridge between the process (*i.e.* mixing exponent) and choice of positive bandwidth h. For example, if $d = 2$, $\theta = \frac{1}{3}$, and $\beta > \max\{\frac{s+1}{s-2}, \frac{2s-2}{s-2}\}$, then we take $h = n^{-1/6}$ in Theorem 2.2 and obtain the convergence rate $|\widehat{m}_n(x) - m(x)| = O_P((\ln n)^{1/2}n^{-1/3})$. Similarly, if $d = 2$, $\theta = \frac{1}{3}$, and $\beta > \frac{2s-2}{s-2}$, then we choose $h = n^{-1/6}$ in Theorem 2.3 and establish the uniform convergence rate $\sup_{x\in S'_f}|\widehat{m}_n(x) - m(x)| = O_p((\ln n)^{1/2}n^{-1/3})$.

3 Conclusion

On the one hand, similar to Theorem 2.1, Hansen [12] investigated the kernel average estimator

$$\widehat{\Psi}(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i K_h(x - X_i),$$

and obtained the variance bound $\mathrm{Var}(\widehat{\Psi}(x)) \leq \frac{\Theta}{nh^d}$, where Θ is a positive constant. For the details, see Theorem 1 of Hansen [12]. Under some other conditions, Hansen [12] also gave the uniform convergence rates such as $\sup_{\|x\|\leq c_n}|\widehat{\Psi}(x) - E\widehat{\Psi}(x)| = O_P(a_n)$, where $a_n = \sqrt{\frac{\ln n}{nh^d}} \to 0$ and $\{c_n\}$ is a sequence of positive constant (see Theorems 2-5 of Hansen [12]). On the other hand, under the conditions such as the geometrically α-mixing and $\{Y_n, n \geq 1\}$ is bounded as well as the density function $f(x)$ of X_1, Shen and Xie [1] obtained the complete convergence such as $|\widehat{m}_n(x) - m(x)| \xrightarrow{a.c.} 0$, if $\frac{\ln^2 n}{nh^d} \to 0$ (see Theorem 3.1 of Shen and Xie [1]), the uniform complete convergence such as $\sup_{x\in S'_f}|\widehat{m}_n(x) - m(x)| \xrightarrow{a.c.} 0$, if $\frac{\ln^2 n}{nh^d} \to 0$ (see Theorem 4.1 of Shen and Xie [1]). In this paper, we do not need the bounded conditions of $\{Y_n, n \geq 1\}$ and $f(x)$ of X_1, and we also investigate the convergence of the internal estimator $\hat{m}_n(x)$. Under some weak conditions such as the arithmetically α-mixing and $E|Y_1|^s < \infty$, $s > 2$, we establish the convergence rate in Theorem 2.2 such as $|\widehat{m}_n(x) - m(x)| = O_P(a_n) + O(h^2)$ if $a_n = \sqrt{\frac{\ln n}{nh^d}} \to 0$, and uniform convergence rate in Theorem 2.3 such as $\sup_{x\in S'_f}|\widehat{m}_n(x) - m(x)| = O_p(a_n) + O(h^2)$ if $a_n = \sqrt{\frac{\ln n}{nh^d}} \to 0$. In Theorem 2.2 and Theorem 2.3, we have $|\widehat{m}_n(x) - E\widehat{m}_n(x)| = O_P(a_n)$ and $\sup_{x\in S'_f}|\widehat{m}_n(x) - E\widehat{m}_n(x)| = O_P(a_n)$, where the convergence rates are the same as that obtained by Hansen [12]. So, we relatively generalize the results in Shen and Xie [1].

4 Some lemmas and the proofs of the main results

Lemma 4.1 (Hall and Heyde, [31], Corollary A.2, *i.e.* Davydov's lemma) *Suppose that X and Y are random variables which are $\mathscr{G}$-measurable and $\mathscr{H}$-measurable, respectively, and $E|X|^p < \infty$, $E|Y|^q < \infty$, where $p, q > 1$, $p^{-1} + q^{-1} < 1$. Then*

$$\left|E(XY) - EXEY\right| \leq 8\left(E|X|^p\right)^{1/p}\left(E|Y|^q\right)^{1/q}\left[\alpha(\mathscr{G}, \mathscr{H})\right]^{1-p^{-1}-q^{-1}}.$$

