Ann Inst Stat Math (2017) 69:121–144
DOI 10.1007/s10463-015-0529-6

Strong consistency of wavelet estimators for errors-in-variables regression model

Huijun Guo[1] · **Youming Liu**[1]

Received: 9 November 2014 / Revised: 23 May 2015 / Published online: 4 August 2015

Abstract This paper studies the strong consistency of some estimators for an errors-in-variables regression model. We first provide an extension of Meister's theorem. Then, the same problem is dealt with under the Fourier-oscillating noises. Finally, we prove two strong consistency theorems for wavelet estimators corresponding to non-oscillating and Fourier-oscillating noises.

Keywords Errors-in-variables · Fourier-oscillating noise · Regression function · Strong consistency · Wavelets

1 Introduction

In this paper, we study the following errors-in-variables regression problem. Let (W_j, Y_j) $(j = 1, 2, \ldots, n)$ be independent and identically distributed (i.i.d.) data valued in $\mathbb{R}^d \times \mathbb{R}$ from the model

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

The errors ε_j and δ_j are independent of each other and independent of X_j. The functions f_X and f_δ denote the densities of X_j and δ_j, respectively. The regression errors ε_j satisfies $E\varepsilon_j = 0$. The problem is to approximate the regression function

✉ 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

m by some estimator $\hat{m}_n$ (depending on (W_j, Y_j), $j = 1, 2, \ldots, n$) in the sense of strong convergence.

This above model has some practical 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.

Fan and Truong (1993) apply a deconvolution technique to extend Nadaraya–Watson estimator to the general case. More precisely, they provide optimal weak convergence rates of their estimator to m for two types of noises δ_j, when m has some regularity, $d = 1$ and $f_\delta^{ft}(t)$ has no zeros. Here and after, $t \cdot x := \sum_{i=1}^{d} t_i x_i$ 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.

However, many noise densities have zeros in the Fourier transform domain. For example, in the experiment of Sun et al., the measurement errors are assumed to be uniformly distributed (Sun et al. 2002). Nonparametric function estimators under Fourier-oscillating noises are investigated (Hall and Meister 2007; Meister 2008; Delaigle and Meister 2011; Guo and Liu 2014). To get some convergence rates, they assume the estimated functions to be located in some Sobolev or Besov spaces.

Because we do not know if the estimated functions are smooth or not in some practical applications, it is more reasonable to consider the consistency for an estimator (Shen and Xie 2013). For the model (1) with $f_\delta^{ft}(t) \neq 0$ and $d = 1$, Meister (2009) shows a strong consistency theorem without assuming any regularity of the regression and density functions m, f_X. However, he requires the boundedness of f_X and the continuity of f_X at x.

In this small work, we first remove those conditions and rewrite the theorem for $d \geq 1$. Then, we extend our result to the model with Fourier-oscillating noises by developing Delaigle–Meister's technique (Delaigle and Meister 2011) from one-dimensional to multidimensional cases. Wavelet estimators are widely used in regression estimation (Li et al. 2008; Chesneau 2010; Gencay and Gradojevic 2011; Chaubey et al. 2013). We finally study the strong consistency of wavelet estimators.

The current paper is organized as follows: The first section gives an extension of Meister's theorem. We discuss the same problem with Fourier-oscillating noises using a kernel method in the second part. The strong consistency of wavelet estimators is proved in Sect. 4. More precisely, we use the Meyer's scaling function to define our estimator when $f_\delta^{ft}(t) \neq 0$, and use the Daubechies' to do for Fourier-oscillating noises. The last part provides some concluding remarks.

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 $B \lesssim A$.

2 An extension of Meister's Theorem

This section is devoted to generalizing the following theorem by Meister (2009). For the model (1), define $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ with

$$f_{X,n}(x) = \frac{1}{n}\sum_{l=1}^{n}\frac{1}{2\pi}\int_{\mathbb{R}}\exp(-itx)K^{ft}(ht)\exp(itW_l)\Big/ f_\delta^{ft}(t)\mathrm{d}t \tag{3}$$

and

$$p_n(x) = \frac{1}{n}\sum_{l=1}^{n}Y_l\frac{1}{2\pi}\int_{\mathbb{R}}\exp(-itx)K^{ft}(ht)\exp(itW_l)\Big/ f_\delta^{ft}(t)\mathrm{d}t. \tag{4}$$

Then, we state Meister's theorem (Theorem 3.2, Meister 2009, page 114) as follows:

Theorem 1 *Consider the errors-in-variables regression problem defined in* (1) *with* $d = 1$ *under the conditions that* $f_\delta^{ft}(t) \neq 0$ *for all* $t \in \mathbb{R}$*; the functions* $p = m \cdot f_X$ *and* f_X *are bounded on the whole real line and continuous at some* $x \in \mathbb{R}$*;* $p \in L^1(\mathbb{R})$*; the continuous and bounded kernel function* $K \in L^1(\mathbb{R})$ *satisfies* $\int_{\mathbb{R}} K(z)dz = 1$ *and* K^{ft} *is supported on* $[-1, 1]$*. Furthermore, select the bandwidth* $h = h_n \to 0$ *such that*

$$h \cdot \min_{|t|\leq \frac{1}{h}} \left|f_\delta^{ft}(t)\right| \geq n^{-\xi}$$

for some $\xi \in (0, \frac{1}{2})$ *and all* n *sufficiently large; assume that the (2s)th moment of* Y_1 *exists for some* $s > 1/(1-2\xi)$*, and that*

$$|m(x)| \leq C_1 \quad \textit{and} \quad f_X(x) \geq C_2$$

holds for some constants $C_1, C_2 > 0$*. Then, the estimator* $\hat{m}_n$ *satisfies*

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

We aim to remove the boundedness assumption of both f_X and p, to relax the continuity of those functions to a Lebesgue point. Moreover, the above theorem will be extended to multidimensional cases.

It is well known that $f \in L^1(\mathbb{R}^d)$ denotes an equivalence class of functions, which means two functions different from a measure zero set belong to the same class. In particular, $f \in L^1(\mathbb{R}^d)$ is allowed to take value ∞ on a measure zero set. Hence, it does not make sense to talk about a function value at a given point for $f \in L^1(\mathbb{R}^d)$ in general. However, to define a Lebesgue point x of $f \in L^1(\mathbb{R}^d)$, we can choose a representation element of f (still denoted by f) such that $|f(x)| < +\infty$ for each $x \in \mathbb{R}^d$, because the Lebesgue measure of the set $\{x, |f(x)| = +\infty\}$ must be zero for $f \in L^1(\mathbb{R}^d)$.

Let $f \in L^1(\mathbb{R}^d)$ and $B(x,r)$ denote the Euclidean ball centered at $x \in \mathbb{R}^d$ with radius $r > 0$. If $\lim_{r\to 0} r^{-d} \int_{B(x,r)} |f(y) - f(x)|dy = 0$, then x is called a Lebesgue point of f. Clearly, a continuous point of f must be a Lebesgue point. Although $f \in L^1(\mathbb{R}^d)$ may not have any continuous points, almost every $x \in \mathbb{R}^d$ is a Lebesgue point (Stein and Shakarchi 2005). The following lemma (Stein and Shakarchi 2005) is needed in the proof of Theorem 2.

Lemma 1 *If* $f \in L^1(\mathbb{R}^d)$ *and* $\{K_h\}_{h>0}$ *satisfy (i)* $\int_{\mathbb{R}^d} K_h(x)\mathrm{d}x = 1$; *(ii)* $|K_h(x)| \lesssim h^{-d}$ *for all* $h > 0$; *(iii)* $|K_h(x)| \lesssim h/|x|^{d+1}$ *for all* $h > 0$ *and* $x \in \mathbb{R}^d$, *then*

$$\lim_{h\to 0} f * K_h(x) = f(x)$$

holds for each Lebesgue point x *of* f.

Example 1 Let $K(x) = C \prod_{s=1}^{d} [2x_s^{-1} \sin(2^{-1}x_s)]^{d+1}$ with C being the normalized constant such that $\int_{\mathbb{R}^d} K(x)\mathrm{d}x = 1$. Define $K_h(x) = h^{-d} K(h^{-1}x)$. Then, K_h satisfies all conditions of Lemma 1.

Clearly, the function $K \in L^1(\mathbb{R}^d)$ and $|K_h(x)| \lesssim h^{-d}$ by the definitions of K and K_h. For condition (iii), it suffices to check the boundedness of the function

$$F(x_1, x_2, \ldots, x_d) := \left(\prod_{s=1}^{d} \frac{\sin x_s}{x_s}\right) \left(x_1^2 + x_2^2 + \cdots + x_d^2\right)^{1/2}$$

on $\mathbb{R}^d$.

When $|x_s| > 1$ for $s = 1, 2, \ldots, d$, it is easy to see that

$$\begin{aligned} F^2(x_1, \ldots, x_d) &\lesssim \left(\prod_{s=1}^{d} \frac{1}{x_s^2}\right) \left(x_1^2 + x_2^2 + \cdots + x_d^2\right) \\ &= \frac{1}{x_2^2 \cdots x_d^2} + \cdots + \frac{1}{x_1^2 \cdots x_{d-1}^2} \le d. \end{aligned}$$

Hence, F is bounded. In the other cases, we can assume that $|x_s| \le 1$ for $1 \le s \le l$, and $|x_s| > 1$ otherwise, where $1 \le l \le d$. By $(|a| + |b|)^\theta \lesssim |a|^\theta + |b|^\theta$ $(\theta > 0)$,

$$\begin{aligned} |F(x_1, x_2, \ldots, x_d)| &\lesssim \left(\prod_{s=1}^{d} \left|\frac{\sin x_s}{x_s}\right|\right) \left[\left(x_1^2 + \cdots + x_l^2\right)^{1/2} + \left(x_{l+1}^2 + \cdots + x_d^2\right)^{1/2}\right] \\ &\lesssim 1 + \left(\prod_{s=l+1}^{d} \left|\frac{\sin x_s}{x_s}\right|\right) \left(x_{l+1}^2 + \cdots + x_d^2\right)^{1/2}. \end{aligned}$$

The same arguments as in the first case conclude the boundedness of F.

