Ann Inst Stat Math (2019) 71:553–576
https://doi.org/10.1007/s10463-018-0653-1

Regression estimation under strong mixing data

Huijun Guo[1] · **Youming Liu**[1]

Received: 28 July 2016 / Revised: 30 July 2017 / Published online: 27 February 2018

Abstract This paper studies multivariate wavelet regression estimators with errors-in-variables under strong mixing data. We firstly prove the strong consistency for non-oscillating and Fourier-oscillating noises. Then, a convergence rate is provided for non-oscillating noises, when an estimated function has some smoothness. Finally, the consistency and convergence rate are discussed for a practical wavelet estimator.

Keywords Regression estimation · Errors-in-variables · Strong mixing · Practical estimator · Wavelets

1 Introduction

The current paper considers the following errors-in-variables regression problem. Let data $(W_j, Y_j) \in \mathbb{R}^d \times [-T, T]$ ($j = 1, 2, \ldots, n$ and $T > 0$) be from the model

$$Y_j = m(X_j) + \varepsilon_j, \quad W_j = X_j + \delta_j. \tag{1}$$

The errors ε_j and δ_j are independent of each other and independent of X_j. The functions f_X (unknown) and f_δ (known) denote the densities of X_j and δ_j, respectively. The regression errors ε_j satisfy $E\varepsilon_j = 0$ and $E\varepsilon_j^2 < \infty$. The goal is to estimate the regression function m by some estimator $\hat{m}_n$ (depending on (W_j, Y_j), $j = 1, 2, \ldots, n$).

✉ Youming Liu
liuym@bjut.edu.cn

Huijun Guo
guohuijun@emails.bjut.edu.cn

[1] Department of Applied Mathematics, Beijing University of Technology, Pingle Yuan 100, Beijing 100124, People's Republic of China

This above model has many applications in the field of medical statistics (Carroll et al. 2006, 2007) and econometrics (Schennach 2004). For a special case $\delta_j = 0$, the Nadaraya–Watson estimator works well. By deconvolution technique, Fan and Truong (1993) extend the Nadaraya–Watson estimation to the regression model with errors-in-variables.

Chesneau (2010) studies firstly that model by wavelet method. As a generalization of Meister's theorem (Meister 2009), the strong consistency of a wavelet estimator is obtained, when f_δ^{ft} has some zeros (Guo and Liu 2017). Here and after, $t \cdot x := \sum_{i=1}^{\mathrm{d}} t_i x_i$ for $t = (t_1, \ldots, t_d)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, and

$$f^{ft}(t) := \int_{\mathbb{R}^d} f(x) e^{it \cdot x} \mathrm{d}x \tag{2}$$

denotes the Fourier transform of $f \in L^1(\mathbb{R}^d)$. A standard method extends that definition to $L^2(\mathbb{R}^d)$ functions. Recently, Chichignoud et al. (2017) show a convergence rate for an adaptive wavelet regression estimator over anisotropic Hölder classes.

All above work assumes the independence of the given data (W_j, Y_j) $(j = 1, 2, \ldots, n)$. Regression estimation with strong mixing data has received a lot of attentions, such as Masry (1993), Shen and Xie (2013), Chaubey et al. (2013) and Chesneau et al. (2015). It should be pointed out that Chesneau (2014) provides more general theorems on wavelet thresholding method under strong mixing data, which mainly concern the mean integrated square error of an estimator. Motivated by those work, we consider the strong consistency and convergence rate of wavelet regression estimators under strong mixing data for model (1).

For a strictly stationary process $\{Z_j, \ j \in \mathbb{Z}\}$, its kth $(k \geq 1)$ strong mixing coefficient is defined by

$$\alpha_Z(k) = \sup_{(A,B) \in \mathcal{F}_Z^{-\infty, \ 0} \times \mathcal{F}_Z^{k, \ +\infty}} |P(A \cap B) - P(A)P(B)|,$$

where $\mathcal{F}_Z^{-\infty, \ 0}$ and $\mathcal{F}_Z^{k, +\infty}$ are the σ-algebras generated by Z_l for $l \leq 0$ and $l \geq k$, respectively. A process $\{Z_j, \ j \in \mathbb{Z}\}$ is said to be strong mixing, if $\lim_{k \to +\infty} \alpha_Z(k) = 0$. Clearly, independent and identically distributed (i.i.d.) data are strong mixing.

Throughout this paper, the observed data $\{(W_j, Y_j), \ j = 1, 2, \ldots, n\}$ are assumed to be geometrically strong mixing, which means with some positive constants μ_0 and μ_1,

$$\alpha_{(W,Y)}(k) \leq \mu_0 \exp(-\mu_1 k)$$

for each $k \geq 1$.

Example 1 Let $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ with

$$\{\varepsilon_t, \ t \in \mathbb{Z}\} \overset{i.i.d.}{\sim} N(0, \ \sigma^2) \text{ and } a_k = \begin{cases} 2^{-k}, & k \geq 0, \\ 0, & k < 0. \end{cases}$$

Then, it can be proved by Theorem 2 and Corollary 1 of Doukhan (1994, p. 58) that $\{X_t,\ t \in \mathbb{Z}\}$ is a geometrically strong mixing sequence.

Example 2 Let $\{\varepsilon(t), t \in \mathbb{Z}\} \overset{i.i.d.}{\sim} N_r(\mathbf{0}, \Sigma)$ (r-dimensional normal distribution) and $\{Y(t),\ t \in \mathbb{Z}\}$ satisfy the auto-regression moving average equation

$$\sum_{i=0}^{p} B(i)Y(t-i) = \sum_{k=0}^{q} A(k)\varepsilon(t-k)$$

with $l \times r$ and $l \times l$ matrices $A(k)$, $B(i)$, respectively, as well as $B(0)$ being the identity matrix. If the absolute values of the zeros of the determinant $\det P(z) := \det \sum_{i=0}^{p} B(i)z^i$ ($z \in \mathbb{C}$) are strictly greater than 1, then $\{Y(t),\ t \in \mathbb{Z}\}$ is geometrically strong mixing (Mokkadem 1988).

To give an example of model (1), we introduce a simple but important lemma, which is also used in our later discussions.

Lemma 1 *Let $\{Z_i, i \in \mathbb{Z}\}$ be a strong mixing sequence valued in $\mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}^m$ be Borel measurable. Then, $\{f(Z_i), i \in \mathbb{Z}\}$ is strong mixing and $\alpha_{f(Z)}(k) \leq \alpha_Z(k)$.*

Proof Clearly, the strict stationarity of $\{f(Z_i), i \in \mathbb{Z}\}$ follows from the same property of $\{Z_i, i \in \mathbb{Z}\}$. Then for $k \in \mathbb{Z}$ and $n \in \mathbb{N}^+$ (the positive integer set),

$$\begin{aligned}\mathcal{F}_{f(Z)}^{k,k+n} &= \sigma(f(Z_k), \ldots,\ f(Z_{k+n}))\\ &= \left\{(f(Z_k), \ldots,\ f(Z_{k+n}))^{-1}(B):\ B \in \mathcal{B}(\mathbb{R}^{m(n+1)})\right\},\end{aligned}$$

where $\mathcal{B}(\mathbb{R}^{m(n+1)})$ stands for the Borel σ-algebra on $\mathbb{R}^{m(n+1)}$. Because the Borel measurability of f implies that of $g(z_k, z_{k+1}, \ \ldots,\ z_{k+n}) := (f(z_k),\ f(z_{k+1}), \ldots, f(z_{k+n}))$, $g^{-1}(B) \in \mathcal{B}(\mathbb{R}^{d(n+1)})$ (for $B \in \mathcal{B}(\mathbb{R}^{m(n+1)})$) and

$$\begin{aligned}\mathcal{F}_{f(Z)}^{k,k+n} &= \left\{(g(Z_k,\ \ldots,\ Z_{k+n}))^{-1}(B):\ B \in \mathcal{B}(\mathbb{R}^{m(n+1)})\right\}\\ &= \left\{(Z_k,\ \ldots,\ Z_{k+n})^{-1}(g^{-1}(B)):\ B \in \mathcal{B}(\mathbb{R}^{m(n+1)})\right\}\\ &\subset \left\{(Z_k,\ \ldots,\ Z_{k+n})^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^{d(n+1)})\right\}\\ &\subset \sigma(Z_k,\ \ldots,\ Z_{k+n}) = \mathcal{F}_Z^{k,k+n}.\end{aligned}$$

Hence, $\alpha_{f(Z)}(k) \leq \alpha_Z(k)$ and $\{f(Z_i), i \in \mathbb{Z}\}$ is strong mixing. □

Example 3 For a fixed $T > 0$ and $z = (z_1,\ z_2, \ldots,\ z_d) \in \mathbb{R}^d$, define $f : \mathbb{R}^d \to \mathbb{R}^d$ by

$$f(z) = zI_{\mathbb{R}^{d-1}\times[-T,T]}(z) := (z_1, \ldots,\ z_{d-1},\ z_d I_{[-T,T]}(z_d)),$$

where $I_{[-T,T]}(\cdot)$ denotes the indicator function on $[-T, T]$. Then, f is a Borel measurable mapping. Choose a strong mixing sequence $\{Z_i,\ i \in \mathbb{Z}\}$ valued in $\mathbb{R}^d$ (say $Z_i = X_i$ or $Y(i)$) in Example 1 or Example 2. According to Lemma 1, $\{f(Z_i), i \in \mathbb{Z}\}$ is strong mixing as well.

This paper is organized as follows: In Sect. 2, we prove the strong consistency of our wavelet estimators for both non-oscillating and Fourier-oscillating noises. Section 3 provides a convergence rate of a wavelet regression estimator for non-oscillating noises, when the estimated function belongs to some Hölder class. The same problem is considered for a practical wavelet estimator in the last section.

2 Strong consistency

This section studies the strong consistency of wavelet regression estimators under strong mixing data. We begin with two lemmas. The first one generalizes a conclusion of Theorem 3.2 in Meister (2009).

Lemma 2 (Guo and Liu 2017) *Let $f \in C(\mathbb{R}^d)$ (the continuous function set on $\mathbb{R}^d$) and $f(t) \neq 0$ for each $t \in \mathbb{R}^d$. Then, there exists a positive sequence $h_n \to 0$ such that*

$$h_n^d \cdot \min_{t\in[-1/h_n,\ 1/h_n]^d} |f(t)| \geq n^{-\frac{1}{4}}$$

holds for sufficiently large n.

We use $\|x\|_\infty$ to denote the L^∞ norm of a measurable and essentially bounded function $x(t)$. For $\tau > 0$, $\lfloor \tau \rfloor$ stands for the largest integer smaller than or equal to τ, while $\lceil \tau \rceil$ does for the smallest integer larger than or equal to τ.