Lemma 4.2 (Liebscher [32], Proposition 5.1) *Let $\{X_n, n \geq 1\}$ be a stationary α-mixing sequence with mixing coefficient $\alpha(k)$. Assume that $EX_i = 0$ and $|X_i| \leq S < \infty$, a.s., $i = 1, 2, \ldots, n$. Then, for $n, m \in N$, $0 < m \leq n/2$, and all $\varepsilon > 0$,*

$$P\left(\left|\sum_{i=1}^{n} X_i\right| > \varepsilon\right) \leq 4\exp\left\{-\frac{\varepsilon^2}{16(\frac{n}{m}D_m + \frac{1}{3}\varepsilon Sm)}\right\} + 32\frac{S}{\varepsilon}n\alpha(m),$$

where $D_m = \max_{1\leq j\leq 2m} \operatorname{Var}(\sum_{i=1}^{j} X_i)$.

Lemma 4.3 (Shen and Xie [1], Lemma 3.2) *Under Assumption 2.3, for $x \in S_f^0$, one has*

$$\left|E\widehat{m}_n(x) - m(x)\right| = O\left(h^2\right).$$

Proof of Theorem 2.1 For $x \in S_f$, let $Z_i := \frac{Y_i K_h(x-X_i)}{f(X_i)}$, $1 \leq i \leq n$. Consider now

$$\widehat{m}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i K_h(x - X_i)}{f(X_i)} = \frac{1}{n}\sum_{i=1}^{n} Z_i, \quad n \geq 1.$$

For any $1 \leq r \leq s$ and $s > 2$, it follows from (2.4) and (2.6) that

$$\begin{aligned}
h^{d(r-1)}E|Z_1|^r &= h^{d(r-1)}E\left|\frac{K_h(x - X_1)Y_1}{f(X_1)}\right|^r \\
&= h^{d(r-1)}E\left(\frac{|K_h(x - X_1)|^r}{f^r(X_1)}E\left(|Y_1|^r|X_1\right)\right) \\
&= \int_{S_f}\left|K\left(\frac{x-u}{h}\right)\right|^r E\left(|Y_1|^r|X_1 = u\right)\frac{1}{h^d}\frac{f(u)}{f^r(u)}\,du \\
&\leq \int_{S_f}\left|K\left(\frac{x-u}{h}\right)\right|^r \left(E\left(|Y_1|^s|X_1 = u\right)f(u)\right)^{r/s}\frac{1}{h^d}\frac{1}{f^{r-1+r/s}(u)}\,du \\
&\leq \frac{(B_0)^{r/s}\bar{K}^{r-1}\mu}{(\inf_{x\in S_f} f)^{r-1+r/s}} := \bar{\mu}(r, s) < \infty. \qquad (4.1)
\end{aligned}$$

For $j \geq j^*$, by (2.5), one has

$$\begin{aligned}
E|Z_1 Z_{j+1}| &= E\left|\frac{K_h(x - X_1)K_h(x - X_{j+1})Y_1 Y_{j+1}}{f(X_1)f(X_{j+1})}\right| \\
&= E\left(\frac{|K_h(x - X_1)K_h(x - X_{j+1})|}{f(X_1)f(X_{j+1})}E\left(|Y_1 Y_{j+1}||X_1, X_{j+1}\right)\right) \\
&= \int_{S_f}\int_{S_f}\left|K\left(\frac{x - u_1}{h}\right)K\left(\frac{x - u_{j+1}}{h}\right)\right|E\left(|Y_1 Y_{j+1}||X_1 = u_1, X_j = u_{j+1}\right)
\end{aligned}$$

$$\begin{aligned}
&\times \frac{1}{h^{2d}} \frac{1}{f(u_1)f(u_{j+1})} f_j(u_1, u_{j+1})\, du_1\, du_{j+1} \\
&\le \frac{B_1}{(\inf_{x\in S_f} f)^2} \int_{R^d}\int_{R^d} \bigl|K(u_1)K(u_{j+1})\bigr|\, du_1\, du_{j+1} \le \frac{B_1\mu^2}{(\inf_{x\in S_f} f)^2} < \infty.
\end{aligned} \tag{4.2}$$

