İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
Vol. 6, No. 1, January 2013, pp. 10–26
ISSN 1300-4077 | 13 | 1 | 10 | 26

İSTATİSTİK

A NOTE ON THE ADAPTIVE ESTIMATION OF A QUADRATIC FUNCTIONAL FROM DEPENDENT OBSERVATIONS

Christophe Chesneau[1], Maher Kachour[3] and Fabien Navarro[1,2]
[1]LMNO, Université de Caen Basse-Normandie, Département de Mathématiques,
UFR de Sciences, 14032 Caen Cedex, France. E-mail: christophe.chesneau@unicaen.fr.
[2]GREYC CNRS-ENSICAEN-Université de Caen,
14050 Caen Cedex, France. E-mail: fabien.navarro@ensicaen.fr.
[3]École supérieure de commerce IDRAC, 47, rue Sergent Michel Berthet CP 607,
69258 Lyon Cedex 09, France. E-mail: maher.kachour@idraclyon.com.

Abstract : We investigate the estimation of the integral of the square of a multidimensional unknown function f under mild assumptions on the model allowing dependence on the observations. We develop an adaptive estimator based on a plug-in approach and wavelet projections. Taking the mean absolute error and assuming that f has a certain degree of smoothness, we prove that our estimator attains a sharp rate of convergence. Applications are given for the biased density model, the nonparametric regression model and a GARCH-type model under some mixing dependence conditions (α-mixing or β-mixing). A simulation study considering nonparametric regression models with dependent observations illustrates the usefulness of the proposed estimator.

Key words: Quadratic functional estimation, Plug-in approach, Wavelets, Rates of convergence, Mixing dependence.
History: Submitted: 13 June 2012; Revised: 20 September 2012; Accepted: 9 October 2012

1. Introduction

Let d be a positive integer, $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, Z be a random variable vector on $\mathbb{R}^d$, $f : [0,1]^d \to \mathbb{R}$ be an unknown squared integrable function related to Z (as a density function, a regression function,...) and Q_f be the quadratic functional:

$$Q_f = \int_{[0,1]^d} f^2(\mathbf{x})d\mathbf{x}. \tag{1.1}$$

We aim to estimate Q_f from n identical distributed observations $Z_1, \ldots, Z_n$ of Z.

When $d = 1$ and $Z_1, \ldots, Z_n$ are independent, this problem has been addressed in many papers for a wide variety of models under various settings. See, e.g., [2], [16], [27], [20], [34], [25, 26], [14], [28], [6, 7], [21], [33] and [5]. The multidimensional case has been considered by [1] for the density model. When $d = 1$ and $Z_1, \ldots, Z_n$ are dependent, the estimation of Q_f has been investigated by [24] for the density model and by [4] for the density deconvolution model. A common feature is that when f has a certain degree of smoothness the parametric rate of convergence "$1/\sqrt{n}$" is achievable.

The main contribution of this paper is to present new theoretical results in a general multidimensional nonparametric setting. "General" in the sense that it includes a wide variety of models with possible dependent $Z_1, \ldots, Z_n$. In the first part, we develop a simple adaptive estimator for Q_f based on a plug-in approach and wavelet methodology. We refer to, e.g., [23] and [38] for detailed discussions on the performances of wavelet estimators and some of their advantages over traditional methods. The asymptotic performances of our estimator are evaluated under the mean absolute error (MAE) over a wide range of function class for f. Under mild assumptions on $Z_1, \ldots, Z_n$, we prove that it attains a sharp rate of convergence (which can be $1/\sqrt{n}$ in some situations). Then we apply our general result to three different models under mixing dependence conditions. To be more specific, we consider the biased density model with α-mixing observations, the nonparametric regression model with α-mixing observations and a GARCH-type model with β-mixing observations. These mixing dependence structures are reasonably weak and particularly interesting in the considered nonparametric models thanks to their numerous applications in dynamic economic systems and financial time series. See, e.g., [40], [22] and [17]. Let us mention that, to the best of our knowledge, the obtained results are new for these statistical frameworks. Finally, a small simulation study is provided in the context of nonparametric regression models with dependent observations illustrating the usefulness of the proposed estimator in finite sample situations.

The rest of the paper is organized as follows. In Section 2, we present some preliminaries on wavelets. In Section 3 we describe our wavelet estimator and study its asymptotic properties. Applications are given in Section 4. Section 5 is devoted to a simulation study. Finally, the proofs are postponed to Section 6.