Lemma 3 (Bosq and Blanke 2007) *Let $\{Z_i, i \in \mathbb{Z}\}$ be a strictly stationary and real-valued process with $EZ_i = 0$ and $\sup_{1\leq i\leq n} \|Z_i\|_\infty \leq M_n$ ($M_n > 0$). Then for $\varepsilon > 0$ and $\eta \in [1,\ n/2]$,*

$$P\left(\left|\sum_{i=1}^{n} Z_i\right| > n\varepsilon\right) \leq 4\exp\left(-\frac{n^2\varepsilon^2/\eta}{32\sigma^2(\eta) + \frac{4M_n}{3}n^2\eta^{-2}\varepsilon}\right) + \frac{16M_n}{\varepsilon}\alpha_Z\left(\left\lfloor \frac{n}{2\eta} \right\rfloor\right),$$

where $\sigma^2(\eta) = (\lfloor\tau\rfloor + 2)[Var\ Z_1 + 2\sum_{l=1}^{\lfloor\tau\rfloor+1} |Cov(Z_0, Z_l)|]$ with $\tau = \frac{n}{2\eta}$.

Lemma 3 implies the classical Bernstein inequality, which plays a key role for estimation with i.i.d. data. In fact, when $\{Z_i\}$ are i.i.d., the term containing $\alpha_Z(\cdot)$ disappears. By taking $\eta = \frac{n}{2}$, we find $\tau = 1$ and $\sigma^2(\eta) = 3\text{Var}\ Z_1 := 3\sigma^2$. Then, the estimation of Lemma 3 reduces to

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| > \varepsilon\right) \leq 4\exp\left(-\frac{n\varepsilon^2}{48\sigma^2 + \frac{8}{3}M\varepsilon}\right)$$

with $|Z_i| \leq M$. This is the desired conclusion up to some constants (Härdle et al. 1998).

In order to simplify the proof of Theorem 1, we recall the corresponding result of Guo and Liu (2017) with i.i.d. data.

For a non-oscillating noise (which means $f_\delta^{ft}(t) \neq 0, t \in \mathbb{R}^d$), choose the scaling function $\varphi(x) := \prod_{l=1}^{d} \tilde{\varphi}(x_l)$, where $\tilde{\varphi} := \varphi_M$ stands for the one-dimensional Meyer scaling function with

$$\text{supp } \varphi_M^{ft} \subset \left[-\frac{4\pi}{3}, \frac{4\pi}{3}\right] \text{ and } \varphi_M^{ft} \in C^\infty$$

(Daubechies 1992). As usual in wavelet analysis, denote $\varphi_{j,k}(x) := 2^{dj/2}\varphi(2^j x - k)$ with $j \in \mathbb{Z}$ and $k \in \mathbb{Z}^d$. Then, the estimators for f_X and $p := mf_X$ are defined by

$$f_{X,n}(x) = \sum_{k \in \mathbb{Z}^d} \hat{\alpha}_{j,k}\varphi_{j,k}(x) \quad \text{and} \quad p_n(x) = \sum_{k \in \mathbb{Z}^d} \hat{\gamma}_{j,k}\varphi_{j,k}(x), \tag{3}$$

respectively, with

$$\hat{\alpha}_{j,k} = \frac{1}{n}\sum_{l=1}^{n} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{it\cdot W_l}\overline{[\varphi_{j,k}]^{ft}(t)}/f_\delta^{ft}(t)\mathrm{d}t, \tag{4}$$

$$\hat{\gamma}_{j,k} = \frac{1}{n}\sum_{l=1}^{n} Y_l \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{it\cdot W_l}\overline{[\varphi_{j,k}]^{ft}(t)}/f_\delta^{ft}(t)\mathrm{d}t. \tag{5}$$

Here, the long bar stands for complex conjugate. It is easy to show $E\hat{\alpha}_{j,k} = \alpha_{j,k}$, $E\hat{\gamma}_{j,k} = \gamma_{j,k}$ and

$$Ef_{X,n}(x) = P_j f_X(x) \text{ and } Ep_n(x) = P_j p(x), \tag{6}$$

where

$$\alpha_{j,k} = \int_{\mathbb{R}^d} f_X(x)\varphi_{j,k}(x)\mathrm{d}x \text{ and } \gamma_{j,k} = \int_{\mathbb{R}^d} p(x)\varphi_{j,k}(x)\mathrm{d}x.$$

A Fourier-oscillating noise means that there exists a constant $C > 0$ such that

$$\left|f_\delta^{ft}(t_1, \ldots, t_d)\right| \geq C \prod_{s=1}^{d} \left|\sin\left(\frac{\pi t_s}{\lambda_s}\right)\right|^{\upsilon_s} (1 + |t_s|)^{-\alpha_s} \tag{7}$$

with $\lambda_s > 0$, $\alpha_s \geq 0$ and $\upsilon_s \in \mathbb{Z}^+ \cup \{0\}$. In that case, the authors of Guo and Liu (2017) assume $f_X \in L^2(\mathbb{R}^d)$ and supp $f_X \subset \Omega := [a, b]^d$. They replace the Meyer's function φ_M in former case by the Daubechies' D_{2N} (for large N) to obtain the scaling function φ. Let

$$\mathcal{K}_j := \{k = (k_1, k_2, \ldots, k_d) \in \mathbb{Z}^d : \text{ supp } \varphi_{j,k_s} \cap [a, b] \neq \emptyset, s = 1, 2, \ldots, d\}$$

and $\tilde{J} := (\tilde{J}_1, \tilde{J}_2 \ldots, \tilde{J}_d)$, where $\tilde{J}_s = \lceil \frac{(\tilde{b}-\tilde{a})\lambda_s}{2\pi} \rceil$ with $\tilde{a} := a - 2N + 1$ and $\tilde{b} := b + 2N - 1$. Then, the estimators for f_X and p are defined, respectively, by

$$f_{X,n}(x) := \sum_{k \in \mathcal{K}_j} \hat{\alpha}_{j,k}\varphi_{j,k}(x) \text{ and } p_n(x) := \sum_{k \in \mathcal{K}_j} \hat{\gamma}_{j,k}\varphi_{j,k}(x), \tag{8}$$

respectively, where

$$\hat{\alpha}_{j,k} := \frac{1}{n}\sum_{l=1}^{n} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{\xi}(t) e^{it\cdot W_l} \frac{\overline{[\varphi_{j,k}]^{ft}(t)}}{f_\delta^{ft}(t)} \mathrm{d}t, \tag{9}$$

$$\hat{\gamma}_{j,k} := \frac{1}{n}\sum_{l=1}^{n} Y_l \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{\xi}(t) e^{it\cdot W_l} \frac{\overline{[\varphi_{j,k}]^{ft}(t)}}{f_\delta^{ft}(t)} \mathrm{d}t \tag{10}$$

with

$$\tilde{\xi}(t) = \left(\sum_{m=\mathbf{0}}^{\tilde{J}} \eta_m e^{i\frac{2\pi m}{\lambda}\cdot t}\right) \left[\prod_{s=1}^{d} (e^{\frac{2\pi i t_s}{\lambda_s}} - 1)^{v_s}\right]. \tag{11}$$

Here, $\sum_{m=\mathbf{0}}^{\tilde{J}} := \sum_{m_1=0}^{\tilde{J}_1} \sum_{m_2=0}^{\tilde{J}_2} \cdots \sum_{m_d=0}^{\tilde{J}_d}$ and the constants η_m depend only on $\tilde{J}$, v, d.

The main theorem in Guo and Liu (2017) states as follows:

For problem (1) with i.i.d. data $(W_j, Y_j)(Y_j \in \mathbb{R})$, if $p := mf_X \in L(\mathbb{R}^d)$, $E|Y_1|^4 < \infty$ and x is a Lebesgue point of p and f_X ($f_X(x) \neq 0$), then the following assertions hold true:

(a) When f_δ^{ft} has no zeros on $\mathbb{R}^d$, $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ given by (3)–(5) satisfies $\lim_{n\to\infty} \hat{m}_n(x) \overset{a.s.}{=} m(x)$;

(b) When (7) holds and f_X, $p \in L^2(\mathbb{R}^d)$ have compact support Ω, $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ given by (8)–(11) satisfies $\lim_{n\to\infty} \hat{m}_n(x) \overset{a.s.}{=} m(x)$ $(x \in \Omega)$ as well.

Our Theorem 1 extends this above results from i.i.d. to geometrically strong mixing data. Since we assume $Y \in [-T, T]$, $E|Y_1|^4 < +\infty$ and $p := mf_X \in L(\mathbb{R}^d)$ hold automatically.

In the proofs of Theorem 1—3, we frequently use the following notations. For two variables A and B, $A \lesssim B$ denotes $A \leq CB$ for some positive constant C in later discussions; $A \gtrsim B$ means $B \lesssim A$; we use $A \sim B$ to stand for both $A \lesssim B$ and $B \lesssim A$.

Theorem 1 *For problem* (1) *with* $\alpha_{(W,Y)}(k) \leq \mu_0 \exp(-\mu_1 k)$ $(\mu_0, \mu_1 > 0)$*, if x is a Lebesgue point of* $p := mf_X$ *and* $f_X(f_X(x) \neq 0)$*, then the following assertions hold true:*

(a) When f_δ^{ft} *has no zeros on* $\mathbb{R}^d$, $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ *given by* (3)–(5) *satisfies* $\lim_{n\to\infty} \hat{m}_n(x) \overset{a.s.}{=} m(x)$;

(b) When (7) *holds and* $f_X \in L^2(\mathbb{R}^d)$ *has compact support* Ω*,* $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ *given by* (8)–(11) *satisfies* $\lim_{n\to\infty} \hat{m}_n(x) \overset{a.s.}{=} m(x)$ *for* $x \in \Omega$*.*

Proof Since $\hat{m}_n(x) = \frac{p_n(x)}{f_{X,n}(x)}$ and $m(x) = \frac{p(x)}{f_X(x)}$, it suffices to show $\lim_{n\to\infty} p_n(x) \overset{a.s.}{=} p(x)$ and $\lim_{n\to\infty} f_{X,n}(x) \overset{a.s.}{=} f_X(x)$. From the definitions of $p_n(x)$ and $f_{X,n}(x)$, one need only prove the first limit. For $\varepsilon > 0$,

$$\begin{aligned} &P(|p_n(x) - p(x)| > 4\varepsilon) \\ &\quad \le P(|p_n(x) - Ep_n(x)| > 2\varepsilon) + P(|Ep_n(x) - p(x)| > 2\varepsilon). \end{aligned} \tag{12}$$

By the same arguments as Theorem 4 and 5 of Guo and Liu (2017), $\lim_{n\to\infty} Ep_n(x) = p(x)$ holds for each Lebesgue point of p in both cases (a) and (b). In fact, careful observations find that the whole proof does not use the independence of (W_j, Y_j). Instead, it does use the identical distribution property, which is fortunately implied by the strict stationarity of (W_j, Y_j).

The independence of (W_j, Y_j) plays a key role for the estimate of $P(|p_n(x) - Ep_n(x)| > 2\varepsilon)$ in the proof of Theorem 4, 5 (Guo and Liu 2017). In fact, the authors estimate $E|p_n(x) - Ep_n(x)|^4$ for an upper bound of $P(|p_n(x) - Ep_n(x)| > 2\varepsilon)$. In this current case, one estimates $P(|p_n(x) - Ep_n(x)| > 2\varepsilon)$ directly for both cases (a) and (b).