Define the covariances $\gamma_j = \operatorname{Cov}(Z_1, Z_{j+1})$, $j > 0$. Assume that n is sufficiently large so that $h^{-d} \ge j^*$. We now bound the γ_j separately for $j \le j^*$, $j^* < j \le h^{-d}$, and $h^{-d} < j < \infty$. First, for $1 \le j \le j^*$, by the Cauchy-Schwarz inequality and (4.1) with $r = 2$,

$$|\gamma_j| \le \sqrt{\operatorname{Var}(Z_1) \cdot \operatorname{Var}(Z_{j+1})} = \operatorname{Var}(Z_1) \le EZ_1^2 \le \bar{\mu}(2, s)h^{-d}. \tag{4.3}$$

Second, for $j^* < j \le h^{-d}$, in view of (4.1) ($r = 1$) and (4.2), we establish that

$$|\gamma_j| \le E|Z_1 Z_{j+1}| + \bigl(E|Z_1|\bigr)^2 \le \frac{B_1\mu^2}{(\inf_{x\in S_f} f)^2} + \bar{\mu}^2(1, s). \tag{4.4}$$

Third, for $j > h^{-d}$, we apply Lemma 4.1, (2.1) and (4.1) with $r = s$ ($s > 2$) and we thus obtain

$$\begin{aligned}
|\gamma_j| &\le 8\bigl(\alpha(j)\bigr)^{1-2/s}\bigl(E|Z_1|^s\bigr)^{2/s} \le 8A^{-1-2/s} j^{-\beta(1-2/s)}\bigl(\bar{\mu}(s,s)h^{-d(s-1)}\bigr)^{2/s} \\
&\le 8A^{-1-2/s}\bar{\mu}^{\frac{2}{s}}(s,s) j^{-(2-2/s)} h^{-2d(s-1)/s}.
\end{aligned} \tag{4.5}$$

Consequently, in view of the property of second-order stationarity and (4.3)-(4.5), for n sufficiently large, we establish

$$\begin{aligned}
\operatorname{Var}\bigl(\widehat{m}_n(x)\bigr) &= \frac{1}{n^2}\operatorname{Var}\Biggl(\sum_{i=1}^n Z_i\Biggr) = \frac{1}{n^2}\Biggl(n\gamma_0 + 2n\sum_{j=1}^{n-1}\Bigl(1 - \frac{j}{n}\Bigr)\gamma_j\Biggr) \\
&\le \frac{1}{n^2}\Biggl(nh^{-d}\bar{\mu}(2,s) + 2n\sum_{1\le j\le j^*}|\gamma_j| + 2n\sum_{j^*<j\le h^{-d}}|\gamma_j| + 2n\sum_{h^{-d}<j}|\gamma_j|\Biggr) \\
&\le \frac{1}{n}\bar{\mu}(2,s)h^{-d} + \frac{2}{n}j^*\bar{\mu}(2,s)h^{-d} + \frac{2}{n}\bigl(h^{-d} - j^*\bigr)\Biggl(\bar{\mu}^2(1,s) + \frac{B_1\mu^2}{(\inf_{x\in S_f} f)^2}\Biggr) \\
&\quad + \frac{2}{n}\sum_{h^{-d}<j<\infty} 8A^{1-2/s}\bar{\mu}^{\frac{2}{s}}(s,s) j^{-(2-2/s)} h^{-2d(s-1)/s} \\
&\le \Biggl(\bar{\mu}(2,s) + 2j^*\bar{\mu}(2,s) + 2\Biggl(\bar{\mu}^2(1,s) + \frac{B_1\mu^2}{(\inf_{x\in S_f} f)^2}\Biggr) \\
&\quad + \frac{16A^{1-2/s}\bar{\mu}^{\frac{2}{s}}(s,s)}{(s-2)/s}\Biggr)\frac{1}{nh^d} \\
&:= \frac{\Theta}{nh^d},
\end{aligned} \tag{4.6}$$

where the final inequality uses the fact that $\sum_{j=k+1}^{\infty} j^{-\delta} \le \int_k^\infty x^{-\delta}\,dx = \frac{k^{1-\delta}}{\delta-1}$ for $\delta > 1$ and $k \ge 1$.