We shall use the kernel function $K(x) = C\prod_{s=1}^{d}[2x_s^{-1}\sin(2^{-1}x_s)]^{d+1}$ in Example 1 to define our estimator $\hat{m}_n(x)$ for $m(x)$. More precisely,

$$\hat{m}_n(x) := p_n(x)/f_{X,n}(x) \tag{5}$$

with

$$f_{X,n}(x) := \frac{1}{n}\sum_{l=1}^{n}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)K^{ft}(ht)\exp(it\cdot W_l)\big/ f_\delta^{ft}(t)\mathrm{d}t \tag{6}$$

and

$$p_n(x) := \frac{1}{n}\sum_{l=1}^{n}Y_l\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)K^{ft}(ht)\exp(it\cdot W_l)\big/ f_\delta^{ft}(t)\mathrm{d}t. \tag{7}$$

This above estimator $\hat{m}_n(x)$ is the multidimensional version of Meister's except for our choosing a special kernel function K for convenient discussions. Note that $K_0(x) := [2x^{-1}\sin(2^{-1}x)]^{d+1}$ satisfies

$$K_0^{ft}(t) = 2\pi\,\underbrace{I_{[-1/2,\ 1/2]} * \cdots * I_{[-1/2,\ 1/2]}}_{d+1}(t),$$

where $I_{[-1/2,\ 1/2]}(t)$ stands for the indicator function of the set $[-1/2,\ 1/2]$. Then, supp $K_0^{ft} \subset [-\frac{d+1}{2},\ \frac{d+1}{2}]$ and the function $K^{ft}(t) = C\prod_{s=1}^{d}K_0^{ft}(t_s)$ so that supp $K^{ft} \subset [-\frac{d+1}{2},\ \frac{d+1}{2}]^d$. This with $f_\delta^{ft}(t) \neq 0$ $(t \in \mathbb{R}^d)$ shows that $f_{X,n}(x) \neq 0$ for almost every $x \in \mathbb{R}^d$. Hence, $\hat{m}_n(x)$ is $a.e.$ well defined.

To prove the consistency of $\hat{m}_n(x)$, we need a lemma, which comes from the proof of Theorem 3.2 in Meister (2009). We rewrite it in a multidimensional version and give a short proof.

Lemma 2 *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 bandwidth sequence $h_n \to 0$ such that*

$$h_n^d \cdot \min_{t\in[-1/h_n,\ 1/h_n]^d}|f(t)| \geq n^{-\frac{1}{6}}$$

holds for sufficiently large n.

Proof For $n \in \mathbb{Z}^+$ (the positive integer set), define

$$\Omega_n := \left\{h > 0 : h^d \min_{t\in[-1/h,\ 1/h]^d}|f(t)| \geq n^{-1/6}\right\}.$$

Since $f \in C(\mathbb{R}^d)$ and $f(t) \neq 0$, $\Omega_n \neq \emptyset$ and $d_n = \inf\Omega_n \geq 0$. Moreover, there exists $h_n \in \Omega_n$ such that $d_n \leq h_n \leq d_n + \frac{1}{n}$ for each $n \in \mathbb{N}^+$. Note that $\Omega_n \subset \Omega_{n+1}$.

Then, $d_{n+1} \leq d_n$. This with $d_n \geq 0$ shows $\{d_n\}_{n\in\mathbb{N}^+}$ convergent. It suffices to show $\lim_{n\to\infty} d_n = 0$ to conclude Lemma 2.

If $\lim_{n\to\infty} d_n = d_0 > 0$, there would exist $d_* \in (0,\ d_0)$ such that for each $n \in \mathbb{N}^+$,

$$d_*^d \cdot \min_{t\in[-1/d_*,\ 1/d_*]^d} |f(t)| < n^{-\frac{1}{6}}. \tag{8}$$

Let n in (8) tends to $+\infty$, one receives $d_*^d \cdot \min_{t\in[-1/d_*,\ 1/d_*]^d} |f(t)| = 0$. By $f \in C([-1/d_*,\ 1/d_*]^d)$, $f(t_0) = 0$ for some $t_0 \in \mathbb{R}^d$. This contradicts with $f(t) \neq 0$ for $t \in \mathbb{R}^d$. □

Remark 1 From the above proof, we find that Lemma 2 still holds, when $n^{-\frac{1}{6}}$ is replaced by $0 < \varepsilon_n \downarrow 0$. However, the current version is enough for our later discussions.

Now, we are in the position to state the main result of this section:

Theorem 2 *Consider the problem defined by* (1) *with* $f_\delta^{ft}(t) \neq 0$ *for each* $t \in \mathbb{R}^d$. *If* $p := mf_X \in L^1(\mathbb{R}^d)$, $E|Y_1|^4 < +\infty$, *x is a Lebesgue point of f_X and p ($f_X(x) \neq 0$), then with $h = h_n$ for $f = f_\delta^{ft}$ in Lemma 2, $\hat{m}_n(x)$ defined by* (5)–(7) *satisfies*

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

Proof It is sufficient to prove $\lim_{n\to\infty} p_n(x) \overset{a.s.}{=} p(x)$ and $\lim_{n\to\infty} f_{X,n}(x) \overset{a.s.}{=} f_X(x)$, since $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$, $m(x) = p(x)/f_X(x)$ with $f_X(x) \neq 0$. One shows $\lim_{n\to\infty} p_n(x) \overset{a.s.}{=} p(x)$ only. The proof for $\lim_{n\to\infty} f_{X,n}(x) \overset{a.s.}{=} f_X(x)$ is similar and even simpler.

Clearly, $|p_n(x) - p(x)| \leq |p_n(x) - Ep_n(x)| + |Ep_n(x) - p(x)|$ and $P[|p_n(x) - p(x)| \geq \varepsilon] \leq P[|p_n(x) - Ep_n(x)| \geq \varepsilon/2] + P[|Ep_n(x) - p(x)| \geq \varepsilon/2]$ for each $\varepsilon > 0$. By Markov's inequality,

$$P\left[|p_n(x) - p(x)| \geq \varepsilon\right] \lesssim \varepsilon^{-4} E\left|p_n(x) - Ep_n(x)\right|^4 + I_{[\varepsilon/2,\ \infty)}(|p(x) - Ep_n(x)|). \tag{9}$$

To estimate the second term in (9), one observes first from (7) and $h = h_n$ that

$$Ep_n(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-it\cdot x) K^{ft}(h_n t) E\left[Y_1 \exp(it\cdot W_1)\right] / f_\delta^{ft}(t) \mathrm{d}t. \tag{10}$$

Since $W_1 = X_1 + \delta_1$ and X_1 is independent of δ_1,

$$E\left(Y_1 e^{it\cdot W_1}\right) = E\left(Y_1 e^{it\cdot X_1}\right) E e^{it\cdot\delta_1} = E\left(Y_1 e^{it\cdot X_1}\right) f_\delta^{ft}(t).$$

On the other hand, it follows from $E\varepsilon_1 = 0$, the independence between X_1 and ε_1 that

$$E\left(Y_1|X_1 = x\right) = E\left(m(X_1) + \varepsilon_1|X_1 = x\right) = m(x) + E\left(\varepsilon_1|X_1 = x\right) = m(x).$$

Hence, $E(Y_1 e^{it\cdot X_1}) = E(E(Y_1|X_1)e^{it\cdot X_1}) = \int_{\mathbb{R}^d} m(u)e^{it\cdot u} f_X(u)du = p^{ft}(t)$. Then, (10) reduces to

$$Ep_n(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-it\cdot x)K^{ft}(h_n t)p^{ft}(t)\mathrm{d}t.$$

According to p, $K \in L^1(\mathbb{R}^d)$, one finds that $p * K_{h_n} \in L^1(\mathbb{R}^d)$ and $(p * K_{h_n})^{ft}(t) = p^{ft}(t)\cdot K^{ft}(h_n t)$. Moreover,

$$Ep_n(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp(-it\cdot x)(p * K_{h_n})^{ft}(t)\mathrm{d}t.$$

Clearly, $|(p * K_{h_n})^{ft}(t)| \lesssim |K^{ft}(h_n t)| \in L^1(\mathbb{R}^d)$. Then, by the inverse Fourier transform theorem (Stein and Shakarchi 2005),

$$Ep_n(x) = (p * K_{h_n})(x)$$

holds at each Lebesgue point x of p. This with Lemma 1 and Example 1 shows that $\lim_{n\to\infty}(p * K_{h_n})(x) = p(x)$. Since $\lim_{n\to\infty} h_n = 0$, the second term of (9) vanishes as $n \to \infty$ and

$$P\left[|p_n(x) - p(x)| \geq \varepsilon\right] \lesssim \varepsilon^{-4} E|p_n(x) - Ep_n(x)|^4. \tag{11}$$