(a) By (3) and (5),

$$p_n(x) := \frac{1}{n}\sum_{l=1}^{n} \frac{Y_l}{(2\pi)^d} \int_{\mathbb{R}^d} 2^{-dj/2} e^{it\cdot W_l} \left[\sum_{k\in\mathbb{Z}^d} e^{-i2^{-j}k\cdot t} \varphi_{j,k}(x)\right] \frac{\overline{\varphi^{ft}(2^{-j}t)}}{f_\delta^{ft}(t)} \mathrm{d}t. \tag{13}$$

With $U_l := Z_l - EZ_l$ and

$$Z_l := \frac{Y_l}{(2\pi)^d} \int_{\mathbb{R}^d} 2^{-dj/2} e^{it\cdot W_l} \left[\sum_{k\in\mathbb{Z}^d} e^{-i2^{-j}k\cdot t} \varphi_{j,k}(x)\right] \frac{\overline{\varphi^{ft}(2^{-j}t)}}{f_\delta^{ft}(t)} \mathrm{d}t,$$

$p_n(x) - Ep_n(x) = \frac{1}{n}\sum_{l=1}^{n} U_l$ and

$$P(|p_n(x) - Ep_n(x)| > 2\varepsilon) = P\left(\left|\frac{1}{n}\sum_{l=1}^{n} U_l\right| > 2\varepsilon\right). \tag{14}$$

Obviously, $EU_l = 0$. According to $|Y_l| \le T$, $|\sum_{k\in\mathbb{Z}^d} \varphi_{j,k}(x)| \lesssim 2^{dj/2}$ and supp $\varphi^{ft} \subset [-\frac{4\pi}{3}, \frac{4\pi}{3}]^d$,

$$|Z_l| \lesssim 2^{jd} \left[\min_{t\in[-4\pi 2^j/3,\ 4\pi 2^j/3]^d} \left|f_\delta^{ft}(t)\right|\right]^{-1}. \tag{15}$$

By $f_\delta^{ft}(t) \neq 0$ and Lemma 2, there exists a positive sequence $h_n \to 0$ such that

$$h_n^d \min_{t\in[-1/h_n,\ 1/h_n]^d} \left|f_\delta^{ft}(t)\right| \geq n^{-\frac{1}{4}}.$$

Since $h_n \to 0$, $\frac{3}{4\pi h_n} > 1$ for large n and

$$j := \left\lfloor \log_2\left(\frac{3}{4\pi h_n}\right)\right\rfloor > 0. \tag{16}$$

Clearly, $j \leq \log_2(\frac{3}{4\pi h_n})$ and $\frac{4\pi}{3}2^j \leq \frac{1}{h_n}$. Then, (15) reduces to

$$|Z_l| \lesssim \left[h_n^d \min_{t\in[-1/h_n,\ 1/h_n]^d} |f_\delta^{ft}(t)|\right]^{-1} \lesssim n^{1/4}.$$

Hence, $|U_l| \lesssim n^{1/4}$, $Var\ U_l \lesssim n^{1/2}$ and $|Cov\ (U_0, U_l)| \lesssim n^{1/2}$.

Note that $|\sum_{k\in\mathbb{Z}^d} \varphi_{j,k}(x)| \lesssim 2^{dj/2}$, supp $\varphi^{ft} \subset [-\frac{4\pi}{3}, \frac{4\pi}{3}]^d$ and $f_\delta^{ft}(t) \neq 0$ ($t \in \mathbb{R}^d$). Then as a function of t, $\sum_{k\in\mathbb{Z}^d} e^{-i2^{-j}k\cdot t}\varphi_{j,k}(x)\overline{\varphi^{ft}(2^{-j}t)}/f_\delta^{ft}(t) \in L(\mathbb{R}^d)$ for fixed $x \in \mathbb{R}^d$, and its Fourier transform should be continuous. Furthermore,

$$g(w, y) := y(2\pi)^{-d} \int_{\mathbb{R}^d} 2^{-dj/2} e^{it\cdot w} \left[\sum_{k\in\mathbb{Z}^d} e^{-i2^{-j}k\cdot t}\varphi_{j,k}(x)\right] \overline{\varphi^{ft}(2^{-j}t)}/f_\delta^{ft}(t)\mathrm{d}t$$

defines a continuous function on $\mathbb{R}^d \times [-T, T]$, so does Re $g(w, y)$. It follows from $U_l := g(W_l, Y_l) - Eg(W_l, Y_l)$ and Lemma 1 that {Re U_l, $l \in \mathbb{Z}$} is strong mixing. With the given assumption $\alpha_{(W,Y)}(q) \leq \mu_0 e^{-\mu_1 q}$, one chooses $\tau_0 > \frac{2}{\mu_1}$ and denotes $\eta := \lfloor \frac{n}{2\tau_0 \ln n}\rfloor$, $\tau := \frac{n}{2\eta}$. Since $|\text{Re}\ U_l| \leq |U_l| \lesssim n^{1/4}$, $|\text{Re}\ U_l| \leq c_1 n^{1/4} := M_n$ for some $c_1 > 0$. Then, Lemma 3 tells that

$$P\left(\left|\frac{1}{n}\sum_{l=1}^n \text{Re}\ U_l\right| > \varepsilon\right) \leq P_{1,n} + P_{2,n}, \tag{17}$$

where $P_{1,n} := 4\exp(-\frac{n^2\varepsilon^2/\eta}{32\sigma^2(\eta)+\frac{4M_n}{3}n^2\eta^{-2}\varepsilon})$, $P_{2,n} := \frac{16M_n}{\varepsilon}\alpha_{\text{Re}\,U}(\lfloor\frac{n}{2\eta}\rfloor)$ and

$$\sigma^2(\eta) = (\lfloor\tau\rfloor + 2)\left[Var(\text{Re}\ U_1) + 2\sum_{l=1}^{\lfloor\tau\rfloor+1} |Cov(\text{Re}\ U_0,\ \text{Re}\ U_l)|\right].$$

According to Lemma 1, $\alpha_{\text{Re}\,U}(\lfloor\frac{n}{2\eta}\rfloor) \leq \alpha_{(W,Y)}(\lfloor\frac{n}{2\eta}\rfloor) \leq \mu_0 e^{-\mu_1\lfloor\frac{n}{2\eta}\rfloor}$. Because $\eta := \lfloor\frac{n}{2\tau_0 \ln n}\rfloor$ implies $\frac{n}{2\eta} \geq \tau_0 \ln n$, $\alpha_{\text{Re}\,U}(\lfloor\frac{n}{2\eta}\rfloor) \leq \mu_0 e^{-\mu_1(\tau_0 \ln n - 1)} = \mu_0 e^{\mu_1} n^{-\mu_1\tau_0}$. Furthermore,

$$P_{2,n} \leq \frac{16}{\varepsilon} c_1 \mu_0 e^{\mu_1} n^{-(\mu_1 \tau_0 - 1/4)} \lesssim n^{-7/4} \tag{18}$$

thanks to $M_n = c_1 n^{1/4}$ and $\tau_0 > \frac{2}{\mu_1}$.

Since $Var(\text{Re } U_l) \lesssim n^{1/2}$ and $|Cov(\text{Re } U_0, \text{ Re } U_l)| \lesssim n^{1/2}$, $\sigma^2(\eta) = (\lfloor\tau\rfloor + 2)[Var(\text{Re } U_1) + 2\sum_{l=1}^{\lfloor\tau\rfloor+1} |Cov(\text{Re } U_0, \text{ Re } U_l)|] \lesssim (\lfloor\tau\rfloor + 2)(2\lfloor\tau\rfloor + 3)n^{1/2} \leq c_0\tau^2 n^{1/2}$ with some constant $c_0 > 0$. This with $\tau = \frac{n}{2\eta}$ and $M_n = c_1 n^{1/4}$ shows

$$P_{1,n} \leq 4\exp\left(-\frac{n^2\varepsilon^2/\eta}{32c_0(\frac{n}{2\eta})^2 n^{1/2} + \frac{4}{3}c_1 n^{1/4} n^2 \eta^{-2}\varepsilon}\right) \leq 4\exp\left(-\frac{\varepsilon^2 n^{-\frac{1}{2}}}{8c_0 + \frac{4}{3}c_1\varepsilon}\eta\right).$$

By $\eta = \lfloor\frac{n}{2\tau_0 \ln n}\rfloor \geq \frac{n}{4\tau_0 \ln n}$ for n large enough, one obtains

$$P_{1,n} \leq 4\exp\left(-\frac{\varepsilon^2 n^{-\frac{1}{2}}}{8c_0 + \frac{4}{3}c_1\varepsilon} \cdot \frac{n}{4\tau_0 \ln n}\right) \lesssim n^{-7/4}. \tag{19}$$

Therefore, it follows from (17)–(19) that $P(|\frac{1}{n}\sum_{l=1}^{n} \text{Re } U_l| > \varepsilon) \lesssim n^{-\frac{7}{4}}$. Similarly, $P(|\frac{1}{n}\sum_{l=1}^{n} \text{Im } U_l| > \varepsilon) \lesssim n^{-\frac{7}{4}}$. Hence, (14) reduces to

$$P(|p_n(x) - Ep_n(x)| > 2\varepsilon) \lesssim n^{-\frac{7}{4}}. \tag{20}$$

This with $\lim_{n\to+\infty} Ep_n(x) = p(x)$ leads to $P(\ |p_n(x) - p(x)| > 4\varepsilon\ ) \lesssim n^{-\frac{7}{4}}$ for n large enough. Then, the conclusion $\lim_{n\to\infty} p_n(x) \overset{a.s.}{=} p(x)$ reaches thanks to Borel–Cantelli lemma. The proof of (a) is done.