Thus, (2.7) is completely proved. □

Proof of Corollary 2.1 It is easy to see that

$$\big|\widehat{m}_n(x)-m(x)\big| \le \big|\widehat{m}_n(x)-E\widehat{m}(x)\big| + \big|E\widehat{m}_n(x)-m(x)\big|,$$

which can be treated as 'variance' part and 'bias' part, respectively.

On the one hand, by the proof of Theorem 3.1 of Shen and Xie [1], one has $|E\widehat{m}_n(x)-m(x)| \to 0$. On the other hand, we apply Theorem 2.1 and obtain that $|\widehat{m}_n(x)-E\widehat{m}_n(x)| \xrightarrow{P} 0$. So, (2.8) is proved finally. □

Proof of Theorem 2.2 Let $\tau_n = a_n^{-1/(s-1)}$ and define

$$R_n = \widehat{m}_n(x) - \frac{1}{n}\sum_{i=1}^n \frac{Y_i K_h(x-X_i)}{f(X_i)} I\big(|Y_i| \le \tau_n\big) = \frac{1}{n}\sum_{i=1}^n \frac{Y_i K_h(x-X_i)}{f(X_i)} I\big(|Y_i| > \tau_n\big).$$

Obviously, we have

$$\begin{aligned}
E|R_n| &\le E\biggl(\frac{|K_h(x-X_1)|}{f(X_1)} E\big(|Y_1| I\big(|Y_1|>\tau_n\big)|X_1\big)\biggr) \\
&\le \frac{1}{(\inf_{S_f} f)} \int_{S_f} \biggl|K\biggl(\frac{x-u}{h}\biggr)\biggr| E\big(|Y_1| I\big(|Y_1|>\tau_n\big)|X_1=u\big)\frac{f(u)}{h^d}\,du \\
&= \frac{1}{(\inf_{S_f} f)} \int_{S_f} \big|K(u)\big| E\big(|Y_1| I\big(|Y_1|>\tau_n\big)|X_1=x-hu\big) f(x-hu)\,du \\
&\le \frac{1}{(\inf_{S_f} f)} \frac{1}{\tau_n^{s-1}} \int_{S_f} \big|K(u)\big| E\big(|Y_1|^s|X_1=x-hu\big) f(x-hu)\,du \\
&\le \frac{1}{(\inf_{S_f} f)} \frac{\mu B_0}{\tau_n^{s-1}}.
\end{aligned} \tag{4.7}$$

Combining Markov's inequality with (4.7), one has

$$|R_n - ER_n| = O_P\big(\tau_n^{-(s-1)}\big) = O_P(a_n). \tag{4.8}$$

Denote

$$\widetilde{m}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{Y_i K_h(x-X_i)}{f(X_i)} I\big(|Y_i| \le \tau_n\big) := \frac{1}{n}\sum_{i=1}^n \tilde{Z}_n, \quad n \ge 1. \tag{4.9}$$

It can be seen that

$$\begin{aligned}
\big|\widehat{m}_n(x)-m(x)\big| &\le \big|\widehat{m}_n(x)-E\widehat{m}_n(x)\big| + \big|E\widehat{m}_n(x)-m(x)\big| \\
&\le \big|\widetilde{m}_n(x)-E\widetilde{m}_n(x)\big| + |R_n - ER_n| + \big|E\widehat{m}_n(x)-m(x)\big|.
\end{aligned} \tag{4.10}$$

Similar to the proof of (4.6), it can be argued that

$$\operatorname{Var}\Biggl(\sum_{i=1}^j \tilde{Z}_i\Biggr) \le C_2 j h^{-d},$$

which implies

$$D_m = \max_{1\le j\le 2m} \operatorname{Var}\left(\sum_{i=1}^{j} \tilde{Z}_i\right) \le C_3 m h^{-d}.$$