2. Preliminaries on wavelets

In this section, we briefly present the wavelet tensor-product bases on $[0,1]^d$ and the considered function space in term of wavelet coefficients.

2.1. Wavelet tensor-product bases on $[0,1]^d$

For the purpose of this paper, we use a compactly supported wavelet-tensor product basis on $[0,1]^d$ based on the Daubechies wavelets.

Let N be a positive integer, ϕ be "father" Daubechies-type wavelet and ψ be a "mother" Daubechies-type wavelet of the family $db2N$. In particular, mention that ϕ and ψ have compact supports (see [30]).

For any $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$, we consider 2^d functions as follows:

- A scale function Φ defined by

$$\Phi(\mathbf{x}) = \prod_{v=1}^{d} \phi(x_v),$$

- $2^d - 1$ wavelet functions $(\Psi_u)_{u\in\{1,\ldots,2^d-1\}}$ defined by

$$\Psi_u(\mathbf{x}) = \begin{cases} \psi(x_u) \prod_{\substack{v=1 \\ v\neq u}}^{d} \phi(x_v) & \text{when } u \in \{1,\ldots,d\}, \\ \prod_{v\in A_u} \psi(x_v) \prod_{v\notin A_u} \phi(x_v) & \text{when } u \in \{d+1,\ldots,2^d-1\}, \end{cases}$$

where $(A_u)_{u\in\{d+1,\ldots,2^d-1\}}$ forms the set of all the non void subsets of $\{1,\ldots,d\}$ of cardinal superior or equal to 2.

For any integer j and any $\mathbf{k} = (k_1, \ldots, k_d)$, we set

$$\Phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\Phi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d),$$

for any $u \in \{1, \ldots, 2^d - 1\}$,

$$\Psi_{j,\mathbf{k},u}(\mathbf{x}) = 2^{jd/2}\Psi_u(2^j x_1 - k_1, \ldots, 2^j x_d - k_d).$$

We set $D_j = \{0, \ldots, 2^j - 1\}^d$. Then, with an appropriate treatment at the boundaries, there exists an integer τ such that the system

$$\mathcal{S} = \{\Phi_{\tau,\mathbf{k}}, \mathbf{k} \in D_\tau;\ (\Psi_{j,\mathbf{k},u})_{u\in\{1,\ldots,2^d-1\}},\quad j \in \mathbb{N} - \{0,\ldots,\tau-1\},\ \mathbf{k} \in D_j\}$$

forms an orthonormal basis of $\mathbb{L}_2([0,1]^d) = \{h : [0,1]^d \to \mathbb{R};\ \int_{[0,1]^d} h^2(\mathbf{x})d\mathbf{x} < \infty\}$.

A function $h \in \mathbb{L}_2([0,1]^d)$ can be expressed via $\mathcal{S}$ as wavelet series as

$$h(\mathbf{x}) = \sum_{\mathbf{k}\in D_\tau} \alpha_{\tau,\mathbf{k}}\Phi_{\tau,\mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1}\sum_{j=\tau}^{\infty}\sum_{\mathbf{k}\in D_j} \beta_{j,\mathbf{k},u}\Psi_{j,\mathbf{k},u}(\mathbf{x}), \qquad \mathbf{x}\in[0,1]^d, \tag{2.1}$$

where

$$\alpha_{j,\mathbf{k}} = \int_{[0,1]^d} h(\mathbf{x})\Phi_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x}, \qquad \beta_{j,\mathbf{k},u} = \int_{[0,1]^d} h(\mathbf{x})\Psi_{j,\mathbf{k},u}(\mathbf{x})d\mathbf{x}. \tag{2.2}$$

The feature of (2.1) is to provide a set of wavelet approximation coefficients, i.e., $\{\alpha_{\tau,\mathbf{k}};\ \mathbf{k}\in D_\tau\}$, and wavelet detail coefficients, i.e., $\{\beta_{j,\mathbf{k},u};\ j\ge\tau,\ \mathbf{k}\in D_j,\ u\in\{1,\dots,2^d-1\}\}$. For further details about wavelet bases, we refer to [32], [11] and [30].