(b) The condition $f_X \in L^2(\mathbb{R}^d)$ is needed for the definition of $\hat{m}_n$, when f_δ^{ft} has some zeros (see (16) in Guo and Liu 2017). Similar to case (a), one defines $U_l := Z_l - EZ_l$ and

$$Z_l := Y_l \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} 2^{-dj/2} \tilde{\xi}(t) e^{it\cdot W_l} \left[\sum_{k\in\mathcal{K}_j} e^{-i2^{-j}k\cdot t} \varphi_{j,k}(x)\right] \overline{\varphi^{ft}(2^{-j}t)}/f_\delta^{ft}(t)\mathrm{d}t$$

with $\tilde{\xi}(t)$ given in (11). Then, $EU_l = 0$ and

$$|Z_l| \lesssim \int_{\mathbb{R}^d} |\tilde{\xi}(t)\overline{\varphi^{ft}(2^{-j}t)}|/|f_\delta^{ft}(t)|\mathrm{d}t, \tag{21}$$

because $|Y_l| \leq T$ and $|\sum_{k\in\mathcal{K}_j} \varphi_{j,k}(x)| \lesssim 2^{dj/2}$. Recall that $\tilde{J}$ is fixed and independent of t, the constants η_k are only dependent of $\tilde{J},\ v,\ d$. Then, (11) implies

$$|\tilde{\xi}(t)\Bigg| \lesssim \prod_{s=1}^{d} |\exp\left(\frac{2\pi i t_s}{\lambda_s}\right) - 1\Bigg|^{v_s}, \tag{22}$$

where the compactness of f_X is needed to guarantee $\tilde{J}$ independent of t. This with (21) and (7) shows that

$$|Z_l| \lesssim \int_{\mathbb{R}^d} \prod_{s=1}^{d} (1+|t_s|)^{\alpha_s} \left|\varphi_N^{ft}(2^{-j}t_s)\right| \mathrm{d}t$$
$$\lesssim \prod_{s=1}^{d} 2^j \int_{\mathbb{R}} (1+|2^j t_s|)^{\alpha_s} \left|\varphi_N^{ft}(t_s)\right| \mathrm{d}t_s \lesssim 2^{\sum_{s=1}^{d}(1+\alpha_s)j}$$

holds for large N. Here, $\varphi^{ft}(t) = \prod_{s=1}^{d} \varphi_N^{ft}(t_s)$ because φ is defined by the tensor product of φ_N.

By choosing j with

$$2^{\sum_{s=1}^{d}(1+\alpha_s)j} \backsimeq n^{\frac{1}{4}}, \tag{23}$$

one gets $|Z_l| \lesssim n^{1/4}$ and $|U_l| \lesssim n^{1/4}$. Moreover, $Var\ U_l \lesssim n^{1/2}$ and $|Cov(U_0, U_l)| \lesssim n^{1/2}$. Then, it follows from (22) and $|\sum_{k\in\mathcal{K}_j} \varphi_{j,k}(x)| \lesssim 2^{dj/2}$ that

$$|2^{-dj/2}\tilde{\xi}(t)[\sum_{k\in\mathcal{K}_j} e^{-i2^{-j}k\cdot t}\varphi_{j,k}(x)]\overline{\varphi^{ft}(2^{-j}t)}/f_\delta^{ft}(t)|$$
$$\lesssim \left[\prod_{s=1}^{d} \left|\exp\left(\frac{2\pi i t_s}{\lambda_s}\right) - 1\right|^{v_s}\right] |\overline{\varphi^{ft}(2^{-j}t)}|/|f_\delta^{ft}(t)|.$$

According to (7) and $\exp(\frac{2\pi i t_s}{\lambda_s}) - 1 = \exp\left(\frac{\pi i t_s}{\lambda_s}\right) \cdot 2i \sin\frac{\pi i t_s}{\lambda_s}$, $2^{-dj/2}\tilde{\xi}(t)[\sum_{k\in\mathcal{K}_j} e^{-i2^{-j}k\cdot t}\varphi_{j,k}(x)]\overline{\varphi^{ft}(2^{-j}t)}/f_\delta^{ft}(t) \in L(\mathbb{R}^d)$ for fixed x and j. Hence, its Fourier transform is continuous and

$$g(w, y) := y(2\pi)^{-d} \int_{\mathbb{R}^d} 2^{-dj/2}\tilde{\xi}(t)e^{it\cdot w}\left[\sum_{k\in\mathcal{K}_j} e^{-i2^{-j}k\cdot t}\varphi_{j,k}(x)\right]\overline{\varphi^{ft}(2^{-j}t)}/f_\delta^{ft}(t)\mathrm{d}t$$

defines a continuous function on $\mathbb{R}^d \times [-T,\ T]$. The remaining proofs are the same as those in case (a). □

Remark 1 From the choices of j in (16) and (23), we find that j goes to $+\infty$, as $n \to \infty$.

Compared with Theorem 4, 5 in Guo and Liu (2017), the estimates of the bias term are similar; the stochastic term $P\{|p_n(x) - Ep_n(x)| > 2\varepsilon\}$ is done by estimating an upper bound of $E|p_n(x) - Ep_n(x)|^4$ in i.i.d. case, where the independence plays a key role. However, it does not seem work in strong mixing situation. Therefore, we use the generalized Bernstein inequality (Lemma 3).

3 Convergence rate

In this part, we provide a convergence rate of $\hat{m}_n(x) = \frac{p_n(x)}{f_{X,n}(x)}$ defined by (3)–(5), when the estimated function $m(x)$ is in some Hölder class and the noise density f_δ satisfies that

$$|f_\delta^{ft}(t)| \gtrsim \prod_{i=1}^{d}(1+|t_i|)^{-\alpha_i} \text{ and } |\partial^\beta f_\delta^{ft}(t)| \lesssim \prod_{i=1}^{d}(1+|t_i|)^{-\alpha_i-\beta_i}, \tag{24}$$

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$ $(\alpha_i > 2)$ and $\beta = (\beta_1,\ \beta_2, \dots, \beta_d)$ $(\beta_i = 0,\ 1,\ 2)$ for $i = 1,\ 2, \dots, d$.

Let $f_{(W_0,Y_0,W_l,Y_l)}$ and $f_{(W,Y)}$ stand for the density of $(W_0, Y_0, W_l, Y_l)(l \in \mathbb{Z})$ and (W, Y), respectively, and

$$\sup_{l\in\mathbb{Z}} \sup_{\substack{(w_0,\, y_0) \in \mathbb{R}^d \times [-T,\, T] \\ (w_l,\, y_l) \in \mathbb{R}^d \times [-T,\, T]}} |h(w_0, y_0, w_l, y_l)| \le C \tag{25}$$

with constant $C > 0$ and

$$h(w_0, y_0, w_l, y_l) := f_{(W_0,Y_0,W_l,Y_l)}(w_0, y_0, w_l, y_l) - f_{(W,Y)}(w_0, y_0) f_{(W,Y)}(w_l, y_l).$$

Our Theorem 2 can be considered as an extension of Theorem 3.3 in Meister (2009).

Recall that the estimator $\hat{m}_n(x)$ is defined by the scaling function $\varphi(x) = \prod_{s=1}^{d} \tilde{\varphi}(x_l)$ (see (3)–(5)), where $\tilde{\varphi}$ stands for the one-dimensional Meyer's function. Then, the kernel function $\tilde{K}(x, y) := \sum_{k\in\mathbb{Z}} \tilde{\varphi}(x-k)\tilde{\varphi}(y-k)$ satisfies $\lfloor s \rfloor$-order moment condition $\int_{\mathbb{R}} \tilde{K}(x, y)(y-x)^j \mathrm{d}y = \delta_{0,j},\ j = 0, 1, \dots, \lfloor s \rfloor$. Moreover, $K(x, y) := \prod_{l=1}^{d} \tilde{K}(x_l, y_l)$ has the same property,

$$\int_{\mathbb{R}^d} K(x, y)(y-x)^\gamma \mathrm{d}y = \delta_{0,|\gamma|} \quad (|\gamma| = 0, 1, \dots, \lfloor s \rfloor). \tag{26}$$

Since $\tilde{\varphi}$ is the Meyer's scaling function, there exists a bounded and radically decreasing L^1 function Φ such that

$$|K(x, y)| \le C\Phi\left(\frac{|x-y|}{2}\right) \text{ (a.e.)} \quad \text{and} \quad \int_{\mathbb{R}^d} \Phi(|u|)|u|^s \mathrm{d}u < +\infty \tag{27}$$

with some positive constant C depending only on Φ, see Kelly et al. (1994).

The next lemma plays an important role in the proof of Theorem 2. It is a generalization of Lemma 6 in Fan and Koo (2002) from $d = 1$ to $d \ge 1$.

Lemma 4 *Let* $(K^*\varphi)_{j,k}(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{ix\cdot t}\overline{[\varphi_{j,k}]^{ft}(t)}/f_\delta^{ft}(t)\mathrm{d}t$ *and* (24) *hold. Then with* $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$,

$$|(K^*\varphi)_{j,k}(x)| \lesssim 2^{j(|\alpha|+d/2)} \prod_{l=1}^{d} \left(1 + |2^j x_l - k_l|\right)^{-2}.$$

Proof Denote $(K_j^-\varphi)(x) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{ix\cdot t}\varphi^{ft}(-t)/f_\delta^{ft}(2^j t)\mathrm{d}t$. Then, $(K^*\varphi)_{j,k}(x) = 2^{jd/2}(K_j^-\varphi)(2^j x - k)$. It is sufficient to prove

$$|(K_j^-\varphi)(x)| \lesssim 2^{j|\alpha|} \prod_{l=1}^{d} (1 + |x_l|)^{-2}. \tag{28}$$

Let $\mathcal{L} := \{l : \ 1 \le l \le d, \ |x_l| > 1\}$ and $\mathcal{L}^c := \{1, \ 2, \ \ldots, \ d\}\backslash\mathcal{L}$. Then by integral by parts twice with respect to t_l $(l \in \mathcal{L})$,

$$(K_j^-\varphi)(x) = \frac{\prod_{l\in\mathcal{L}}(ix_l)^{-2}}{(2\pi)^d} \int_{\mathbb{R}^d} e^{it\cdot x} \partial^{\tilde{\beta}} \left[\varphi^{ft}(-t)/f_\delta^{ft}(2^j t)\right] \mathrm{d}t,$$

where $\tilde{\beta} = (\tilde{\beta}_1, \ \tilde{\beta}_2, \ldots, \ \tilde{\beta}_d)$ with $\tilde{\beta}_l = 2$ for $l \in \mathcal{L}$, while $\tilde{\beta}_l = 0$ for $l \in \mathcal{L}^c$. Because $|x_l|^{-2} \lesssim (1 + |x_l|)^{-2}$ for $|x_l| > 1$,

$$|(K_j^-\varphi)(x)| \lesssim \left[\prod_{l\in\mathcal{L}} (1 + |x_l|)^{-2}\right] \int_{\mathbb{R}^d} |\partial^{\tilde{\beta}} \left[\varphi^{ft}(-t)/f_\delta^{ft}(2^j t)\right]| \mathrm{d}t. \tag{29}$$

For $\beta = (\beta_1, \ \beta_2, \ldots, \beta_d) \in \mathbb{N}^d$ with $\beta_i \le 2$,

$$\left|\partial^\beta \left[f_\delta^{ft}(2^j t)\right]\right| = \left|\left[\partial^\beta f_\delta^{ft}\right](2^j t) \prod_{l=1}^{d} 2^{j\beta_l}\right| \lesssim \prod_{l=1}^{d} \left(1 + |2^j t_l|\right)^{-\alpha_l - \beta_l} 2^{j\beta_l}$$

thanks to the second inequality of (24). According to the first one of (24),

$$\left|\partial^{\tilde{\beta}} \frac{\varphi^{ft}(-t)}{f_\delta^{ft}(2^j t)}\right| \lesssim \sum_{\beta=\mathbf{0}}^{\tilde{\beta}} \left\{\left|\left[\partial^{\tilde{\beta}-\beta}\varphi^{ft}\right](-t)\right| \left[\prod_{l=1}^{d} (1 + |2^j t_l|)^{\alpha_l - \beta_l} 2^{j\beta_l}\right]\right\}$$

with $\sum_{\beta=\mathbf{0}}^{\tilde{\beta}} = \sum_{\beta_1=0}^{\tilde{\beta}_1} \sum_{\beta_2=0}^{\tilde{\beta}_2} \cdots \sum_{\beta_d=0}^{\tilde{\beta}_d}$.