Meanwhile, one has $|\tilde{Z}_i - E\tilde{Z}_i| \le \frac{C_1\tau_n}{h^d}$, $1 \le i \le n$. Setting $m = a_n^{-1}\tau_n^{-1}$ and using (2.9), $h = n^{-\theta/d}$, and Lemma 4.2 with $\varepsilon = a_n n$, we obtain for n sufficiently large

$$\begin{aligned}
&P\big(\big|\widetilde{m}_n(x) - E\widetilde{m}_n(x)\big| > a_n\big) \\
&\quad= P\left(\left|\sum_{i=1}^{n}(\tilde{Z}_i - E\tilde{Z}_i)\right| > na_n\right) \\
&\quad\le 4\exp\left\{-\frac{na_n^2}{16(C_3h^{-d} + \frac{1}{3}C_1h^{-d})}\right\} + 32\frac{C_1\tau_n}{a_n n h^d} nA(a_n\tau_n)^{\beta} \\
&\quad\le 4\exp\left\{-\frac{\ln n}{16(C_3 + \frac{1}{3}C_1)}\right\} + C_4 h^{-d} a_n^{\frac{\beta(s-2)-s}{s-1}} \\
&\quad\le o(1) + C_5 n^{\theta} n^{(\theta-1)\frac{\beta(s-2)-s}{2(s-1)}} (\ln n)^{\frac{\beta(s-2)-s}{2(s-1)}} \\
&\quad= o(1) + C_5 n^{\frac{\beta(\theta-1)(s-2)+\theta s+s-2\theta}{2(s-1)}} (\ln n)^{\frac{\beta(s-2)-s}{2(s-1)}} = o(1),
\end{aligned} \tag{4.11}$$

in view of $s > 2$, $0 < \theta < 1$, $\beta > \max\{\frac{\theta s+s-2\theta}{(1-\theta)(s-2)}, \frac{2s-2}{s-2}\}$, and $\frac{\beta(\theta-1)(s-2)+\theta s+s-2\theta}{2(s-1)} < 0$.

Consequently, by (4.8), (4.10), (4.11), and Lemma 4.3, we establish the result of (2.10). □

Proof of Theorem 2.3 We use some similar notation in the proof of Theorem 2.2. Obviously, one has

$$\sup_{x\in S_f'}\big|\widehat{m}_n(x) - m(x)\big| \le \sup_{x\in S_f'}\big|\widehat{m}_n(x) - E\widehat{m}_n(x)\big| + \sup_{x\in S_f'}\big|E\widehat{m}_n(x) - m(x)\big|. \tag{4.12}$$

By the proof of (3.21) of Shen and Xie [1], we establish that

$$\big|E\widehat{m}_n(x) - m(x)\big| \le h^2\frac{b}{2}\sum_{1\le i,j\le d}\int_{R^d} K(v)|v_iv_j|\,dv \le C_0h^2,$$

which implies

$$\sup_{x\in S_f'}\big|E\widehat{m}_n(x) - m(x)\big| = O\big(h^2\big). \tag{4.13}$$

Since $\hat{m}_n(x) = R_n(x) + \tilde{m}_n(x)$,

$$\sup_{x\in S_f'}\big|\widehat{m}_n(x) - E\widehat{m}_n(x)\big| \le \sup_{x\in S_f'}\big|\widetilde{m}_n(x) - E\widetilde{m}_n(x)\big| + \sup_{x\in S_f'}|R_n - ER_n|. \tag{4.14}$$

It follows from the proof of (4.8) that

$$\sup_{x\in S_f'}|R_n - ER_n| = O_p(a_n). \tag{4.15}$$

Since S_f' is a compact set, there exists a $\xi > 0$ such that $S_f' \subset B := \{x : \|x\| \le \xi\}$. Let ν_n be a positive integer. Take an open covering $\bigcup_{j=1}^{\nu_n^d} B_{jn}$ of B, where $B_{jn} \subset \{x : \|x - x_{jn}\| \le \frac{\xi}{\nu_n}\}$, $j = 1, 2, \ldots, \nu_n^d$, and their interiors are disjoint. So it follows that

$$\begin{aligned}
&\sup_{x\in S_f'} \bigl|\widetilde{m}_n(x) - E\widetilde{m}_n(x)\bigr| \\
&\quad\le \max_{1\le j\le \nu_n^d} \sup_{x\in B_{jn}\cap S_f'} \bigl|\widetilde{m}_n(x) - E\widetilde{m}_n(x)\bigr| \\
&\quad\le \max_{1\le j\le \nu_n^d} \sup_{x\in B_{jn}\cap S_f'} \bigl|\widetilde{m}_n(x) - \widetilde{m}_n(x_{jn})\bigr| + \max_{1\le j\le \nu_n^d} \bigl|\widetilde{m}_n(x_{jn}) - E\widetilde{m}_n(x_{jn})\bigr| \\
&\qquad + \max_{1\le j\le \nu_n^d} \sup_{x\in B_{jn}\cap S_f'} \bigl|E\widetilde{m}_n(x_{jn}) - E\widetilde{m}_n(x)\bigr| \\
&\quad:= I_1 + I_2 + I_3.
\end{aligned} \tag{4.16}$$