The remaining proofs are completely same as those in Meister (2009), except for $\mathbb{R}^d$ data replacing $\mathbb{R}$ ones. One includes a proof for completeness. By (7),

$$\begin{aligned} p_n(x) - Ep_n(x) = (2\pi)^{-d} n^{-1} \sum_{l=1}^{n} \int_{\mathbb{R}^d} \exp(-it\cdot x)K^{ft}(h_n t)\Big[Y_l \exp(it\cdot W_l) \\ -EY_l \exp(it\cdot W_l)\Big]\Big/ f_\delta^{ft}(t)\mathrm{d}t. \end{aligned}$$

With the notation $\Psi_l(t) := Y_l \exp(it\cdot W_l) - EY_l \exp(it\cdot W_l)$, this above identity reduces to

$$p_n(x) - Ep_n(x) = (2\pi)^{-d} n^{-1} \sum_{l=1}^{n} \int_{\mathbb{R}^d} \exp(-it\cdot x)K^{ft}(h_n t)\Psi_l(t)\big/ f_\delta^{ft}(t)\mathrm{d}t.$$

Furthermore,

$$E|p_n(x) - Ep_n(x)|^4 \lesssim n^{-4} \sum_{l_1=1}^{n} \cdots \sum_{l_4=1}^{n} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \left\{ \prod_{r=1}^{2} \exp(-it_{2r-1} \cdot x) \right.$$

$$\left. \exp(it_{2r} \cdot x) K^{ft}(h_n t_{2r-1}) K^{ft}(-h_n t_{2r}) \left[f_\delta^{ft}(t_{2r-1}) f_\delta^{ft}(-t_{2r}) \right]^{-1} \right\}$$

$$E \prod_{r=1}^{2} \Psi_{l_{2r-1}}(t_{2r-1}) \Psi_{l_{2r}}(-t_{2r}) \mathrm{d}t_1 \ldots \mathrm{d}t_4. \tag{12}$$

Whenever the set $\{l_1, \ldots, l_4\}$ contains more than 2 different elements, at least one of the Ψ_{l_r} is stochastically independent of all other $\Psi_{l_{r'}}$ with $r' \neq r$, so that

$$E \prod_{r=1}^{2} \Psi_{l_{2r-1}}(t_{2r-1}) \Psi_{l_{2r}}(-t_{2r}) = 0.$$

For $\sharp\{l_1, \ldots, l_4\} \leq 2$, Jensen's inequality tells

$$\left| E \prod_{r=1}^{2} \Psi_{l_{2r-1}}(t_{2r-1}) \Psi_{l_{2r}}(-t_{2r}) \right| \leq 16 E|Y_1|^4.$$

Note that $\#L_n := \#\{(l_1, l_2, l_3, l_4),\ \sharp\{l_1, \ldots, l_4\} \leq 2\} \lesssim n^2$, supp $K^{ft} \subset [-\frac{d+1}{2}, \frac{d+1}{2}]^d$ and $|K^{ft}(t)| \lesssim 1$. Then, (12) becomes

$$E\,|p_n(x) - Ep_n(x)|^4 \lesssim n^{-2} \left[h_n^d \min_{t \in \left[-\frac{d+1}{2h_n},\, \frac{d+1}{2h_n}\right]^d} \left| f_\delta^{ft}(t) \right| \right]^{-4} E|Y_1|^4. \tag{13}$$

Applying Lemma 2, one obtains that

$$E\,|p_n(x) - Ep_n(x)|^4 \lesssim n^{-2}\, n^{2/3} = n^{-4/3}. \tag{14}$$

Finally, it follows from (11) and (14) that $P[|p_n(x) - p(x)| \gtrsim \varepsilon] \lesssim n^{-4/3}$ and $\sum_{n=1}^{\infty} P[|p_n(x) - p(x)| \gtrsim \varepsilon] < \infty$ for any $\varepsilon > 0$. This concludes $\lim_{n\to\infty} p_n(x) \overset{a.s.}{=} p(x)$ thanks to Borel–Cantelli lemma in probability theory. □

3 Consistency with Fourier-oscillating noises

In this section, we consider the same problem as in Sect. 2, but allow for f_δ^{ft} having some zeros. More precisely, we assume that for positive numbers λ_s (as in Delaigle and Meister 2011), $v_s \in \mathbb{Z}^+ \cup \{0\}$ $(1 \leq s \leq d)$, and

$$C\left(\mathbb{R}^d\right) \ni f_\delta^{ft}(t_1,\ldots,t_d)\prod_{s=1}^{d}\left(\sin\left(\frac{\pi t_s}{\lambda_s}\right)\right)^{-v_s} \neq 0 \text{ for each } t\in\mathbb{R}^d. \tag{15}$$

The price to pay requires $f_X \in L^2(\mathbb{R}^d)$ compactly supported, say supp $f_X \subset \Omega := [a,\ b]^d$. In general, if the Fourier transform of a noise density contains the factor $\prod_{s=1}^d(\sin\frac{\pi t_s}{\lambda_s})^{v_s}$, we call that noise Fourier-oscillating. When $f_\delta^{ft}(t) \neq 0$ for each $t\in\mathbb{R}^d$, we say the noise non-oscillating.

Motivated by Delaigle and Meister (2011), we introduce an auxiliary function $\tilde{f}_X \in L^2(\mathbb{R}^d)$ defined by

$$\tilde{f}_X^{ft}(t) := \left[\prod_{s=1}^{d}(\exp(i\frac{2\pi t_s}{\lambda_s})-1)^{v_s}\right]\cdot f_X^{ft}(t). \tag{16}$$

Note that $f_X\in L^2(\mathbb{R}^d)$ implies $f_X^{ft}\in L^2(\mathbb{R}^d)$, $|\prod_{s=1}^d(\exp(i2\pi t_s\lambda_s^{-1})-1)^{v_s}| \lesssim 1$. Then, $\tilde{f}_X$ in (16) exists uniquely in $L^2(\mathbb{R}^d)$.

Since $[\exp(i2\pi t_s\lambda_s^{-1})-1]^{v_s} = \sum_{k_s=0}^{v_s}\binom{v_s}{k_s}(-1)^{v_s-k_s}\exp(2\pi i k_s t_s\lambda_s^{-1})$,

$$\tilde{f}_X^{ft}(t) = \sum_{k_1=0}^{v_1}\cdots\sum_{k_d=0}^{v_d}\left[\prod_{s=1}^{d}\binom{v_s}{k_s}(-1)^{v_s-k_s}\exp\left(\frac{2\pi i k_s t_s}{\lambda_s}\right)\right]\cdot f_X^{ft}(t).$$

Taking the inverse Fourier transform, we find that

$$\begin{aligned}\tilde{f}_X(x_1,\ldots,x_d) = \sum_{k_1=0}^{v_1}\cdots\sum_{k_d=0}^{v_d}&\left[\prod_{s=1}^{d}\binom{v_s}{k_s}(-1)^{v_s-k_s}\right] f_X\\ &\left(x_1-\frac{2\pi k_1}{\lambda_1},\ldots,x_d-\frac{2\pi k_d}{\lambda_d}\right)\end{aligned}$$

holds almost everywhere on $\mathbb{R}^d$. With the notation $\frac{k}{\lambda} := (\frac{k_1}{\lambda_1},\frac{k_2}{\lambda_2},\ldots,\frac{k_d}{\lambda_d})$,

$$\tilde{f}_X(x_1,\ldots,x_d) := \sum_{k=\mathbf{0}}^{v}\left[\prod_{s=1}^{d}\binom{v_s}{k_s}(-1)^{v_s-k_s}\right] f_X\left(x-\frac{2\pi k}{\lambda}\right), \tag{17}$$

where $k=(k_1,\ldots,k_d)$, $v=(v_1,\ldots,v_d)$, $\mathbf{0}=(0,0,\ldots,0)$ and

$$\sum_{k=\mathbf{0}}^{v} := \sum_{k_1=0}^{v_1}\cdots\sum_{k_d=0}^{v_d}.$$

Let $J := (J_1,\ldots,J_d)$ and J_s stand for the smallest integer larger than or equal to $\frac{(b-a)\lambda_s}{2\pi}$ (denoted by $\lceil\frac{(b-a)\lambda_s}{2\pi}\rceil$), $1\le s\le d$. Then, those terms for $k_s > J_s$ do not contribute to the summation in (17), when $x \in \Omega := [a, b]^d$. It is important to note that f_X can be represented by linear combinations of $\tilde{f}_X(x - \frac{2\pi k}{\lambda})$, as in Delaigle and Meister (2011) for one-dimensional case. In fact, there exist $\eta_{k_1,\ldots,k_d}$, $(0 \le k_s \le J_s)$ such that for $x \in \Omega$,

$$f_X(x) = \sum_{k_1=0}^{J_1} \cdots \sum_{k_d=0}^{J_d} \eta_{k_1,\ldots,k_d} \tilde{f}_X\left(x_1 - \frac{2\pi k_1}{\lambda_1}, \ldots, x_d - \frac{2\pi k_d}{\lambda_d}\right). \tag{18}$$