Since $\alpha_l - \beta_l \ge 0$ and $1 + |2^j t_l| \le 2^j + |2^j t_l| = 2^j(1 + |t_l|)$,

$$\left|\partial^{\tilde{\beta}} \frac{\varphi^{ft}(-t)}{f_\delta^{ft}(2^j t)}\right| \lesssim 2^{j|\alpha|} \sum_{\beta=\mathbf{0}}^{\tilde{\beta}} \left|\left[\partial^{\tilde{\beta}-\beta}\varphi^{ft}\right](-t)\right| \prod_{l=1}^{d} (1 + |t_l|)^{\alpha_l}.$$

This with (29) leads to

$$|(K_j^- \varphi)(x)| \lesssim 2^{j|\alpha|} \left[\prod_{l\in\mathcal{L}}(1+|x_l|)^{-2}\right] \int_{\mathbb{R}^d} \sum_{\beta=\mathbf{0}}^{\tilde{\beta}} \left\{\left|\left[\partial^{\tilde{\beta}-\beta}\varphi^{ft}\right](-t)\right| \prod_{l=1}^{d}(1+|t_l|)^{\alpha_l}\right\} \mathrm{d}t.$$

Note that $\varphi^{ft} \in C^\infty(\mathbb{R}^d)$ is compactly supported, and $1 \lesssim (1+|x_l|)^{-2}$ holds for $|x_l| \le 1$. Then, desired conclusion (28) follows. This completes the proof. □

A function f on $\mathbb{R}^d$ is said to satisfy Hölder condition of order s ($s \in \mathbb{R}^+\backslash\mathbb{N}$), if there exists a constant $C > 0$ such that for $|\beta| = \lfloor s \rfloor$ and $y, z \in \mathbb{R}^d$,

$$\left|\partial^\beta f(y) - \partial^\beta f(z)\right| \le C|y-z|^{s-|\beta|}.$$

When the above inequality holds only for $y,\ z \in Q(x,r) := \{y = (y_1,\ y_2, \ldots, y_d)|\ y_i \in (x_i - r, x_i + r), i = 1,\ 2, \ldots,\ d\}$ with $r > 0$, we call f satisfying local Hölder condition of order s at the point x. Obviously, Hölder condition of order s implies local one of the same order at each fixed point $x \in \mathbb{R}^d$. Then, Lemma 2.7 in Meister (2009) can be easily generalized to high-dimensional cases. We state it without giving proof.

Lemma 5 *Let $s \in \mathbb{R}^+\backslash\mathbb{N}$ and $\mathcal{F}$ be a set consisting of local Hölder functions of order s at $x \in \mathbb{R}^d$. If there exist constants $C', C'' > 0$ such that with $\beta \in \mathbb{N}^d$ and $|\beta| = \lfloor s \rfloor$,*

$$|f(y)| \le C' \quad \text{and} \quad |\partial^\beta f(y) - \partial^\beta f(z)| \le C''|y-z|^{s-|\beta|} \tag{30}$$

for each $f \in \mathcal{F}$ and $y,\ z \in Q(x,r)$. Then with some constant $C > 0$ (depending only on C', C'', s and r), $|\partial^\gamma f(y)| \le C$ holds uniformly for $f \in \mathcal{F}$, $y \in Q(x,r)$ and $|\gamma| = 0,\ 1,\ \ldots,\ \lfloor s \rfloor$.

We also need the following lemma in the proof of Theorem 2.

Lemma 6 (Davydov 1970) *Let $\{X_i\}_{i\in\mathbb{Z}}$ be a strong mixing process with the mixing coefficient $\alpha(k)$ ($k \ge 0$) and f be a Borel measurable function. If $E|f(X_0)|^p$ and $E|f(X_0)|^q$ exist for $p, q > 0$ with $\frac{1}{p} + \frac{1}{q} < 1$, then there exists a constant $C > 0$ such that*

$$|Cov(f(X_0), f(X_k))| \le C\alpha(k)^{1-\frac{1}{p}-\frac{1}{q}}[E|f(X_0)|^p]^{\frac{1}{p}}[E|f(X_0)|^q]^{\frac{1}{q}}.$$

Before stating the main theorem of this section, we introduce a functional set $\mathcal{P}_{x,s}$ for $x \in \mathbb{R}^d$ and $s \in \mathbb{R}^+\backslash\mathbb{N}$,

$$\mathcal{P}_{x,s} = \{(m,\ f_X) :\ f_X \text{ and } mf_X \text{ satisfy local Hölder condition of order } s \text{ at } x;\ \|f_X\|_\infty + \|m^2 f_X\|_\infty \le C_1;\ f_X(x) \ge C_2\} \tag{31}$$

for some positive constants $C_1, C_2 > 0$.

When $(m,\ f_X) \in \mathcal{P}_{x,s}$, we find easily that

(i) $mf_X \in L(\mathbb{R}^d)$;
(ii) there exists $C_3 > 0$ such that $\sup_{(m, f_X)\in\mathcal{P}_{x,s}} |m(x)| \le C_3$.

In fact, the conclusion (i) follows directly from $|m(x)| \le T$ (due to $Y \in [-T, T]$) and $f_X \in L(\mathbb{R}^d)$; to see (ii), we realize that both f_X and mf_X are continuous at x because of the local Hölder assumptions. This with $f_X(x) \ge C_2 > 0$ shows the continuity of m at x. On the other hand, $C_2|m^2(x)| \le \|m^2 f_X\|_\infty \le C_1$ implies $|m(x)| \le (C_1 C_2^{-1})^{1/2} := C_3$. Hence, conclusion (ii) holds.

Theorem 2 *Consider problem* (1) *with* $\alpha_{(W,Y)}(k) \le \mu_0 \exp(-\mu_1 k)$ $(k \ge 1)$, (24) *and* (25) *holding. Suppose* $2^j \sim n^{\frac{1}{2s+2|\alpha|+d}}$ *(therefore* j *goes to* $+\infty$ *as* $n \to +\infty$*), then the estimator* $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ *defined by* (3)–(5) *satisfies*

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} \sup_{(m, f_X)\in\mathcal{P}_{x,s}} P\left[|\hat{m}_n(x) - m(x)|^2 \ge c n^{-\frac{2s}{2s+2|\alpha|+d}}\right] = 0$$

with $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$.

Proof According to the proof of Theorem 3.2 in Meister (2009),

$$P\left[\, |\hat{m}_n(x) - m(x)|^2 > \varepsilon\right] \le P\left[\, |p_n(x) - p(x)|^2 \gtrsim \varepsilon\right] + P\left[\, |f_{X,n}(x) - f_X(x)|^2 \gtrsim \varepsilon\right]$$

for $\varepsilon > 0$ small enough. Taking $\varepsilon_n := n^{-\frac{2s}{2s+2|\alpha|+d}} \to 0$ and using Markov inequality, one knows that

$$\sup_{(m, f_X)\in\mathcal{P}_{x,s}} P[\, |\hat{m}_n(x) - m(x)|^2 \ge c\varepsilon_n] \lesssim (c\varepsilon_n)^{-1} \sup_{(m, f_X)\in\mathcal{P}_{x,s}} \left\{E|p_n(x) - p(x)|^2 + E|f_{X,n}(x) - f_X(x)|^2\right\}. \quad (32)$$

In order to estimate $E|p_n(x)-p(x)|^2+E|f_{X,n}(x)-f_X(x)|^2$, it suffices to deal with the variance terms $Var\ p_n(x)$ and $Var\ f_{X,n}(x)$, as well as the bias terms $|Ep_n(x)-p(x)|^2$ and $|Ef_{X,n}(x) - f_X(x)|^2$.

First, one estimates $Var\ p_n(x)$. Similar and even simpler arguments apply to $Var f_{X,n}(x)$. Denote

$$\Psi(w, y) := y\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{it\cdot w} \sum_{k\in\mathbb{Z}^d} \varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(t)}/f_\delta^{ft}(t)\mathrm{d}t.$$

Then, $p_n(x) = \frac{1}{n}\sum_{l=1}^n \Psi(W_l, Y_l)$ and $Var\ p_n(x) = \frac{1}{n^2}\sum_{l=1}^n\sum_{l'=1}^n Cov(\Psi(W_l, Y_l), \Psi(W_{l'}, Y_{l'})) \le \frac{1}{n}Var(\Psi(W_1, Y_1)) + \frac{2}{n^2}\sum_{l=2}^n\sum_{l'=1}^{l-1} |Cov(\Psi(W_l, Y_l), \Psi(W_{l'}, Y_{l'}))|$. By the strict stationarity of $\{(W_l, Y_l)\}_{l\in\mathbb{Z}}$,

$$Var\ p_n(x) \le \frac{1}{n} Var\ \Psi(W_1, Y_1) + \frac{2}{n} \sum_{l=1}^{n} |Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))|. \quad (33)$$

For the first term of (33), one observes

$$Var\ \Psi(W_1, Y_1) \le E \left| Y_1 (2\pi)^{-d} \int_{\mathbb{R}^d} e^{it\cdot W_1} \sum_{k\in\mathbb{Z}^d} \varphi_{j,k}(x) \frac{\overline{(\varphi_{j,k})^{ft}(t)}}{f_\delta^{ft}(t)} \mathrm{d}t \right|^2. \quad (34)$$

According to $W_1 = \delta_1 + X_1$ and the independence between δ_1 and Y_1, the right-hand side of (34) can be rewritten as

$$(2\pi)^{-2d} \int_{\mathbb{R}^d} E(|Y_1|^2|X_1 = u) E \left| \int_{\mathbb{R}^d} e^{it\cdot(\delta_1+u)} \sum_{k\in\mathbb{Z}^d} \varphi_{j,k}(x) \frac{\overline{(\varphi_{j,k})^{ft}(t)}}{f_\delta^{ft}(t)} \mathrm{d}t \right|^2 f_X(u)\mathrm{d}u.$$

Clearly, $E(|Y_1|^2|X_1 = u) = E[(Y_1 - E(Y_1|X_1 = u))^2|X_1 = u] + [E(Y_1|X_1 = u)]^2$; Because X_1 is independent of ε_1 and $E\varepsilon_1 = 0$, $m(u) = E(Y_1|X_1 = u)$ and $E(|Y_1|^2|X_1 = u) = E\varepsilon_1^2 + m^2(u)$. This with $\|f_X\|_\infty + \|m^2 f_X\|_\infty \le C_1$ and $E\varepsilon^2 < \infty$ shows

$$\|E(|Y_1|^2|X_1 = \cdot) f_X(\cdot)\|_\infty \lesssim 1. \quad (35)$$

Hence,

$$Var\ \Psi(W_1, Y_1) \lesssim (2\pi)^{-d} E \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} e^{it\cdot(\delta_1+u)} \sum_{k\in\mathbb{Z}^d} \varphi_{j,k}(x) \frac{\overline{(\varphi_{j,k})^{ft}(t)}}{f_\delta^{ft}(t)} \mathrm{d}t \right|^2 \mathrm{d}u.$$