By the definition of $\widetilde{m}_n(x)$ in (4.9) and the Lipschitz condition of K,

$$\begin{aligned}
\bigl|\widetilde{m}_n(x) - \widetilde{m}_n(x_{jn})\bigr| &\le \Bigl(\inf_{S_f'} f\Bigr)^{-1} \frac{\tau_n}{nh^d} \sum_{i=1}^n \biggl|K\biggl(\frac{x - X_i}{h}\biggr) - K\biggl(\frac{x_{jn} - X_i}{h}\biggr)\biggr| \\
&\le \frac{L\tau_n}{h^{d+1}\inf_{S_f'} f} \|x - x_{jn}\|, \quad x \in S_f'.
\end{aligned}$$

Taking $\nu_n = \lfloor \frac{\tau_n}{h^{d+1}a_n} \rfloor + 1$, we obtain

$$\sup_{x\in B_{jn}\cap S_f'} \bigl|\widetilde{m}_n(x) - \widetilde{m}_n(x_{jn})\bigr| \le \frac{L\xi}{\inf_{S_f'} f} a_n, \quad 1 \le j \le \nu_n^d, \tag{4.17}$$

and

$$I_1 = \max_{1\le j\le \nu_n^d} \sup_{x\in B_{jn}\cap S_f'} \bigl|\widetilde{m}_n(x) - \widetilde{m}_n(x_{jn})\bigr| = O(a_n). \tag{4.18}$$

In view of $|E\widetilde{m}_n(x_{jn}) - E\widetilde{m}_n(x)| \le E|\widetilde{m}_n(x_{jn}) - \widetilde{m}_n(x)|$, we have by (4.17)

$$I_3 = \max_{1\le j\le \nu_n^d} \sup_{x\in B_{jn}\cap S_f'} \bigl|E\widetilde{m}_n(x_{jn}) - E\widetilde{m}_n(x)\bigr| \le \frac{L\xi}{\inf_{S_f'} f} a_n = O(a_n). \tag{4.19}$$

For $1 \le i \le n$ and $1 \le j \le \nu_n^d$, denote $\tilde{Z}_i(j) = \frac{Y_i K_h(x_{jn} - X_i)}{f(X_i)} I(|Y_i| \le \tau_n)$. Then similar to the proof of (4.11), we obtain by Lemma 4.2 with $m = a_n^{-1}\tau_n^{-1}$ and $\varepsilon = Mna_n$ for n sufficiently large

$$\begin{aligned}
P\bigl(|I_2| > Ma_n\bigr) &= P\Bigl(\max_{1\le j\le \nu_n^d} \bigl|\widetilde{m}_n(x_{jn}) - E\widetilde{m}_n(x_{jn})\bigr| > Ma_n\Bigr) \\
&\le \sum_{j=1}^{\nu_n^d} P\Biggl(\Biggl|\sum_{i=1}^n \bigl(\tilde{Z}_i(j) - E\tilde{Z}_i(j)\bigr)\Biggr| > Mna_n\Biggr)
\end{aligned}$$

$$
\begin{aligned}
&\le 4v_n^d \exp\left\{-\frac{M^2 n a_n^2}{16(C_3 h^{-d} + \frac{1}{3} C_1 M h^{-d})}\right\} + 32 v_n^d \frac{C_1 \tau_n}{M a_n h^d} A(a_n \tau_n)^\beta \\
&= I_{21} + I_{22},
\end{aligned}
\tag{4.20}
$$

where the value of M will be given in (4.22).