2.2. Function space As usual in nonparametric estimation, we shall assume that f has a certain degree of smoothness. In this study, it is characterized by the set of functions $\mathcal{L}^s(M)$ defined by

$$\mathcal{L}^s(M) = \left\{ h\in\mathbb{L}_2([0,1]^d);\ (2.2) \text{ satisfies } \sum_{\mathbf{k}\in D_\tau}\alpha_{\tau,\mathbf{k}}^2 + \sup_{j\ge\tau} 2^{2js}\sum_{u=1}^{2^d-1}\sum_{\mathbf{k}\in D_j}\beta_{j,\mathbf{k},u}^2 \le M \right\},$$

where $s > 0$ and $M > 0$.

Under suitable assumptions on s, $\mathcal{L}^s(M)$ corresponds to the so-called Besov ball $B^s_{2,\infty}(M)$. It includes a wide variety of functions. A simple example in the case $d = 1$ is the following: let $h \in \mathbb{L}_2([0,1]^d$ such that its derivatives exist and are continuous up to order ℓ with $\ell \in \{0,\dots,N-1\}$, and there exists a constant $C > 0$ satisfying $|h^{(\ell)}(x) - h^{(\ell)}(y)| \le C|x-y|^\omega$, $(x,y)\in[0,1]^2$, $\omega\in(0,1)$. Then there exists a constant $C > 0$ such that $|\beta_{j,\mathbf{k},1}| \le C2^{-j(\omega+\ell+1/2)}$ for any $j\ge\tau$ and $\mathbf{k}\in D_j$. Hence $h\in\mathcal{L}^s(M)$ with $s = \omega+\ell$. Further details about such function spaces can be found in, e.g., [15], [32], [23] and [30].

3. Estimator and result

3.1. Estimator

Let τ be the integer mentioned in Section 2. Let us expand f as (2.1). Thanks to the orthonormality of the wavelet basis $\mathcal{S}$, we can express Q_f as

$$Q_f = \sum_{\mathbf{k}\in D_\tau}\alpha_{\tau,\mathbf{k}}^2 + \sum_{u=1}^{2^d-1}\sum_{j=\tau}^{\infty}\sum_{\mathbf{k}\in D_j}\beta_{j,\mathbf{k},u}^2, \tag{3.1}$$

where

$$\alpha_{j,\mathbf{k}} = \int_{[0,1]^d} f(\mathbf{x})\Phi_{j,\mathbf{k}}(\mathbf{x})d\mathbf{x}, \qquad \beta_{j,\mathbf{k},u} = \int_{[0,1]^d} f(\mathbf{x})\Psi_{j,\mathbf{k},u}(\mathbf{x})d\mathbf{x}.$$

In view of (3.1), using the plug-in approach, we consider the following wavelet-based estimator:

$$\hat{Q} = \sum_{\mathbf{k}\in D_\tau}\hat{\alpha}_{\tau,\mathbf{k}}^2 + \sum_{u=1}^{2^d-1}\sum_{j=\tau}^{j_*}\sum_{\mathbf{k}\in D_j}\hat{\beta}_{j,\mathbf{k},u}^2, \tag{3.2}$$

where $\hat{\alpha}_{\tau,k}$ and $\hat{\beta}_{j,k}$ denote two estimators of $\alpha_{\tau,k}$ and $\beta_{j,k}$ respectively and j_* denotes a positive integer.

We formulate the following assumption. From $Z_1, \ldots, Z_n$, we suppose that we are able to construct $\hat{\alpha}_{\tau,k}$ and $\hat{\beta}_{j,k}$ satisfying: for any integer $j \geq \tau$ and $\mathbf{k} \in D_j$, there exist a positive sequence $(w_n)_{n\in\mathbb{N}^*}$ with $\lim_{n\to\infty} w_n = 0$, a real number $\delta \geq 0$ and a constant $C > 0$ such that

$$\mathrm{E}\left((\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})^2\right) \leq C w_n, \qquad \mathrm{E}\left((\hat{\beta}_{j,\mathbf{k},u} - \beta_{j,\mathbf{k},u})^2\right) \leq C 2^{j\delta d} w_n. \tag{3.3}$$

We then consider the integer j_* satisfying

$$w_n^{-1/(2d(1+\delta))} < 2^{j_*+1} \leq 2 w_n^{-1/(2d(1+\delta))}. \tag{3.4}$$

Note that, contrary to the wavelet-based estimators constructed from a U-statistics (see, e.g., [27], [34] and [1]), $\hat{Q}$ is not an unbiased estimator of Q_f.