By the compactness of φ^{ft} and the first inequality of (24), $\sum_k \varphi_{j,k}(x)\overline{(\varphi_{j,k})^{ft}(t)} / f_\delta^{ft}(t) \in L^2(\mathbb{R}^d)$ for fixed x and j. Then,

$$Var\ \Psi(W_1, Y_1) \lesssim \int_{\mathbb{R}^d} \left| \sum_{k\in\mathbb{R}^d} \varphi_{j,k}(x) \frac{\overline{(\varphi_{j,k})^{ft}(t)}}{f_\delta^{ft}(t)} \right|^2 \mathrm{d}t \lesssim \int_{\mathbb{R}^d} \left| \frac{\overline{\varphi^{ft}(2^{-j}t)}}{f_\delta^{ft}(t)} \right|^2 \mathrm{d}t$$

$$\lesssim \int_{\left[-\frac{4\pi 2^j}{3}, \frac{4\pi 2^j}{3}\right]^d} |\varphi^{ft}(2^{-j}t)|^2 \prod_{i=1}^{d} (1 + |t_i|)^{2\alpha_i} \mathrm{d}t \lesssim 2^{j(2|\alpha|+d)}, \quad (36)$$

where the Parseval identity is used in the first inequality; the second one holds because $[\varphi_{j,k}]^{ft}(t) = 2^{-dj/2} e^{i2^{-j}k\cdot t} \varphi^{ft}(2^{-j}t)$ and $|\sum_{k\in\mathbb{Z}^d} \varphi(x - k)| \lesssim 1$; for the last two inequalities, one uses (24) and supp $\varphi^{ft} \subset [-\frac{4\pi}{3}, \frac{4\pi}{3}]^d$.

To estimate the second term in (33), one writes

$$\sum_{l=1}^{n} |Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))| := T_1 + T_2 \tag{37}$$

with $T_1 = \sum_{l=1}^{2^{dj}-1} |Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))|$ and $T_2 = \sum_{l=2^{dj}}^{n} |Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))|$. Here, $1 \le 2^{jd} \le n$ holds for large n because of the given assumption $2^j \sim n^{\frac{1}{2s+2|\alpha|+d}}$. Clearly,

$$\begin{aligned} &|Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))| \\ &\le \int_{\mathbb{R}^d}\int_{-T}^{T}\int_{\mathbb{R}^d}\int_{-T}^{T} |h(w_0, y_0, w_l, y_l)||\Psi(w_0, y_0)\Psi(w_l, y_l)|\mathrm{d}y_0\mathrm{d}w_0 dy_l\mathrm{d}w_l, \end{aligned}$$

where $h(w_0, y_0, w_l, y_l) := f_{(W_0,Y_0,W_l,Y_l)}(w_0, y_0, w_l, y_l) - f_{(W,Y)}(w_0, y_0) f_{(W,Y)}(w_l, y_l)$. This with (25) leads to

$$|Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))| \lesssim \left[\int_{\mathbb{R}^d}\int_{-T}^{T} |\Psi(w, y)|\mathrm{d}y\mathrm{d}w\right]^2.$$

Note that $\Psi(w, y) = \sum_{k\in\mathbb{Z}^d} y(K^*\varphi)_{j,k}(w)\varphi_{j,k}(x)$ and $(K^*\varphi)_{j,k}$ is defined in Lemma 4. Then,

$$|\Psi(w, y)| \lesssim |y| \sum_{k\in\mathbb{Z}^d} 2^{j(|\alpha|+d/2)} \left[\prod_{l=1}^{d}\Big(1 + |2^j w_l - k_l|\Big)^{-2}\right] \cdot |\varphi_{j,k}(x)| \tag{38}$$

thanks to Lemma 4. Hence, $|Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))|$ can be bounded by

$$\left\{\sum_{k\in\mathbb{Z}^d} 2^{j(|\alpha|+d/2)} \int_{\mathbb{R}^d}\int_{-T}^{T} |y| \left[\prod_{l=1}^{d}(1 + |2^j w_l - k_l|)^{-2}\right] \mathrm{d}y\mathrm{d}w |\varphi_{j,k}(x)|\right\}^2 \lesssim 2^{2j|\alpha|},$$

which implies

$$T_1 \lesssim 2^{j(2|\alpha|+d)}. \tag{39}$$

For an upper bound of T_2, one takes $p = q = \frac{2}{1-\zeta}$ with $\zeta \in (0, 1)$ in Lemma 6 and obtains that

$$\begin{aligned} &|Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))| \lesssim [\alpha_{(W,Y)}(l)]^{\zeta} [E|\Psi(W_0, Y_0)|^{2/(1-\zeta)}]^{1-\zeta} \\ &\lesssim [\alpha_{(W,Y)}(l)]^{\zeta} \cdot \left[\sup_{(w,y)\in\mathbb{R}^d\times[-T,T]} |\Psi(w, y)|\right]^{2\zeta} \cdot \Big[\, E|\Psi(W_0, Y_0)|^2 \,\Big]^{1-\zeta}. \end{aligned}$$

The arguments from (34) to (36) show $E|\Psi(W_0, Y_0)|^2 \lesssim 2^{j(2|\alpha|+d)}$. On the other hand, (38) tells

$$\sup_{(w,y)\in\mathbb{R}^d\times[-T,T]} |\Psi(w, y)| \lesssim 2^{j(|\alpha|+d)}. \tag{40}$$

Hence, $|Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))| \lesssim [\alpha_{(W,Y)}(l)]^{\zeta} 2^{j(2|\alpha|+d+d\zeta)}$ and

$$T_2 \lesssim 2^{j(2|\alpha|+d)} 2^{dj\zeta} \sum_{l=2^{dj}}^{n} [\alpha_{(W,Y)}(l)]^{\zeta} \lesssim 2^{j(2|\alpha|+d)} \sum_{l=2^{dj}}^{n} l^{\zeta} [\alpha_{(W,Y)}(l)]^{\zeta} \lesssim 2^{j(2|\alpha|+d)}$$

thanks to the given condition $\alpha_{(W,Y)}(l) \le \mu_0 \exp(-\mu_1 l)$. Substituting this and (39) into (37), one receives

$$\sum_{l=1}^{n} |Cov(\Psi(W_0, Y_0), \Psi(W_l, Y_l))| \lesssim 2^{j(2|\alpha|+d)}.$$

This with (36) and (33) indicates $Var\ p_n(x) \lesssim n^{-1} 2^{j(2|\alpha|+d)}$. Similarly, $Var\ f_{X,n}(x) \lesssim n^{-1} 2^{j(2|\alpha|+d)}$. According to the choice $2^j \sim n^{\frac{1}{2s+2|\alpha|+d}}$,

$$Var\ f_{X,n}(x) + Var\ p_n(x) \lesssim n^{-1} 2^{j(2|\alpha|+d)} \lesssim n^{-\frac{2s}{2s+2|\alpha|+d}} := \varepsilon_n. \tag{41}$$

It remains to estimate the bias terms $|Ep_n(x) - p(x)| + |Ef_{X,n}(x) - f_X(x)|$. By (6), $Ef_{X,n}(x) = P_j f_X(x)$ and $Ep_n(x) = P_j p(x)$. On the other hand, $P_j f = K_j f$, where $K_j f(x) = \int_{\mathbb{R}^d} K_j(x, y) f(y) \mathrm{d}y$ with $K_j(x, y) := 2^{dj} K(2^j x, 2^j y)$ and $K(x, y) := \sum_{k\in\mathbb{Z}^d} \varphi(x-k)\varphi(y-k)$. Then, it is sufficient to estimate

$$|K_j f(x) - f(x)|^2$$

with f being f_X or p.

Since f satisfies the local Hölder condition of order s ($s \in \mathbb{R}^+ \setminus \mathbb{N}$) at x, there exist constants $C, r > 0$ such that $|\partial^\beta f(y) - \partial^\beta f(z)| \le C|y - z|^{s-|\beta|}$ holds for $y,\ z \in Q(x, r)$ and $|\beta| = \lfloor s \rfloor$. Denote

$$B_j := |K_j f(x) - f(x)|^2 = \left| \int_{\mathbb{R}^d} K_j(x, y)[f(y) - f(x)] \mathrm{d}y \right|^2 \lesssim B_{j1} + B_{j2}$$

with $B_{j1} := |\int_{Q(x,r)} K_j(x, y)[f(y) - f(x)]\mathrm{d}y|^2$ and $B_{j2} := |\int_{\mathbb{R}^d\setminus Q(x,r)} K_j(x, y) \big[f(y) - f(x)\big]\mathrm{d}y|^2$. By $\|f\|_\infty \le C_1$ and $K_j(x, y) = \prod_{\eta=1}^{d} \tilde{K}_j(x_\eta,\ y_\eta)$,

$$
\begin{aligned}
B_{j2} &\lesssim \left[\int_{\mathbb{R}^d\setminus Q(x,r)} |K_j(x,y)|\mathrm{d}y\right]^2\\
&\le \left[\sum_{l=1}^{d}\int_{\mathbb{R}^{d-1}}\int_{|y_l-x_l|>r}\prod_{\eta=1}^{d}|\tilde{K}_j(x_\eta,y_\eta)|\mathrm{d}y_l\mathrm{d}y_l^{-}\right]^2\\
&\le \left[\sum_{l=1}^{d}\int_{|y_l-x_l|>r}|\tilde{K}_j(x_l,y_l)|\left|\frac{y_l-x_l}{r}\right|^s\mathrm{d}y_l\right]^2
\end{aligned}
$$

with $\mathrm{d}y_l^{-} := \mathrm{d}y_1\cdots\mathrm{d}y_{l-1}\mathrm{d}y_{l+1}\cdots\mathrm{d}y_d$. This with (27) leads to

$$
B_{j2} \lesssim \left[\sum_{l=1}^{d}\int_{|y_l-x_l|>r} 2^j\tilde{\Phi}\left(2^{j-1}|x_l-y_l|\right)|y_l-x_l|^s\mathrm{d}y_l\right]^2 \lesssim 2^{-2js}. \tag{42}
$$

To find an upper bound of B_{j1}, one considers firstly the case $s\in(0,1)$, for which the local Hölder condition of f implies that

$$
B_{j1} = \left|\int_{Q(x,r)} K_j(x,y)[f(x)-f(y)]\mathrm{d}y\right|^2 \le \left[\int_{Q(x,r)}|K_j(x,y)||y-x|^s\mathrm{d}y\right]^2.
$$

According to (27),

$$
B_{j1} \lesssim \left[\int_{\mathbb{R}^d} 2^{jd}\Phi(2^{j-1}|y-x|)|y-x|^s\mathrm{d}y\right]^2 \lesssim 2^{-2js}. \tag{43}
$$