In view of $0 < \theta < 1$, $s > 2$, $h = n^{-\theta/d}$, and $a_n = (\frac{\ln n}{nh^d})^{1/2}$, one has $h^{-d(d+1)} = n^{\theta(d+1)}$ and $a_n^{-\frac{sd}{s-1}} = (\ln n)^{-\frac{sd}{2(s-1)}} n^{\frac{sd(1-\theta)}{2(s-1)}}$. Therefore, by $v_n = \lfloor \frac{\tau_n}{h^{d+1} a_n} \rfloor + 1$ and $\tau_n = a_n^{-\frac{1}{s-1}}$, we obtain for n sufficiently large

$$
\begin{aligned}
I_{21} &= 4v_n^d \exp\left\{-\frac{M^2 n a_n^2}{16(C_3 h^{-d} + \frac{1}{3} M C_1 h^{-d})}\right\} \\
&\le C_4 h^{-d(d+1)} a_n^{-\frac{sd}{s-1}} \exp\left\{-\frac{M^2 \ln n}{16(C_3 + \frac{1}{3} M C_1)}\right\} \\
&\le C_5 (\ln n)^{-\frac{sd}{2(s-1)}} n^{\theta(d+1) + \frac{sd(1-\theta)}{2(s-1)} - \frac{M^2}{16(C_3 + \frac{1}{3} M C_1)}} = o(1),
\end{aligned}
\tag{4.21}
$$

where M is sufficiently large such that

$$
\frac{M^2}{16(C_3 + \frac{1}{3} M C_1)} \ge \theta(d+1) + \frac{sd(1-\theta)}{2(s-1)}. \tag{4.22}
$$

Meanwhile, by (2.11) and $h = n^{-\theta/d}$, one has for n sufficiently large

$$
\begin{aligned}
I_{22} &= 32 v_n^d \frac{C_1 \tau_n}{M a_n h^d} A(a_n \tau_n)^\beta \le \frac{C_6}{M} \left(\frac{\tau_n}{h^{d+1} a_n}\right)^d a_n^{\frac{\beta(s-2)-s}{s-1}} h^{-d} = \frac{C_6}{M} a_n^{\frac{\beta(s-2)-s(d+1)}{s-1}} h^{-d(d+2)} \\
&= \frac{C_6}{M} (\ln n)^{\frac{\beta(s-2)-s(d+1)}{2(s-1)}} n^{\frac{\beta(\theta-1)(s-2)+s\theta d+3\theta s+sd+s-2\theta d-4\theta}{2(s-1)}} = o(1),
\end{aligned}
\tag{4.23}
$$

in which is used the fact that $s > 2$, $0 < \theta < 1$,

$$
\beta > \max\left\{\frac{s\theta d + 3\theta s + sd + s - 2\theta d - 4\theta}{(1-\theta)(s-2)}, \frac{2s-2}{s-2}\right\},
$$

and

$$
\frac{\beta(\theta-1)(s-2) + s\theta d + 3\theta s + sd + s - 2\theta d - 4\theta}{2(s-1)} < 0.
$$

Thus, by (4.20)-(4.23), we establish that

$$
|I_2| = O_p(a_n). \tag{4.24}
$$

Finally, the result of (2.12) follows from (4.12)-(4.16), (4.18), (4.19), and (4.24) immediately. □

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Acknowledgements
The authors are deeply grateful to the editors and the anonymous referees for their careful reading and insightful comments, which helped in improving the earlier version of this paper. This work is supported by the National Natural Science Foundation of China (11171001, 11426032, 11501005), Natural Science Foundation of Anhui Province (1408085QA02, 1508085QA01, 1508085J06, 1608085QA02), the Provincial Natural Science Research Project of Anhui Colleges (KJ2014A010, KJ2014A020, KJ2015A065), Quality Engineering Project of Anhui Province (2015jyxm054), Applied Teaching Model Curriculum of Anhui University (XJYYKC1401, ZLTS2015053) and Doctoral Research Start-up Funds Projects of Anhui University.