For easily understanding, one proves (18) for $d = 2$ first. Define a vector $\tilde{F}(x)$ of dimension $(J_1 + 1)(J_2 + 1)$ by its transpose

$$\begin{aligned}\tilde{F}^T(x) : = \Bigg(&\tilde{f}_X(x_1,\ x_2),\ \tilde{f}_X\left(x_1,\ x_2 - \frac{2\pi}{\lambda_2}\right), \ldots,\ \tilde{f}_X\left(x_1,\ x_2 - \frac{2\pi J_2}{\lambda_2}\right),\\ &\tilde{f}_X\left(x_1 - \frac{2\pi}{\lambda_1},\ x_2\right),\ \tilde{f}_X\left(x_1 - \frac{2\pi}{\lambda_1},\ x_2 - \frac{2\pi}{\lambda_2}\right), \ldots,\ \tilde{f}_X\left(x_1 - \frac{2\pi}{\lambda_1},\ x_2 - \frac{2\pi J_2}{\lambda_2}\right),\\ &\cdots,\\ &\tilde{f}_X\left(x_1 - \frac{2\pi J_1}{\lambda_1},\ x_2\right),\ \tilde{f}_X\left(x_1 - \frac{2\pi J_1}{\lambda_1},\ x_2 - \frac{2\pi}{\lambda_2}\right), \ldots,\\ &\tilde{f}_X\left(x_1 - \frac{2\pi J_1}{\lambda_1},\ x_2 - \frac{2\pi J_2}{\lambda_2}\right)\Bigg).\end{aligned}$$

Similarly, $F(x)$ is defined by f_X. According to (17), there exists an upper triangular matrix Γ of order $(J_1 + 1)(J_2 + 1)$ such that

$$\tilde{F}(x) = \Gamma F(x)$$

with the diagonal component $(-1)^{v_1+v_2}$. Hence, Γ is invertible and $F(x) = \Gamma^{-1}\tilde{F}(x)$. Then, (18) follows from the definitions of $\tilde{F}(x)$ and $F(x)$.

To show (18) for $d \ge 3$, one uses $e_{k_1,k_2,\ldots,k_d} := (0, \ldots, 0, 1, 0, \ldots, 0)$ to denote the row vector of dimension $(J_1+1)(J_2+1)\ldots(J_d+1)$ with the $l_{k_1,k_2,\ldots,k_d}$-th coordinate being 1 and all others 0, where

$$l_{k_1,k_2,\ldots,k_d} := \sum_{s=1}^{d-1} k_s(J_{s+1} + 1)\ldots(J_d + 1) + k_d + 1.$$

Similar to the case $d = 2$, define two column vectors $\tilde{F}(x)$ and $F(x)$ by

$$\tilde{F}^T(x) = \sum_{s=1}^{d} \sum_{k_s=0}^{J_s} T_{k_1,k_2,\ldots,k_d} \tilde{f}_X(x) e_{k_1,k_2,\ldots,k_d}$$

and

$$F^T(x) = \sum_{s=1}^{d} \sum_{k_s=0}^{J_s} T_{k_1,k_2,\ldots,k_d} f_X(x) e_{k_1,k_2,\ldots,k_d},$$

where the translation operators

$$T_{k_1,k_2,\dots,k_d} f(x) := f\left(x_1 - \frac{2\pi k_1}{\lambda_1}, x_2 - \frac{2\pi k_2}{\lambda_2}, \dots, x_d - \frac{2\pi k_d}{\lambda_d}\right).$$

By (17), $\tilde{F}(x) = \Gamma F(x)$ for some upper triangular matrix Γ of order $(J_1+1)(J_2+1)\dots(J_d+1)$ with the diagonal component $(-1)^{v_1+v_2+\dots+v_d}$. Therefore, (18) holds for $d \geq 3$.

To introduce explicitly our estimators, we need to know the coefficients $\eta_{k_1,k_2,\dots,k_d}$ in (18). It is easy to see

$$\sum_{s=1}^{d}\sum_{k_s=0}^{J_s} \eta_{k_1,k_2,\dots,k_d} e_{k_1,k_2,\dots,k_d} = (1, 0, \dots, 0)\Gamma^{-1}$$

thanks to $F(x) = \Gamma^{-1}\tilde{F}(x)$ and the definitions of $F(x)$ and $\tilde{F}(x)$. Clearly, all elements of the matrix Γ are provided by the coefficients in (17). On the other hand, we can easily calculate Γ^{-1} because Γ is upper triangular.

With the notation $\eta_k := \eta_{k_1,\dots,k_d}$, (18) is rewritten as

$$f_X(x) = \sum_{k=\mathbf{0}}^{J} \eta_k\ \tilde{f}_X\left(x - \frac{2\pi k}{\lambda}\right). \tag{19}$$

Let $K(x) = C\prod_{s=1}^{d}[2x_s^{-1}\sin(2^{-1}x_s)]^{d+1}$ be the kernel function in Example 1 of Sect. 2, h_n be the bandwidth sequence for

$$f(t) := f_\delta^{ft}(t)\prod_{s=1}^{d}\left(\exp\left(2\pi i t_s \lambda_s^{-1}\right) - 1\right)^{-v_s}$$

in Lemma 2. Define

$$\tilde{f}_{X,n}(x) = \frac{1}{n}\sum_{l=1}^{n}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)K^{ft}(h_n t) \times \left[\prod_{s=1}^{d}\left(\exp\left(i\frac{2\pi t_s}{\lambda_s}\right) - 1\right)^{v_s}\right]\exp(it\cdot W_l)\big/ f_\delta^{ft}(t)\mathrm{d}t.$$

Then, compared with (19), it is natural to define that for $x \in \Omega$,

$$f_{X,n}(x) := \frac{1}{n}\sum_{l=1}^{n}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)K^{ft}(h_n t)\xi(t)\exp(it\cdot W_l)\big/ f_\delta^{ft}(t)\mathrm{d}t \tag{20}$$

and

$$p_n(x) := \frac{1}{n}\sum_{l=1}^{n} Y_l \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \exp(-it\cdot x)K^{ft}(h_n t)\xi(t)\exp(it\cdot W_l)\big/ f_\delta^{ft}(t)\mathrm{d}t, \tag{21}$$

where

$$\xi(t) = \left[\sum_{k=\mathbf{0}}^{J} \eta_k \exp\left(i\frac{2\pi k\cdot t}{\lambda}\right)\right]\prod_{s=1}^{d}\left[\exp\left(\frac{2\pi i t_s}{\lambda_s}\right)-1\right]^{v_s}. \tag{22}$$

Theorem 3 *Consider the problem* (1) *with* (15)*. If f_X, $p := mf_X \in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$ have compact support Ω, f_X^{ft}, $p^{ft} \in L^1(\mathbb{R}^d)$, $E|Y_1|^4 < +\infty$ and $f_X(x) \neq 0$, then the estimator $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ defined by* (20)–(22) *satisfies that for almost all $x \in \Omega$,*

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

Proof As in the proof of Theorem 2, it suffices to show

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

Since J is a fixed number (independent of t), $|\xi(t)| \lesssim \prod_{s=1}^{d} |\exp(\frac{2\pi i t_s}{\lambda_s})-1|^{v_s}$ thanks to (22). A careful inspection for the proof of (13) leads to

$$\begin{aligned} &E|p_n(x) - Ep_n(x)|^4 \\ &\quad\lesssim n^{-2}\left\{h_n^d \min_{t\in\left[-\frac{d+1}{2h_n},\, \frac{d+1}{2h_n}\right]^d}\left[|f_\delta^{ft}(t)|\prod_{s=1}^{d}\left|\left(\exp\left(\frac{2\pi i t_s}{\lambda_s}\right)-1\right)^{-v_s}\right|\right]\right\}^{-4} E|Y_1|^4. \end{aligned}$$

Combining this with Lemma 2 and (15), one obtains

$$E|p_n(x) - Ep_n(x)|^4 \lesssim n^{-2}\, n^{2/3} = n^{-4/3}. \tag{23}$$

The remaining and key work is to show that for almost all $x \in \Omega$,

$$\lim_{n\to\infty} Ep_n(x) = p(x). \tag{24}$$

For a non-oscillating noise, $Ep_n(x) = p * K_{h_n}(x)$ and (24) holds automatically due to Lemma 1. Although $Ep_n(x)$ is no longer a convolution form in this current case, one can still prove (24) as follows.

Similar to (16), one defines

$$\tilde{p}^{ft}(t) := \left[\prod_{s=1}^{d}\left(\exp\left(\frac{2\pi i t_s}{\lambda_s}\right)-1\right)^{v_s}\right] p^{ft}(t). \tag{25}$$

Then, $\tilde{p}(x) = \sum_{k=\mathbf{0}}^{v}[\prod_{s=1}^{d}\binom{v_s}{k_s}(-1)^{v_s-k_s}]p(x-\frac{2\pi k}{\lambda})$. These with the assumptions $p,\ p^{ft} \in L^1(\mathbb{R}^d)$ show $\tilde{p},\ \tilde{p}^{ft} \in L^1(\mathbb{R}^d)$. Hence,

$$\tilde{p}(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)\tilde{p}^{ft}(t)\mathrm{d}t \tag{26}$$

holds almost everywhere on $\mathbb{R}^d$.

Since supp $p \subset$ supp $f_X \subset \Omega$, similar arguments to (18) lead to

$$p(x) = \sum_{k=\mathbf{0}}^{J}\eta_k\ \tilde{p}\left(x-\frac{2\pi k}{\lambda}\right) \tag{27}$$

for $x \in \Omega$. Furthermore, it follows from (26) and (27) that

$$p(x) = \sum_{k=\mathbf{0}}^{J}\eta_k\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp\left(-it\cdot\left(x-\frac{2\pi k}{\lambda}\right)\right)\tilde{p}^{ft}(t)\mathrm{d}t.$$

This with (25) shows that

$$p(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)\xi(t)p^{ft}(t)\mathrm{d}t \tag{28}$$

for almost all $x \in \Omega$, where $\xi(t)$ is given by (22). It is important to note that (28) holds only for $x \in \Omega$, not for all $x \in \mathbb{R}^d$. Otherwise, it would contradict with p^{ft} being the Fourier transform of p.