For the case $s>1$, one uses Taylor theorem with $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$, $i:=(i_1,\ldots,i_d)$ and $x^i := x_1^{i_1}\cdots x_d^{i_d}$ ($x\in\mathbb{R}^d$) to get

$$
\begin{aligned}
f(y)-f(x) = &\sum_{k=1}^{\lfloor s\rfloor}\frac{1}{k!}\sum_{|i|=k}\frac{k!}{i_1!\cdots i_d!}\partial^i f(x)(y-x)^i\\
&+\frac{1}{\lfloor s\rfloor!}\sum_{|i|=\lfloor s\rfloor}\frac{\lfloor s\rfloor!}{i_1!\cdots i_d!}\partial^i[f(x+\theta(y-x))-f(x)](y-x)^i
\end{aligned}
$$

with $\theta\in(0,1)$. Hence, $B_{j1}\lesssim B_{j11}+B_{j12}$, where

$$
B_{j11} = \left|\sum_{k=1}^{\lfloor s\rfloor}\sum_{|i|=k}\frac{1}{i_1!\cdots i_d!}\partial^i f(x)\int_{Q(x,r)}K_j(x,y)(y-x)^i\mathrm{d}y\right|^2 \quad \text{and}
$$

$$
B_{j12} = \left|\sum_{|i|=\lfloor s\rfloor}\frac{1}{i_1!\cdots i_d!}\int_{Q(x,r)}K_j(x,y)\partial^i[f(x+\theta(y-x))-f(x)](y-x)^i\mathrm{d}y\right|^2.
$$

By the local Hölder condition of f and $|y_l - x_l|^{i_l} \le |y - x|^{i_l}$ $(l = 1, 2, \ldots, d)$,

$$B_{j12} \lesssim \left[\sum_{|i|=\lfloor s \rfloor} \frac{1}{i_1! \ldots i_d!} \int_{Q(x,r)} |K_j(x, y)||y - x|^s \mathrm{d}y \right]^2.$$

The same arguments as (43) show

$$B_{j12} \lesssim 2^{-2js}. \tag{44}$$

Because $\int_{\mathbb{R}^d} K(x, y)(y - x)^\gamma \mathrm{d}y = 0$ $(|\gamma| = 1, 2, \ldots, \lfloor s \rfloor)$, $\int_{Q(x,r)} K(x, y)(y - x)^\gamma \mathrm{d}y = -\int_{\mathbb{R}^d \setminus Q(x,r)} K(x, y)(y - x)^\gamma \mathrm{d}y$. This with Lemma 5 concludes

$$B_{j11} \lesssim \Big[\sum_{k=1}^{\lfloor s \rfloor} \sum_{|i|=k} \int_{\mathbb{R}^d \setminus Q(x,r)} |K_j(x, y)||y_1 - x_1|^{i_1} \cdots |y_d - x_d|^{i_d} \mathrm{d}y \Big]^2. \tag{45}$$

According to the definitions of K_j and $Q(x, r)$,

$$\begin{aligned}
&\int_{\mathbb{R}^d \setminus Q(x,r)} |K_j(x, y)||y_1 - x_1|^{i_1} \cdots |y_d - x_d|^{i_d} \mathrm{d}y \\
&\le \sum_{l=1}^{d} \int_{\mathbb{R}^{d-1}} \int_{|y_l - x_l| > r} \left| \prod_{\eta=1}^{d} \tilde{K}_j(x_\eta, y_\eta) \right| |y_1 - x_1|^{i_1} \cdots |y_d - x_d|^{i_d} \mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&\lesssim \sum_{l=1}^{d} \left[\prod_{\eta=1, \eta \ne l}^{d} \int_{\mathbb{R}} |\tilde{K}_j(x_\eta, y_\eta)||y_\eta - x_\eta|^{i_\eta} \mathrm{d}y_\eta \right] \left[\int\limits_{|y_l - x_l| > r} |\tilde{K}_j(x_l, y_l)||y_l - x_l|^{i_l} \mathrm{d}y_l \right].
\end{aligned}$$

Note that $|\tilde{K}_j(x, y)| \lesssim 2^j \tilde{\Phi}(2^{j-1}|x - y|)$ and $\int_{\mathbb{R}} \tilde{\Phi}(|u|)|u|^{r'} \mathrm{d}u \lesssim 1$ with $1 \le r' \le s$. Then,

$$\int_{\mathbb{R}} |\tilde{K}_j(x_\eta, y_\eta)||x_\eta - y_\eta|^{i_\eta} \mathrm{d}y_\eta \lesssim \int_{\mathbb{R}} 2^j \tilde{\Phi}(2^{j-1}|x_\eta - y_\eta|)|x_\eta - y_\eta|^{i_\eta} \mathrm{d}y_\eta \lesssim 1.$$

On the other hand, $|y_l - x_l|^{i_l} = |y_l - x_l|^s |y_l - x_l|^{i_l - s} \le |y_l - x_l|^s r^{i_l - s}$ for $|y_l - x_l| > r$ and $i_l < s$. Hence,

$$\int_{|y_l - x_l| > r} |\tilde{K}_j(x_l, y_l)||y_l - x_l|^{i_l} \mathrm{d}y_l \lesssim \int_{|y_l - x_l| > r} |\tilde{K}_j(x_l, y_l)||y_l - x_l|^s \mathrm{d}y_l \lesssim 2^{-js}.$$

Thus, (45) reduces to $B_{j11} \lesssim 2^{-2js}$. This with (43) and (44) proves $B_{j1} \lesssim 2^{-2js}$. Combining it with (42), one knows $B_j = |K_j f(x) - f(x)|^2 \lesssim B_{j1} + B_{j2} \lesssim 2^{-2js}$ for $f = f_X$ or p. Moreover,

$$|Ep_n(x) - p(x)|^2 + |Ef_{X,n}(x) - f_X(x)|^2 \lesssim 2^{-2js} \lesssim n^{-2s/(2s+2|\alpha|+d)} = \varepsilon_n \tag{46}$$

thanks to the choice $2^j \simeq n^{1/(2s+2|\alpha|+d)}$. Substituting (46) and (41) into (32), one receives

$$\sup_{(m,f_X)\in\tilde{\mathcal{P}}_{x,s}} P[|\hat{m}_n(x) - m(x)|^2 \geq c\varepsilon_n] \leq c^{-1}.$$

Finally, the desired conclusion of Theorem 2 follows. □

Remark 2 When $d = 1$, the convergence rate in Theorem 2 coincides with Theorem 3.3 of Meister (2009), where the author studies a kernel estimator with i.i.d. data. It should be pointed out that our estimation for the bias term is similar to Meister's, although the technical conditions $f_X^{ft} \in L(\mathbb{R}^d)$ and $p^{ft} \in L(\mathbb{R}^d)$ are removed.

Remark 3 Because our Theorem 2 allows the data $\{(W_j, Y_j)\}$ strong mixing, the estimation for the variance terms is more difficulty than i.i.d. case, see Meister (2009). In particular, the covariance term in (33) need to be dealt with carefully, which vanishes for i.i.d. case. In addition, we pay a price to assume an extra conditions on f_δ^{ft}, i.e., the second inequality of (24). Compared with Theorem 1, Theorem 2 requires f_δ^{ft} having no zeros. It is interesting to consider the convergence rate for Fourier-oscillating noises.

Remark 4 The assumption $Y_j \in [-T, T]$ plays a key role for the proofs of (15), (39) and (40). Our method can apply to a more general model

$$m^*(x) = E\{\rho(Y)|X = x\}$$

for $Y \in \mathbb{R}$ and $\rho \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, see Chaubey and Shirazi (2015). When $\rho(y) = yI_{[-T,\ T]}(y)$, the model reduces to ours discussed in this current paper. In fact, similar results to Theorems 1 and 2 can be obtained for $m^*(x)$.

In contrast to the estimation in Theorem 1(b), the estimator in Theorem 1(a) and Theorem 2 is not practical, because the summation index k runs over $\mathbb{Z}^d$. The next section studies the consistency and convergence rate for a practical wavelet estimator under some mild conditions.

4 Practical estimation

We truncate the estimator defined by (3)–(5) to obtain a practical estimator $\hat{m}_n^F(x) = p_n^F(x)/f_{X,n}^F(x)$, where

$$f_{X,n}^F(x) = \sum_{k\in\mathcal{K}_n} \hat{\alpha}_{j,k}\varphi_{j,k}(x), \quad p_n^F(x) = \sum_{k\in\mathcal{K}_n} \hat{\gamma}_{j,k}\varphi_{j,k}(x) \ \text{ and } \\ \mathcal{K}_n := \{(k_1, \ldots, k_n) :\ |k_i| \leq K_n,\ i = 1,\ 2,\ \ldots,\ d\} \tag{47}$$

with K_n specified later on. Then, we have the following conclusion.

Theorem 3 *Consider problem* (1) *with* $\alpha_{(W,Y)}(k) \leq \mu_0 \exp(-\mu_1 k)(\mu_0, \mu_1 > 0)$ *and* $|x|^2 f_X(x) \in L^2(\mathbb{R}^d)$.

(a) If f_δ^{ft} *has no zeros on* $\mathbb{R}^d$*, then*

$$\lim_{n\to\infty} \hat{m}_n^F(x) \overset{a.s.}{=} m(x)$$

holds for each Lebesgue point x *of* $p := mf_X$ *and* $f_X(f_X(x) \neq 0)$;

(b) When (24), (25) *hold and* $2^j \sim n^{\frac{1}{2s+2|\alpha|+d}}$,

$$\lim_{c\to\infty} \overline{\lim_{n\to\infty}} \sup_{(m,f_X)\in\mathcal{P}_{x,s}} P\left[|\hat{m}_n^F(x) - m(x)|^2 \geq cn^{-\frac{2s}{2s+2|\alpha|+d}}\right] = 0$$

with $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$.

Proof (a) As in the proof of Theorem 1(a), it is sufficient to prove $\lim_{n\to\infty} p_n^F(x) \overset{a.s.}{=} p(x)$. For $\varepsilon > 0$,

$$P\left(|p_n^F(x) - p(x)| > 6\varepsilon\right) \leq P(|p_n^F(x) - Ep_n^F(x)| > 2\varepsilon) + P\left(|Ep_n^F(x) - P_j p(x)| > 2\varepsilon\right) + P(|P_j p(x) - p(x)| > 2\varepsilon). \quad (48)$$

Similar arguments to Theorem 1(a), the first and third terms of (48) satisfy that

$$P(|p_n^F(x) - Ep_n^F(x)| > 2\varepsilon) \lesssim n^{-7/4}, \quad (49)$$

$$P(|P_j p(x) - p(x)| > 2\varepsilon) = 0 \quad \text{(for large } n) \quad (50)$$

with j defined in (16).

The main work is to estimate the middle one.