Received: 28 January 2016 Accepted: 13 May 2016 Published online: 27 May 2016

References

1. Shen, J, Xie, Y: Strong consistency of the internal estimator of nonparametric regression with dependent data. Stat. Probab. Lett. **83**, 1915-1925 (2013)
2. Nadaraya, EA: On estimating regression. Theory Probab. Appl. **9**, 141-142 (1964)
3. Watson, GS: Smooth regression analysis. Sankhya, Ser. A **26**, 359-372 (1964)
4. Jones, MC, Davies, SJ, Park, BU: Versions of kernel-type regression estimators. J. Am. Stat. Assoc. **89**, 825-832 (1994)
5. Mack, YP, Müller, HG: Derivative estimation in nonparametric regression with random predictor variable. Sankhya, Ser. A **51**, 59-72 (1989)
6. Linton, O, Nielsen, J: A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika **82**, 93-100 (1995)
7. Linton, O, Jacho-Chávez, D: On internally corrected and symmetrized kernel estimators for nonparametric regression. Test **19**, 166-186 (2010)
8. Masry, E: Recursive probability density estimation for weakly dependent stationary processes. IEEE Trans. Inf. Theory **32**, 254-267 (1986)
9. Roussas, GG, Tran, LT, Ioannides, DA: Fixed design regression for time series: asymptotic normality. J. Multivar. Anal. **40**, 262-291 (1992)
10. Tran, LT, Roussas, GG, Yakowitz, S, Van, BT: Fixed-design regression for linear time series. Ann. Stat. **24**, 975-991 (1996)
11. Liebscher, E: Strong convergence of sums of α-mixing random variables with applications to density estimation. Stoch. Process. Appl. **65**, 69-80 (1996)
12. Hansen, BE: Uniform convergence rates for kernel estimation with dependent data. Econom. Theory **24**, 726-748 (2008)
13. Andrews, DWK: Nonparametric kernel estimation for semiparametric models. Econom. Theory **11**, 560-596 (1995)
14. Bakouch, HS, Ristić, MM, Sandhya, E, Satheesh, S: Random products and product auto-regression. Filomat **27**, 1197-1203 (2013)
15. Bosq, D: Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, 2nd edn. Lecture Notes in Statistics, vol. 110. Springer, Berlin (1998)
16. Bosq, D, Blanke, D: Inference and Prediction in Large Dimensions. Wiley, Chichester (2007)
17. Fan, JQ: Design-adaptive nonparametric regression. J. Am. Stat. Assoc. **87**, 998-1004 (1992)
18. Fan, JQ: Local linear regression smoothers and their minimax efficiency. Ann. Stat. **21**, 196-216 (1993)
19. Fan, JQ, Yao, QW: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York (2003)
20. Gao, JT, Lu, ZD, Tjostheim, D: Estimation in semiparametric spatial regression. Ann. Stat. **34**, 1395-1435 (2006)
21. Gao, JT, Tjostheim, D, Yin, JY: Estimation in threshold autoregressive models with a stationary and a unit root regime. J. Econom. **172**, 1-13 (2013)
22. Györfi, L, Kohler, M, Krzyżak, A, Walk, H: A Distribution-Free Theory of Nonparametric Regression. Springer, New York (2002)
23. Ling, NX, Wu, YH: Consistency of modified regression estimation for functional data. Statistics **46**, 149-158 (2012)
24. Ling, NX, Xu, Q: Asymptotic normality of conditional density estimation in the single index model for functional time series data. Stat. Probab. Lett. **82**, 2235-2243 (2012)
25. Ling, NX, Li, ZH, Yang, WZ: Conditional density estimation in the single functional index model for α-mixing functional data. Commun. Stat., Theory Methods **43**, 441-454 (2014)
26. Masry, E: Multivariate local polynomial regression for time series: uniform strong consistency and rates. J. Time Ser. Anal. **17**, 571-599 (1996)
27. Newey, WK: Kernel estimation of partial means and a generalized variance estimator. Econom. Theory **10**, 233-253 (1994)
28. Peligrad, M: Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence conditions. Stoch. Stoch. Rep. **40**, 147-168 (1992)
29. Rosenblatt, M: Remarks on some non-parametric estimates of a density function. Ann. Math. Stat. **27**, 832-837 (1956)
30. Stone, CJ: Consistent nonparametric regression. Ann. Stat. **5**, 595-645 (1977)
31. Hall, P, Heyde, CC: Martingale Limit Theory and Its Application. Academic Press, New York (1980)
32. Liebscher, E: Estimation of the density and the regression function under mixing conditions. Stat. Decis. **19**, 9-26 (2001)