On the other hand, one observes from (21) that

$$Ep_n(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)K^{ft}(h_nt)\xi(t)EY_1\exp(it\cdot W_1)\big/f_\delta^{ft}(t)\mathrm{d}t.$$

By the independent assumptions in model (1), $EY_1e^{it\cdot W_1} = EY_1e^{it\cdot X_1}Ee^{it\cdot\delta_1} = E(E(Y_1e^{it\cdot X_1}|X_1))\ f_\delta^{ft}(t) = E(E(Y_1|X_1)e^{it\cdot X_1})f_\delta^{ft}(t) = p^{ft}(t)f_\delta^{ft}(t)$. Hence,

$$Ep_n(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-it\cdot x)K^{ft}\,(h_nt)\,\xi(t)p^{ft}(t)\mathrm{d}t. \tag{29}$$

According to (28)–(29), $Ep_n(x) - p(x) = (2\pi)^{-d}\int_{\mathbb{R}^d}\exp(-it\cdot x)(K^{ft}(h_nt) - 1)\xi(t)p^{ft}(t)\mathrm{d}t$ and $\int_\Omega|Ep_n(x) - p(x)|^2\mathrm{d}x \le \int_{\mathbb{R}^d}|(2\pi)^{-d}\int_{\mathbb{R}^d}\exp(-it\cdot x)$

$(K^{ft}(h_n t) - 1)\xi(t)p^{ft}(t)\mathrm{d}t|^2\mathrm{d}x$. Using the Parseval identity, the boundedness of $|\xi(t)|$ and $|p^{ft}(t)|$, one receives that

$$\int_\Omega |Ep_n(x) - p(x)|^2 \mathrm{d}x \lesssim \int_{\mathbb{R}^d} |K^{ft}(h_n t) - 1|^2 |p^{ft}(t)|\mathrm{d}t.$$

By the choice of K, $\lim\limits_{h_n \to 0} K^{ft}(h_n t) = 1$ and $|K^{ft}(h_n t) - 1|^2 \lesssim 1$. These with the given assumption $p^{ft} \in L^1(\mathbb{R}^d)$ and the dominated convergence theorem show that

$$\lim_{n\to\infty} \int_\Omega |Ep_n(x) - p(x)|^2 \mathrm{d}x = 0. \tag{30}$$

Since Ep_n depends on h_n (see 29), one rewrites $Ep_n(x) := F_x(h_n)$ and finds that

$$\begin{aligned} |F_x(h_n) - F_x(h_m)| &\le \left| \frac{1}{(2\pi)^d} \int \exp(-it\cdot x) \left(K^{ft}(h_n t) - K^{ft}(h_m t)\right) \xi(t) p^{ft}(t)\mathrm{d}t \right| \\ &\lesssim \int \left| K^{ft}(h_n t) - K^{ft}(h_m t) \right| \left| p^{ft}(t) \right| \mathrm{d}t. \end{aligned}$$

When $n,\ m \to \infty$, $|F_x(h_n) - F_x(h_m)| \to 0$. As a Cauchy sequence of $\mathbb{R}$, $F_x(h_n) = Ep_n(x) \to g(x) (n \to \infty)$ pointwisely. By Fatou's Lemma and (30),

$$\begin{aligned} \int_\Omega |g(x) - p(x)|^2 \mathrm{d}t &= \int_\Omega \lim_{n\to\infty} |Ep_n(x) - p(x)|^2 \, \mathrm{d}x \\ &\le \underline{\lim}_{n\to\infty} \int_\Omega |Ep_n(x) - p(x)|^2 \, \mathrm{d}x = 0, \end{aligned}$$

which means $g \overset{a.e.}{=} p$ and hence, $\lim\limits_{n\to\infty} Ep_n(x) = p(x)$ for almost all $x \in \Omega$. This completes the proof of (24). □

Remark 2 Compared with Theorem 2, Theorem 3 allows for f_δ^{ft} having some zeros. The price to pay requires f_X compactly supported, $f_X,\ p \in L^2(\mathbb{R}^d)$ as well as $f_X^{ft},\ p^{ft} \in L^1(\mathbb{R}^d)$. It should be pointed out that the compact supportedness of f_X cannot be easily removed. In fact, the inequality

$$\begin{aligned} |\xi(t)| &:= \left| \left[\sum_{k=\mathbf{0}}^{J} \eta_k \exp\left(i \frac{2\pi k\cdot t}{\lambda}\right) \right] \prod_{s=1}^{d} \left[\exp\left(\frac{2\pi i t_s}{\lambda_s}\right) - 1 \right]^{v_s} \right| \\ &\lesssim \prod_{s=1}^{d} \left| \exp\left(\frac{2\pi i t_s}{\lambda_s}\right) - 1 \right|^{v_s} \end{aligned} \tag{31}$$

plays a key role in our proof of Theorem 3. Recall that $J := (J_1,\ J_2, \ldots,\ J_d)$ and $J_k = \lceil \frac{(b-a)\lambda_k}{2\pi} \rceil$ $(1 \le k \le d)$ with supp $f_X \subset [a, b]^d$. Then, $\sum_{k=\mathbf{0}}^{J} \eta_k \exp(i \frac{2\pi k\cdot t}{\lambda})$ may become a summation of infinitely many terms, if f_X is not compactly supported. Hence, it seems hard for us to conclude (31) in that case.

4 Consistency for wavelet estimators

This section aims to prove the strong consistency of wavelet regression estimators for both Fourier-oscillating and non-oscillating noises. We begin with a classical notation in wavelet analysis.

A Multiresolution Analysis (Meyer 1992) is a sequence of closed subspaces $\{V_j\}$ of the square integrable function space $L^2(\mathbb{R}^d)$ satisfying the following properties:

1. $V_j \subset V_{j+1}$, $j \in \mathbb{Z}$ (the integer set);
2. $\overline{\cup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^d)$ (the space $\cup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}^d)$);
3. $f(2x) \in V_{j+1}$ if and only if $f(x) \in V_j$ for all $j \in \mathbb{Z}$;
4. There exists $\varphi(x) \in L^2(\mathbb{R}^d)$ (scaling function) such that $\{\varphi(x-k)\}_{k\in\mathbb{Z}^d}$ forms an orthonormal basis of $V_0 = \overline{\mathrm{span}}\{\varphi(x-k)\}_{k\in\mathbb{Z}^d}$.

Many important wavelets are constructed by Multiresolution Analysis, which include Meyer and Daubechies wavelets.

Let P_j be the orthogonal projection operator from $L^2(\mathbb{R}^d)$ to the space V_j with the orthonormal basis $\{\varphi_{j,k}(x),\ k \in \mathbb{Z}^d\} := \{2^{jd/2}\varphi(2^j x - k),\ k \in \mathbb{Z}^d\}$. Then, for $f \in L^2(\mathbb{R}^d)$ and $\alpha_{j,k} := \langle f, \varphi_{j,k}\rangle$,

$$P_j f = \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}\varphi_{j,k} \tag{32}$$

holds in $L^2(\mathbb{R}^d)$. Moreover, $\lim\limits_{j\to\infty} \|P_j f - f\|_{L^2} = 0$. When the scaling function φ satisfies some additional conditions, $P_j f$ converges to f pointwisely.

Lemma 3 (Kelly et al. 1994) *Let a scaling function φ be bounded by an L^1 radially decreasing function. Then, for $f \in L^p(\mathbb{R}^d)(1 \le p \le \infty)$, $P_j f$ converges to f pointwise almost everywhere on $\mathbb{R}^d$.*

Example 2 If φ_0 is an orthonormal scaling function of dimension one, and $|\varphi_0(x)| \le (1+|x|)^{-d-1}$, then $\varphi(x) = \prod_{s=1}^d \varphi_0(x_s)$ is bounded by an L^1 radially decreasing function.

In fact, it is easy to see that $\prod_{s=1}^d (1+|x_s|)^{-2} \le (1+x_1^2+x_2^2+\cdots+x_d^2)^{-1}$, which implies that

$$|\varphi(x)| \le \prod_{s=1}^d (1+|x_s|)^{-d-1} \le \left(1+|x|^2\right)^{-\frac{d+1}{2}} := \Phi(|x|).$$

Clearly, $\Phi(|x|)$ is a radially decreasing function and $\int_{\mathbb{R}^d} \Phi(|x|)\mathrm{d}x < +\infty$.