However,

• one can prove that, if $\hat{\alpha}_{\tau,\mathbf{k}}$ and $\hat{\beta}_{\tau,\mathbf{k},u}$ are unbiased estimators of $\alpha_{\tau,\mathbf{k}}$ and $\beta_{\tau,\mathbf{k},u}$ respectively, under (3.3) and (3.4), $\hat{Q}$ is asymptotically unbiased,

• the simplicity of its construction offers a certain flexibility on the nature of the considered model; if we are able to construct wavelet coefficient estimators satisfying (3.3) (whatever the dependence structure of the observations), assuming that f has a certain degree of smoothness, we can prove good asymptotic results for $\hat{Q}$ (see Theorem 1 below and the applications in Section 4).

3.2. Result

Theorem 1 below investigates the performances of $\hat{Q}$ under the MAE for $f \in \mathcal{L}^s(M)$.

THEOREM 1. *Let us consider the general nonparametric setting described in Section 1. Let Q_f be (1.1) and $\hat{Q}$ be (3.2) under (3.3) and (3.4). Suppose that $f \in \mathcal{L}^s(M)$ with $M > 0$ and $s > (1+\delta)d/2$. Then there exists a constant $C > 0$ such that*

$$E(|\hat{Q} - Q_f|) \leq C\sqrt{w_n}.$$

Theorem 1 shows that, under mild assumptions on the model, our estimator attains the rate of convergence $\sqrt{w_n}$ (which can be the optimal one in the minimax sense, see Remark 5).

REMARK 1. Since $\lim_{n\to\infty} w_n = 0$, Theorem 1 implies the consistency of $\hat{Q}$.

REMARK 2. The construction of $\hat{Q}$ does not depend on the smoothness parameter s of f; $\hat{Q}$ is adaptive.

REMARK 3. In our study we have supposed that the support of f satisfies $sup(f) = [0,1]^d$ only for the sake of simplicity in exposition. Theorem 1 can be extended for any compactly supported function f provided to an adaptation of the wavelet basis.

REMARK 4. In our multidimensional and general nonparametric framework, the construction of an adaptive estimator attaining the rate $\sqrt{w_n}$ for $f \in \mathcal{L}^s(M)$ with $M > 0$ and all $s > 0$ (without restriction as $s > (1+\delta)d/2$) raises new significant technical difficulties. This needs further investigations that we leave for a future work.

I what follows, we show examples of applications of Theorem 1 to three nonparametric problems: the biased density model, the nonparametric regression model and a GARCH-type model, under various dependent structures. The presented results are new in the considered frameworks.

4. Applications of Theorem 1

4.1. Biased density model

Model. Let d be a positive integer, $(Z_t)_{t\in\mathbb{Z}}$ be a strictly stationary random sequence defined on the probability space $([0,1]^d, \mathcal{B}([0,1]^d), \mathrm{P})$. The density of Z_1 is given by

$$g(\mathbf{x}) = \frac{w(\mathbf{x})f(\mathbf{x})}{\mu}, \qquad \mathbf{x} \in [0,1]^d,$$

where w denotes a known positive function and μ is the unknown normalization parameter: $\mu = \int_{[0,1]^d} w(\mathbf{x})f(\mathbf{x})d\mathbf{x}$. Our goal is to estimate the quadratic functional Q_f (1.1) from $Z_1, \ldots, Z_n$.

When $Z_1, \ldots, Z_n$ are independent and $d = 1$, this problem has been studied by [33]. Further details about the weighted density estimation problem can be found in, e.g., [18], [3] and the references therein.

The rest of study is devoted to the estimation of Q_f in the α-mixing case.

Definitions. For $j \in \mathbb{Z}$, define the σ-fields

$$\mathcal{F}^Z_{-\infty,j} = \sigma(Z_k, k \le j), \qquad \mathcal{F}^Z_{j,\infty} = \sigma(Z_k, k \ge j).$$

For any $m \in \mathbb{Z}$, we define the m-th α-mixing coefficient of $(Z_t)_{t\in\mathbb{Z}}$ by

$$\alpha_m = \sup_{(A,B)\in\mathcal{F}^Z_{-\infty,0}\times\mathcal{F}^Z_{m,\infty}} |\mathrm{P}(A\cap B) - \mathrm{P}(A)\mathrm{P}(B)|. \tag{4.1}$$

We say that $(Z_t)_{t\in\mathbb{Z}}$ is α-mixing if and only if $\lim_{m\to\infty} \alpha_m = 0$.