By $Ep_n^F(x) = \sum_{k\in\mathcal{K}_n} \gamma_{j,k}\varphi_{j,k}(x)$ and $P_j p(x) = \sum_{k\in\mathbb{Z}^d} \gamma_{j,k}\varphi_{j,k}(x)$, one knows

$$P\left(\left|Ep_n^F(x) - P_j p(x)\right| > 2\varepsilon\right) = P\left(\left|\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} \gamma_{j,k}\varphi_{j,k}(x)\right| > 2\varepsilon\right). \quad (51)$$

Define $\mathcal{K}_{n,l} := \{(k_1, \ldots, k_d) \in \mathbb{Z}^d | \; |k_l| \leq K_n\}$ for $l \in \{1, 2, \ldots, d\}$. Then, $\mathbb{Z}^d \setminus \mathcal{K}_n \subset \bigcup_{l=1}^d (\mathbb{Z}^d \setminus \mathcal{K}_{n,l})$ thanks to (47). Furthermore,

$$\left|\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} \gamma_{j,k}\varphi_{j,k}(x)\right| \leq \sum_{l=1}^d \sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_{n,l}} |\gamma_{j,k}||\varphi_{j,k}(x)|. \quad (52)$$

Next, one estimates $|\gamma_{j,k}|$. Since $\gamma_{j,k} := \int_{\mathbb{R}^d} \varphi_{j,k}(x)p(x)\mathrm{d}x$, $|\gamma_{j,k}| \leq \int_{\mathbb{R}^d} |\varphi_{j,k}(x)||p(x)|\mathrm{d}x$ and

$$k_l^2|\gamma_{j,k}| \leq \int_{\mathbb{R}} k_l^2|\tilde{\varphi}_{j,k_l}(x_l)| \int_{\mathbb{R}^{d-1}} \left|\prod_{i\neq l} \tilde{\varphi}_{j,k_i}(x_i)\right| |p(x)|\mathrm{d}x_l^- \mathrm{d}x_l$$

with $\mathrm{d}x_l^- = \mathrm{d}x_1 \cdots \mathrm{d}x_{l-1}\mathrm{d}x_{l+1} \cdots \mathrm{d}x_d$. Then, it follows from $k_l^2 = |2^j x_l - k_l + 2^j x_l|^2 \lesssim |2^j x_l - k_l|^2 + |2^j x_l|^2$ that

$$\begin{aligned} &k_l^2|\gamma_{j,k}| \leq I + II, \quad \text{where} \\ &\quad I = \int_{\mathbb{R}} |2^j x_l - k_l|^2 |\tilde{\varphi}_{j,k_l}(x_l)| \int_{\mathbb{R}^{d-1}} \left|\prod_{i\neq l} \tilde{\varphi}_{j,k_i}(x_i)\right| |p(x)|\mathrm{d}x_l^- \mathrm{d}x_l, \\ &\quad II = \int_{\mathbb{R}} |2^j x_l|^2 |\tilde{\varphi}_{j,k_l}(x_l)| \int_{\mathbb{R}^{d-1}} \left|\prod_{i\neq l} \tilde{\varphi}_{j,k_i}(x_i)\right| |p(x)|\mathrm{d}x_l^- \mathrm{d}x_l. \end{aligned}$$

Note that $Y_i \in [-T, T]$ implies that $|m(x)| \leq T$ and $p := mf_X \in L(\mathbb{R}^d)$. Then,

$$I \lesssim 2^{dj/2} \left[\sup_{x_l\in\mathbb{R}} |x_l^2 \tilde{\varphi}(x_l)|\right] \cdot \left[\prod_{i\neq l} \sup_{x_i\in\mathbb{R}} |\tilde{\varphi}(x_i)|\right] \cdot \int_{\mathbb{R}^d} |p(x)|\mathrm{d}x \lesssim 2^{dj/2},$$

where two supremums exist because $\tilde{\varphi}$ is the Meyer's scaling function. On the other hand, the given conditions $|x|^2 p(x) \in L^2(\mathbb{R}^d)$ and Hölder inequality tell that $II = 2^{2j} \int_{\mathbb{R}^d} |\varphi_{j,k}(x)||x_l^2||p(x)|\mathrm{d}x \leq 2^{2j}\|\varphi_{j,k}\|_2 \|x_l^2 p(x)\|_2 \lesssim 2^{2j}$. This with the estimate of I shows

$$k_l^2|\gamma_{j,k}| \lesssim 2^{dj/2} + 2^{2j}. \tag{53}$$

Take $K_n = \lceil (2^{dj/2} + 2^{(1+d/4)j})2^{j/2} \rceil$. Then, (52) reduces to

$$\begin{aligned} \sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} |\gamma_{j,k}||\varphi_{j,k}(x)| &\lesssim \sum_{l=1}^{d} \sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_{n,l}} \frac{1}{k_l^2}(2^{dj/2} + 2^{2j})|\varphi_{j,k}(x)| \\ &\lesssim d\frac{1}{K_n^2}(2^{dj/2} + 2^{2j})2^{dj/2} \lesssim 2^{-j}, \end{aligned} \tag{54}$$

where $\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_{n,l}} |\varphi(2^j x - k)| \lesssim 1$ is used in the second inequality. As $n \to +\infty$, $j \to +\infty$ and $\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} |\gamma_{j,k}||\varphi_{j,k}(x)| \to 0$ thanks to (54). This with (51) shows

$$P(|Ep_n^F(x) - P_j p(x)| > 2\varepsilon) \leq P\left(\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} |\gamma_{j,k}||\varphi_{j,k}(x)| > 2\varepsilon\right) = 0$$

for large n. Combining this with (48)–(50), one obtains that for any $\varepsilon > 0$, $\sum_{n=1}^{+\infty} P(|p_n^F(x) - p(x)| > 6\varepsilon) < +\infty$. Finally, the desired conclusion

$$\lim_{n\to\infty} p_n^F(x) \overset{a.s.}{=} p_n(x)$$

follows from Borel–Cantelli lemma.

(b). As in the proof of Theorem 2, it suffices to estimate $E|p_n^F(x) - p(x)|^2$. The estimation of $E|f_{X,n}^F(x) - f_X(x)|^2$ is similar. Obviously, $E|p_n^F(x) - p(x)|^2$ is bounded by

$$E|p_n^F(x) - Ep_n^F(x)|^2 + |Ep_n^F(x) - P_j p(x)|^2 + |P_j p(x) - p(x)|^2. \tag{55}$$

According to the proof of Theorem 2,

$$|P_j p(x) - p(x)| \lesssim n^{-s/(2s+2|\alpha|+d)}. \tag{56}$$

Similar arguments to that theorem show

$$E|p_n^F(x) - Ep_n^F(x)|^2 \lesssim n^{-2s/(2s+2|\alpha|+d)}. \tag{57}$$

Clearly, the middle term of (55) satisfies

$$|Ep_n^F(x) - P_j p(x)|^2 \lesssim \left[\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} |\gamma_{j,k}||\varphi_{j,k}(x)|\right]^2.$$

On the other hand, the arguments of (54) with a little different choice $K_n = \lceil(2^{dj/2} + 2^{(1+d/4)j})2^{js/2}\rceil$ ($2^j \sim n^{\frac{1}{2s+2|\alpha|+d}}$) concludes

$$\sum_{k\in\mathbb{Z}^d\setminus\mathcal{K}_n} |\gamma_{j,k}||\varphi_{j,k}(x)| \lesssim 2^{-js} \lesssim n^{-s/(2s+2|\alpha|+d)}, \tag{58}$$

where $2^j \sim n^{\frac{1}{2s+2|\alpha|+d}}$ is used in the last inequality. Hence,

$$|Ep_n^F(x) - P_j p(x)|^2 \lesssim n^{-2s/(2s+2|\alpha|+d)}.$$

This with (56)–(57) concludes $E|p_n^F(x) - p(x)|^2 \lesssim n^{-2s/(2s+2|\alpha|+d)}$. The remaining proofs are the same as those in Theorem 2. This completes the proof of Theorem 3(b). □

It is a good problem to study the numerical illustration of our practical estimation. We shall investigate it in future.

Acknowledgements This paper is supported by the Beijing Natural Science Foundation (No. 1172001) and the National Natural Science Foundation of China (No. 11771030). The authors would like to thank the referees for their important comments and suggestions.

References

Bosq, D., Blanke, D. (2007). Inference and prediction in large dimensions. In *Wiley series in probability and statistics*. Chichester: Wiley-Dunod.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement error in nonlinear models* (2nd ed.). Boca Raton: Chapman and Hall CRC Press.

Carroll, R. J., Delaigle, A., Hall, P. (2007). Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *Journal of the Royal Statistical Society Series B*, *69*, 859–878.

Chaubey, Y. P., Shirazi, E. (2015). On MISE of a nonlinear wavelet estimator of the regression function based on biased data under strong mixing. *Communications in Statistics Theory and Methods*, *44*, 885–899.

Chaubey, Y. P., Chesneau, C., Shirazi, E. (2013). Wavelet-based estimation of regression function for dependent biased data under a given random design. *Journal of Nonparametric Statistics*, *25*, 53–71.

Chesneau, C. (2010). On adaptive wavelet estimation of the regression function and its derivatives in an errors-in-variables model. *Current Development in Theory and Applications of Wavelets*, *4*, 185–208.

Chesneau, C. (2014). A general result on the performance of the wavelet hard thresholding estimator under α-mixing dependence. *Journal of Probaility and Statistics*, Article ID: 403764.

Chesneau, C., Fadili, J., Maillot, B. (2015). Adaptive estimation of an additive regression function from weakly dependent data. *Journal of Multivariate Analysis*, *133*, 77–94.

Chichignoud, M., Hoang, V. H., Ngoc, T. M. P., Rivoirard, V. (2017). Adaptive wavelet multivariate regression with errors in variables. *Electronic Journal of Statistics*, *11*, 682–724.

Daubechies, I. (1992). *Ten lectures on wavelets*. Philadelphia: SIAM.

Davydov, Y. (1970). The invariance principle for stationary processes. *Theory of Probability and its Applications*, *15*(3), 498–509.

Doukhan, P. (1994). *Mixing: Properties and examples*. Lecture Notes in Statistics 85. New York: Springer.

Fan, J., Koo, J. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory*, *48*(3), 734–747.

Fan, J., Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, *21*, 1900–1925.

Guo, H. J., Liu, Y. M. (2017). Strong consistency of wavelet estimators for errors-in-variables regression model. *Annals of the Institute of Statistical Mathematics*, *69*, 121–144.

Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications.* Lecture Notes in Statist, 129. Heidelberg: Springer.

Kelly, S. E., Kon, M. A., Raphael, L. A. (1994). Local convergence for wavelet expansions. *Journal of Functional Analysis*, *126*, 102–138.

Masry, E. (1993). Multivariate regression estimation with errors-in-variables for stationary processes. *Journal of Nonparametric Statistics*, *3*(1), 13–36.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics*. Berlin: Springer.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications*, *29*, 309–315.

Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica*, *72*, 33–75.

Shen, J., Xie, Y. (2013). Strong consistency of the internal estimator of nonparametric regression with dependent data. *Statistics and Probability Letters*, *83*, 1915–1925.