To consider first the strong consistency of a wavelet regression estimator when $f_\delta^{ft}(t) \neq 0$ $(t \in \mathbb{R}^d)$, we choose the one-dimensional Meyer scaling function φ_M with supp $\varphi_M^{ft}(t) \subset [-\frac{4\pi}{3}, \frac{4\pi}{3}]$ and $\varphi_M^{ft} \in C^\infty$. Then,

$$\varphi(x) := \prod_{s=1}^d [\varphi_M](x_s)$$

satisfies the requirement of Lemma 3. Let P_j denote the corresponding projection operator in (32). Then, $P_j f_X = \sum_{k\in\mathbb{Z}^d} \alpha_{j,k}\varphi_{j,k}$, $P_j p = \sum_{k\in\mathbb{Z}^d} \gamma_{j,k}\varphi_{j,k}$ with $\alpha_{j,k} = \langle f_X, \varphi_{j,k}\rangle$ and $\gamma_{j,k} = \langle p, \varphi_{j,k}\rangle$. The estimators for f_X and p are defined, respectively, by

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

with

$$\hat{\alpha}_{j,k} = \frac{1}{n}\sum_{l=1}^{n} \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{it\cdot W_l}\overline{[\varphi_{j,k}]^{ft}(t)}\big/ f_\delta^{ft}(t)\mathrm{d}t, \tag{33}$$

$$\hat{\gamma}_{j,k} = \frac{1}{n}\sum_{l=1}^{n} Y_l \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{it\cdot W_l}\overline{[\varphi_{j,k}]^{ft}(t)}\big/ f_\delta^{ft}(t)\mathrm{d}t. \tag{34}$$

Note that $[\varphi_{j,k}]^{ft}$ is compactly supported and bounded, as well as f_δ^{ft} has no zeros. Then, $\hat{\alpha}_{j,k}$ and $\hat{\gamma}_{j,k}$ are well defined. In fact, $E\hat{\alpha}_{j,k} = \alpha_{j,k}$ and $E\hat{\gamma}_{j,k} = \gamma_{j,k}$, which can be easily proved. Moreover, the strong consistency holds for the wavelet estimator $\hat{m}_n(x) := p_n(x)/f_{X,n}(x)$ under the same assumptions as in Theorem 2.

Theorem 4 *Consider the problem* (1) *with* $f_\delta^{ft}(t) \neq 0$ *for each* $t \in \mathbb{R}^d$*. If* $p := mf_X \in L(\mathbb{R}^d)$*,* $E|Y_1|^4 < \infty$ *and* x *is a Lebesgue point of both* p *and* f_X *(*$f_X(x) \neq 0$*), then*

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

Proof Similar to Theorem 2, it is sufficient to prove $\lim_{n\to\infty} p_n(x) \overset{a.s.}{=} p(x)$. By the definition of $p_n(x)$,

$$\begin{aligned} &p_n(x) - Ep_n(x) \\ &= \frac{1}{n}\sum_{l=1}^{n}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(Y_l e^{it\cdot W_l} - EY_l e^{it\cdot W_l}\right)\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(t)}\big/ f_\delta^{ft}(t)\mathrm{d}t. \end{aligned}$$

With the notation $\Psi_l(t) := Y_l e^{it\cdot W_l} - EY_l e^{it\cdot W_l}$, this above identity reduces to

$$\begin{aligned} E|p_n(x) - Ep_n(x)|^4 \lesssim n^{-4}\sum_{l_1=1}^{n}\cdots\sum_{l_4=1}^{n}\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\Bigg\{\prod_{r=1}^{2}\Bigg[\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x) \\ \overline{[\varphi_{j,k}]^{ft}(t_{2r-1})}\Bigg]\Bigg[\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(-t_{2r})}\Bigg]\left[f_\delta^{ft}(t_{2r-1})\, f_\delta^{ft}(-t_{2r})\right]^{-1}\Bigg\} \end{aligned}$$

$$E\prod_{r=1}^{2}\Psi_{l_{2r-1}}(t_{2r-1})\Psi_{l_{2r}}(-t_{2r})\mathrm{d}t_1\ldots\mathrm{d}t_4. \tag{35}$$

Note that $[\varphi_{j,k}]^{ft}(t)=2^{-\frac{dj}{2}}\prod_{s=1}^{d}[\varphi_M]^{ft}(2^{-j}t_s)e^{it_sk_s2^{-j}}$, $\varphi_{j,k}(x)=2^{jd/2}\,\varphi(2^jx-k)$ and $\sup\limits_{x\in\mathbb{R}^d}\sum\limits_{k\in\mathbb{Z}^d}|\varphi(x-k)|\lesssim 1$. Then,

$$\left|\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(t)}\right|\lesssim\left|\prod_{s=1}^{d}[\varphi_M]^{ft}\left(2^{-j}t_s\right)\right|.$$

Combining this with supp $\varphi_M^{ft}\subset[-\frac{4\pi}{3},\ \frac{4\pi}{3}]$, one knows that

$$\int_{\mathbb{R}^d}\left|\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\frac{\overline{[\varphi_{j,k}]^{ft}(t)}}{f_\delta^{ft}(t)}\right|\mathrm{d}t\lesssim\left[\left(\frac{4\pi}{3}2^j\right)^{-d}\min_{t\in\left[-\frac{4\pi 2^j}{3},\ \frac{4\pi 2^j}{3}\right]^d}\left|f_\delta^{ft}(t)\right|\right]^{-1}. \tag{36}$$

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

$$h_n^d\min_{t\in\left[-\frac{1}{h_n},\ \frac{1}{h_n}\right]^d}\left|f_\delta^{ft}(t)\right|\geq n^{-\frac{1}{6}}.$$

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

$$j:=\left\lfloor\log_2\left(\frac{3}{4\pi h_n}\right)\right\rfloor>0 \tag{37}$$

($\lfloor x\rfloor$ denotes the largest integer less than or equal to x). Clearly, $j\leq\log_2(\frac{3}{4\pi h_n})$ and $\frac{4\pi}{3}2^j\leq\frac{1}{h_n}$. Furthermore, (36) becomes

$$\int_{\mathbb{R}^d}\left|\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\frac{\overline{[\varphi_{j,k}]^{ft}(t)}}{f_\delta^{ft}(t)}\right|\mathrm{d}t\lesssim\left[h_n^d\min_{t\in\left[-\frac{1}{h_n},\frac{1}{h_n}\right]^d}\left|f_\delta^{ft}(t)\right|\right]^{-1}\leq n^{\frac{1}{6}}.$$

This with (35) leads to $E|p_n(x)-Ep_n(x)|^4\lesssim n^{-\frac{4}{3}}$, as in the proof of Theorem 2. Hence, $P(|p_n(x)-Ep_n(x)|\geq\varepsilon/2)\leq(\varepsilon/2)^{-4}E|p_n(x)-Ep_n(x)|^4\lesssim n^{-\frac{4}{3}}$. Because $P[|p_n(x)-p(x)|\geq\varepsilon]\leq P[|p_n(x)-Ep_n(x)|\geq\varepsilon/2]+P[|Ep_n(x)-p(x)|\geq\varepsilon/2]$, it remains to show

$$\lim_{n\to\infty}|Ep_n(x)-p(x)|=0. \tag{38}$$

By the definition of $\hat{\gamma}_{j,k}$ in (34), one finds easily

$$E\hat{\gamma}_{j,k} = (2\pi)^{-d} \int_{\mathbb{R}^d} p^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t.$$

If $p \in L^2(\mathbb{R}^d)$, then the Plancherel formula would imply

$$E\hat{\gamma}_{j,k} = \gamma_{j,k} \tag{39}$$

and $Ep_n(x) = P_j p(x)$. According to (37), j goes to $+\infty$ as $n \to \infty$. Then, (38) follows from Lemma 3 and Example 2. However, one assumes $p \in L^1(\mathbb{R}^d)$ only in Theorem 4. Fortunately,

$$(2\pi)^{-d} \int_{\mathbb{R}^d} p^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t = \int_{\mathbb{R}^d} p(x)\overline{\varphi_{j,k}(x)}\mathrm{d}x$$

still holds in the current case, which can be checked by the following arguments.

For $p \in L^1(\mathbb{R}^d)$, there exists $p_n \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ such that $\lim\limits_{n\to\infty} \|p_n - p\|_{L^1(\mathbb{R}^d)} = 0$. Since $p_n, \varphi_{j,k} \in L^2(\mathbb{R}^d)$,

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} p_n^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t = \int_{\mathbb{R}^d} p_n(x)\overline{\varphi_{j,k}(x)}\mathrm{d}x. \tag{40}$$

Clearly, $|\int_{\mathbb{R}^d} p_n^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t - \int_{\mathbb{R}^d} p^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t| \le \int_{\mathbb{R}^d} |p_n^{ft}(t) - p^{ft}(t)|\, |[\varphi_{j,k}]^{ft}(t)|\mathrm{d}t$. Because $\lim\limits_{n\to\infty} \|p_n - p\|_{L^1(\mathbb{R}^d)} = 0$, $\lim\limits_{n\to\infty} (p_n - p)^{ft}(t) = 0$ uniformly and for fixed j, k,

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} p_n^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t = \int_{\mathbb{R}^d} p^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t \tag{41}$$

thanks to $[\varphi_{j,k}]^{ft} \in L^1(\mathbb{R}^d)$. On the other hand, $|\int_{\mathbb{R}^d} p_n(x)\overline{\varphi_{j,k}(x)}\mathrm{d}x - \int_{\mathbb{R}^d} p(x)\, \overline{\varphi_{j,k}(x)}\mathrm{d}x| \lesssim \int_{\mathbb{R}^d} |p_n(x) - p(x)|\mathrm{d}x = \|p_n - p\|_{L^1(\mathbb{R}^d)} \to 0$ due to the boundedness of $\varphi_{j,k}(x)$, which means $\lim\limits_{n\to\infty} \int_{\mathbb{R}^d} p_n(x)\overline{\varphi_{j,k}(x)}\mathrm{d}x = \int_{\mathbb{R}^d} p(x)\overline{\varphi_{j,k}(x)}\mathrm{d}x$ for fixed j, k. This with (40) and (41) concludes the desired identity

$$\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} p^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t = \int_{\mathbb{R}^d} p(x)\overline{\varphi_{j,k}(x)}\mathrm{d}x.$$

The proof of Theorem 4 is finished. □

Remark 3 Chesneau (2010) provides a nice convergence rate of some wavelet estimators over L^2 risk for the same model, when the estimated regression function m is smooth. In fact, his estimators do not depend on the smoothness index (so adaptive). Our Theorem 4 shows the strong consistency of a wavelet estimator in the pointwise sense. Since we do not assume any regularity of the function m, our estimator can be considered adaptive as well in some sense.

Remark 4 Compared with model (1), Gencay and Gradojevic (2011) study a linear regression model in which both the regressor and regressand have measurement errors. Using the discrete wavelet transformation, they employ extensive simulations and demonstrate their approach better than the traditional methods. It is interesting to do further theoretic research in that area.

To extend Theorem 4 from a non-oscillating to a Fourier-oscillating case whose density function f_δ satisfies

$$\left|f_\delta^{ft}(t_1,\ldots,t_d)\right| \geq \prod_{s=1}^{d}\left|\sin\left(\frac{\pi t_s}{\lambda_s}\right)\right|^{v_s}(1+|t_s|)^{-\alpha_s} \tag{42}$$

with $\lambda_s > 0$, $\alpha_s \geq 0$ and $v_s \in \mathbb{Z}^+ \cup \{0\}$, we need to assume $f_X \in L^2(\mathbb{R}^d)$ and $\operatorname{supp} f_X \subset \Omega := [a,\ b]^d$ as in Sect. 3.

Recall that the Meyer scaling function is used in Theorem 4. However, we need the Daubechies scaling function $\varphi_N = D_{2N}$ (for large N) to define our estimator in the current case, because it seems hard to get the asymptotic unbiased property using the Meyer function (see the discussions below). It is well known that the support length of φ_N is $2N-1$. We define $\tilde{\Omega} := [a-2N+1, b+2N-1]^d$. With

$$\tilde{f}_X^{ft}(t) := \left[\prod_{s=1}^{d}\left(\exp\left(i\frac{2\pi t_s}{\lambda_s}\right)-1\right)^{v_s}\right] f_X^{ft}(t), \tag{43}$$

similar arguments to (19) show that

$$f_X(x) = \sum_{k=\mathbf{0}}^{\tilde{J}} \eta_k\ \tilde{f}_X\left(x - \frac{2\pi k}{\lambda}\right), \quad \forall x \in \tilde{\Omega}, \tag{44}$$

where $\eta_k := \eta_{k_1,\ldots,k_d}$, $\frac{k}{\lambda} := (\frac{k_1}{\lambda_1}, \frac{k_2}{\lambda_2}, \ldots, \frac{k_d}{\lambda_d})$, $\tilde{J} = (\tilde{J}_1, \tilde{J}_2 \ldots, \tilde{J}_d)$ with $\tilde{J}_s = \lceil\frac{(\tilde{b}-\tilde{a})\lambda_s}{2\pi}\rceil$ and $\tilde{a} := a-2N+1$, $\tilde{b} := b+2N-1$.

Let $\varphi(x) = \prod_{s=1}^{d}\varphi_N(x_s)$, $\varphi_{j,k}(x) = 2^{jd/2}\varphi(2^j x - k)$ and $K_j := \{k = (k_1,\ k_2,\ldots,k_d) \in \mathbb{Z}^d : \operatorname{supp}\ (\varphi_N)_{j,k_s} \cap [a,b] \neq \emptyset, s = 1,2,\ldots,d\}$. Define an estimator for f_X by

$$f_{X,n}(x) := \sum_{k\in K_j}\hat{\alpha}_{j,k}\varphi_{j,k}(x), \tag{45}$$

where

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

with

$$\xi(t)=\left(\sum_{m=\mathbf{0}}^{\tilde{J}}\eta_m e^{i\frac{2\pi m}{\lambda}\cdot t}\right)\left[\prod_{s=1}^{d}\left(e^{\frac{2\pi i t_s}{\lambda_s}}-1\right)^{v_s}\right]. \tag{46}$$

Then, $E\hat{\alpha}_{j,k}=(2\pi)^{-d}\int_{\mathbb{R}^d}\xi(t)f_X^{ft}(t)\overline{[\varphi_{j,k}]^{ft}(t)}\mathrm{d}t$, thanks to $Ee^{it\cdot W_l}=f_X^{ft}(t)f_\delta^{ft}(t)$. This with (43) and (46) leads to

$$E\hat{\alpha}_{j,k}=\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left[\sum_{k=\mathbf{0}}^{\tilde{J}}\eta_m\tilde{f}_X\left(x-\frac{2\pi m}{\lambda}\right)\right]^{ft}(t)\overline{[\varphi_{j,k}]}^{ft}(t)\mathrm{d}t. \tag{47}$$

By Plancherel's formula, (47) reduces to

$$E\hat{\alpha}_{j,k}=\int_{\mathbb{R}^d}\left[\sum_{k=\mathbf{0}}^{\tilde{J}}\eta_m\tilde{f}_X\left(x-\frac{2\pi m}{\lambda}\right)\right]\varphi_{j,k}(x)\mathrm{d}x \tag{48}$$

because of $f_X\in L^2(\mathbb{R}^d)$ (and therefore $\tilde{f}_X\in L^2(\mathbb{R}^d)$). Note that $\Omega:=[a,\ b]^d$, $\tilde{\Omega}=[a-2N+1,b+2N-1]^d$ and $K_j:=\{k=(k_1,\ k_2,\ldots,k_d)\in\mathbb{Z}^d:\mathrm{supp}\,(\varphi_N)_{j,k_s}\cap[a,b]\neq\emptyset,\ s=1,2,\ldots,d\}$. Then, supp $\varphi_{j,k}\subset\tilde{\Omega}$ for $k\in K_j$ with $j\geq 0$. This with (48) and (44) leads to $E\hat{\alpha}_{j,k}=\int_{\tilde{\Omega}}f_X(x)\varphi_{j,k}(x)\mathrm{d}x=\langle f_X,\varphi_{j,k}\rangle=\alpha_{j,k}$.

Similarly, the estimator $p_n(x)$ of $p(x)$ is defined by

$$p_n(x):=\sum_{k\in K_j}\hat{\gamma}_{j,k}\varphi_{j,k}(x) \tag{49}$$

with

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

Then, for $k\in K_j$ and $j\geq 0$, $E\hat{\gamma}_{j,k}=\gamma_{j,k}$ and $Ep_n(x)=P_jp(x)$. Finally, define

$$\hat{m}_n(x):=p_n(x)/f_{X,n}(x). \tag{50}$$

The same arguments as (35) show that

$$E|p_n(x)-Ep_n(x)|^4\lesssim n^{-4}\sum_{l_1=1}^{n}\cdots\sum_{l_4=1}^{n}\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\left\{\prod_{r=1}^{2}\xi(t_{2r-1})\xi(-t_{2r})\right.$$
$$\times\left[\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(t_{2r-1})}\right]\left[\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(-t_{2r})}\right]\left[f_\delta^{ft}(t_{2r-1})\right.$$

$$f_\delta^{ft}(-t_{2r})\Big]^{-1}\Big\} E\prod_{r=1}^{2}\Psi_{l_{2r-1}}(t_{2r-1})\Psi_{l_{2r}}(-t_{2r})\mathrm{d}t_1\ldots\mathrm{d}t_4, \tag{51}$$

where $\varphi(x) = \prod_{s=1}^{d}\varphi_N(x_s)$ and $\xi(t)$ is defined in (46). As in Theorem 2, $E\prod_{r=1}^{2}\Psi_{l_{2r-1}}(t_{2r-1})\Psi_{l_{2r}}(-t_{2r}) \lesssim 1$, and $E\prod_{r=1}^{2}\Psi_{l_{2r-1}}(t_{2r-1})\Psi_{l_{2r}}(-t_{2r}) = 0$ when $\sharp\{l_1,\ldots,l_4\} > 2$. These with (51) show that

$$\begin{aligned}
&E|p_n(x) - Ep_n(x)|^4\\
&\lesssim n^{-4}\sum_{(l_1,\ldots,l_4)\in L_n}\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\prod_{r=1}^{2}\xi(t_{2r-1})\xi(-t_{2r})\Big[\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(t_{2r-1})}\Big]\\
&\quad\times\Big[\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)\overline{[\varphi_{j,k}]^{ft}(-t_{2r})}\Big]\Big[f_\delta^{ft}(t_{2r-1})f_\delta^{ft}(-t_{2r})\Big]^{-1}\mathrm{d}t_1\ldots\mathrm{d}t_4,
\end{aligned} \tag{52}$$

where $L_n := \{\{l_1,\ldots,l_4\} : \sharp\{l_1,\ldots,l_4\} \le 2\}$. Using (42), $|\xi(t)| \lesssim \prod_{s=1}^{d}|e^{2\pi i t_s\lambda_s^{-1}} - 1|^{v_s}$ and $[\varphi_{j,k}]^{ft}(t) = 2^{-dj/2}\prod_{s=1}^{d}e^{it_sk_s2^{-j}}\varphi_N^{ft}(2^{-j}t_s)$, we obtain that

$$\int_{\mathbb{R}^d}|\xi(t)|\left|\frac{\sum_{k\in\mathbb{Z}^d}\varphi_{j,k}(x)[\varphi_{j,k}]^{ft}(t)}{f_\delta^{ft}(t)}\right|\mathrm{d}t \lesssim \int_{\mathbb{R}^d}\prod_{s=1}^{d}(1+|t_s|)^{\alpha_s}\left|\varphi_N^{ft}(2^{-j}t_s)\right|\mathrm{d}t.$$

Note that for large N,

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

Then, (52) reduces to

$$E\,|p_n(x) - Ep_n(x)|^4 \lesssim n^{-2}\left[2^{\sum_{s=1}^{d}(1+\alpha_s)j}\right]^4 \lesssim n^{-\frac{4}{3}} \tag{53}$$

by choosing j such that $2^{\sum_{s=1}^{d}(1+\alpha_s)j} \backsimeq n^{\frac{1}{6}}$. The remaining proofs are the same as in Theorem 4. We summarize our findings as follows:

Theorem 5 *Consider the problem* (1) *with* (42)*. If f_X, $p := mf_X \in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$ have compact support Ω, $E|Y_1|^4 < \infty$, $x \in \Omega$ is a Lebesgue point of both f_X and p ($f_X(x) \neq 0$), then $\hat{m}_n(x)$ defined by* (45), (49) *and* (50) *satisfies*

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

Remark 5 The noise condition (42) is a little stronger than (15). However, we do not assume f_X^{ft}, $p^{ft} \in L^1(\mathbb{R}^d)$ as in Theorem 3. This shows some differences between kernel estimators and wavelet ones.

5 Concluding remarks

In this paper, we study the strong consistency of an estimator $\hat{m}_n$ of the functions m (on $\mathbb{R}^d$) for the regression problem

$$Y_j = m\left(X_j\right) + \varepsilon_j, \quad W_j = X_j + \delta_j \quad (j = 1,\ 2, \ldots, n),$$

where f_X and f_δ denote the densities of X_j and δ_j, ε_j and δ_j are independent of each other and independent of X_j, as well as $E\varepsilon_j = 0$.

Theorem 2 extends Meister's theorem from one-dimensional to multidimensional setting. We remove the boundedness assumption of f_X and p ($p := mf_X$), relax the continuity to a Lebesgue point of f_X and p; Theorem 3 allows for the noise density f_δ having Fourier-oscillating zeros, which produces a difficulty that Ep_n is no longer a convolution form. However, we find a new proof for $\lim_{n\to\infty} Ep_n(x) = p(x)$.

Theorems 4–5 deal with the strong consistency of wavelet estimators. We use the Meyer wavelet for non-oscillating noises (Theorem 4). All conditions of Theorem 4 and Theorem 2 are exactly same. However, the proofs for $\lim_{n\to\infty} Ep_n(x) = p(x)$ are totally different. The Daubechies wavelet is chosen to study Fourier-oscillating noises (Theorem 5). The key point is that we use compact supportedness of the Daubechies function in the proof of Theorem 5.

We may ask if the strong consistency holds for a thresholding wavelet estimator. The following discussions tell some difficulties, if we use the method in Sect. 4.

For simplicity, we assume the dimension $d = 1$ and consider only the non-oscillating noise (i.e., $f_\delta^{ft}(t) \neq 0$). As usual, a thresholding wavelet estimator is defined by

$$p_n(x) = \sum_k \hat{\alpha}_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j=j_0}^{j_1-1}\sum_k \hat{\beta}^*_{j,k}\psi_{j,k}(x).$$

Here, ψ is the wavelet function corresponding to the scaling one φ and

$$\hat{\alpha}_{j,k} = \frac{1}{n}\sum_{l=1}^{n} Y_l \frac{1}{2\pi}\int_{\mathbb{R}} 2^{-j/2} e^{itW_l} e^{-itk2^{-j}} \overline{\varphi^{ft}(2^{-j}t)} \Big/ f_\delta^{ft}(t)\mathrm{d}t. \tag{54}$$

The coefficients $\hat{\beta}_{j,k}$ are given by replacing φ by ψ in (54), and $\hat{\beta}^*_{j,k} = \hat{\beta}_{j,k} I_{\{|\hat{\beta}_{j,k}|>\lambda\}}$. Similar to (39), $E\hat{\beta}_{j,k} = \beta_{j,k}$ with $\beta_{j,k} = \langle p, \psi_{j,k}\rangle$.

As in the proof of Theorem 4, we need to estimate $P\{|p_n(x) - p(x)| \geq \varepsilon\}$. Clearly,

$$\begin{aligned} P\left(|p_n(x) - p(x)| \geq \varepsilon\right) &\leq P\left(\left|\sum_k \left(\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}\right)\varphi_{j_0,k}(x)\right| \geq \varepsilon/3\right) \\ &+ P\left(\left|\sum_{j=j_0}^{j_1-1}\sum_k \left(\hat{\beta}^*_{j,k} - \beta_{j,k}\right)\psi_{j,k}(x)\right| \geq \varepsilon/3\right) + P\left(\left|\sum_{j=j_1}^{\infty}\sum_k \beta_{j,k}\psi_{j,k}(x)\right| \geq \varepsilon/3\right). \end{aligned}$$

The first and third terms of the above equality can be treated similarly as in the proof of Theorem 4. However, the middle one is totally different, because it involves the coefficients $\hat{\beta}^*_{j,k}$.

Let us look at

$$T = P\left[\left|\sum_{j=j_0}^{j_1-1}\sum_k\left(\hat{\beta}_{j,k}-\beta_{j,k}\right)\psi_{j,k}(x)I_{\left\{\left|\hat{\beta}_{j,k}\right|>\lambda\right\}}\right|\geq\varepsilon\right].$$

By Markov's inequality,

$$T \leq \varepsilon^{-4}(j_1-j_0)^3\sum_{j=j_0}^{j_1-1}E\left|\sum_k\left(\hat{\beta}_{j,k}-\beta_{j,k}\right)\psi_{j,k}(x)I_{\left\{\left|\hat{\beta}_{j,k}\right|>\lambda\right\}}\right|^4. \tag{55}$$

Similar to (35),

$$\begin{aligned} &E\left|\sum_k\left(\hat{\beta}_{j,k}-\beta_{j,k}\right)\psi_{j,k}(x)I_{\left\{\left|\hat{\beta}_{j,k}\right|>\lambda\right\}}\right|^4 \\ &= \frac{1}{n^4}\sum_{l_1=1}^{n}\cdots\sum_{l_4=1}^{n}\frac{2^{-2j}}{(2\pi)^4}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}[E\ F(t,x)]G(t,x)\mathrm{d}t_1\ldots\mathrm{d}t_4, \end{aligned} \tag{56}$$

where $G(t,x)=\prod_{r=1}^{2}\overline{\psi^{ft}(2^{-j}t_{2r-1})\psi^{ft}(-2^{-j}t_{2r})}[f_\delta^{ft}(t_{2r-1})f_\delta^{ft}(-t_{2r})]^{-1}$ and

$$\begin{aligned} F(t,x) = \prod_{r=1}^{2}\Psi_{l_{2r-1}}\left(t_{2r-1}\right)\Psi_{l_{2r}}(-t_{2r})&\left[\sum_k e^{-it_{2r-1}k2^{-j}}\psi_{j,k}(x)I_{\left\{\left|\hat{\beta}_{j,k}\right|>\lambda\right\}}\right] \\ \times\left[\sum_k e^{it_{2r}k2^{-j}}\psi_{j,k}(x)I_{\left\{\left|\hat{\beta}_{j,k}\right|>\lambda\right\}}\right] &\qquad \text{with } \Psi_l(t):=Y_le^{itW_l}-EY_le^{itW_l}. \end{aligned}$$

Although $G(t,x)$ can be estimated by the same way as in (35), $F(t,x)$ is much more complicated: In addition to $\Psi_{l_{2r-1}}(t_{2r-1})\Psi_{l_{2r}}(-t_{2r})$, $F(t,x)$ contains additional stochastic terms related to $I_{\{|\hat{\beta}_{j,k}|>\lambda\}}$. Therefore, the number of the summation terms in (56) does not reduce to $O(n^2)$ in general. Thus, the Borel–Cantelli lemma cannot be applied for our desired conclusion.

Finally, we want to point out that the current paper focuses only on theoretic investigations. There are existing references dealing with numerical experiments for kernel estimators with both non-oscillating and oscillating noises. We shall try to do the same things for our wavelet estimators in future.

Acknowledgements This paper is supported by the National Natural Science Foundation of China (No. 11271038). The authors would like to thank two referees and an editor for their many good suggestions and comments.

References

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., 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.

Delaigle, A., & Meister, A. (2011). Nonparametric function estimation under Fourier-oscillating noise. *Statistica Sinica*, *21*, 1065–1092.

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

Gencay, R., & Gradojevic, N. (2011). Errors-in-variables estimation with wavelets. *Journal of Statistical Computation and Simulation*, *81*, 1545–1564.

Guo, H. J. and Liu, Y. M. (2014). Wavelet estimations for densities and their derivatives with Fourier oscillating noises. Journal of Inequalities and Applications, Article ID: 236.

Hall, P., & Meister, A. (2007). A ridge-parameter approach to deconvolution. *Annals of Statistics*, *35*, 1535–1558.

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

Li, Y. M., Yang, S. C., & Zhou, Y. (2008). Consistency and uniformly asymptotic normality of wavelet estimator in regression model with associated samples. *Statistics and Probability Letters*, *78*, 2947–2956.

Meister, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems*, *24*, 015003.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Berlin: Springer.

Meyer, Y. (1992). *Wavelets and operators*. Cambridge: Cambridge University Press.

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.

Stein, E., & Shakarchi, R. (2005). *Real Analysis*: *Measure Theory, Integration, and Hilbert Spaces*. Princeton: Princeton University Press.

Sun, J., Morrison, H., Harding, P. & Woodroofe, M. (2002). Density and mixture estimation from data with measurement errors. Manuscript. http://sun.cwru.edu/~jiayang/india.ps.