Full details on the α-mixing dependence can be found in, e.g., [35], [17], [8] and [19].

Assumptions. We formulate the following assumptions.

- There exist two constants $c > 0$ and $C > 0$ such that

$$c \le \inf_{\mathbf{x}\in[0,1]^d} w(\mathbf{x}), \qquad \sup_{\mathbf{x}\in[0,1]^d} w(\mathbf{x}) \le C. \tag{4.2}$$

- There exists a constant $C > 0$ such that

$$\sup_{\mathbf{x}\in[0,1]^d} f(\mathbf{x}) \le C. \tag{4.3}$$

- For any $m \in \{1, \ldots, n\}$, let $g_{(Z_0,Z_m)}$ be the density of (Z_0, Z_m). There exists a constant $C > 0$ such that

$$\sup_{m\in\{1,\ldots,n\}} \sup_{(\mathbf{x},\mathbf{y})\in[0,1]^{2d}} |g_{(Z_0,Z_m)}(\mathbf{x},\mathbf{y}) - g(\mathbf{x})g(\mathbf{y})| \le C. \tag{4.4}$$

- There exist two constants $C > 0$ and $q > 1$ such that the m-th α-mixing coefficient (4.1) of $(Z_t)_{t\in\mathbb{Z}}$ satisfies

$$\sum_{m=1}^{n} m^q \alpha_m^q \le C. \tag{4.5}$$

Result. Proposition 1 below explores the performances of $\hat{Q}$ (3.2) with a suitable choice of $\hat{\alpha}_{j,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},u}$ under the MAE for $f \in \mathcal{L}^s(M)$.

PROPOSITION 1. *Let us consider the biased density model framework described above under* (4.2), (4.3), (4.4) *and* (4.5). *Let* Q_f *be* (1.1), $\hat{Q}$ *be* (3.2) *with*

$$\hat{\alpha}_{\tau,\mathbf{k}} = \frac{\hat{\mu}}{n}\sum_{i=1}^{n} \frac{\Phi_{\tau,\mathbf{k}}(Z_i)}{w(Z_i)}, \qquad \hat{\beta}_{j,\mathbf{k},u} = \frac{\hat{\mu}}{n}\sum_{i=1}^{n} \frac{\Psi_{j,\mathbf{k},u}(Z_i)}{w(Z_i)}, \qquad \hat{\mu} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w(Z_i)}\right)^{-1} \tag{4.6}$$

and j_* such that

$$n^{1/(2d)} < 2^{j_*+1} \le 2n^{1/(2d)}.$$

Suppose that $f \in \mathcal{L}^s(M)$ with $M > 0$ and $s > d/2$. Then there exists a constant $C > 0$ such that

$$E(|\hat{Q} - Q_f|) \le C\frac{1}{\sqrt{n}}.$$

The proof of Proposition 1 is based on an adaptation of [9, Proposition 6.2] to the multidimensional case showing that the wavelet coefficients estimators (4.6) satisfy (3.3) with $w_n = 1/n$ and $\delta = 0$, and Theorem 1. For this reason, the details are omitted.

Remark 5. Let us mention that $1/\sqrt{n}$ is the optimal rate of convergence in the minimax sense for the standard density estimation problem (i.e. with $w(\mathbf{x}) = 1$) in the *i.i.d.* case and for $f \in \mathcal{L}^s(M)$ with $s > d/2$. See, e.g., [37, Section 2.7.4.].

4.2. Regression model

Model. Let d be a positive integer, $(Z_t)_{t\in\mathbb{Z}}$ be a strictly stationary bivariate random sequence defined on the probability space $(\mathbb{R}\times[0,1]^d, \mathcal{B}(\mathbb{R}\times[0,1]^d), \mathrm{P})$ where $Z_t = (Y_t, X_t)$,

$$Y_t = f(X_t) + \xi_t, \qquad t \in \mathbb{Z}, \tag{4.7}$$

$(X_t)_{t\in\mathbb{Z}}$ is a stationary random process with a known density $g : \mathbb{R}^d \to [0,\infty)$, $(\xi_t)_{t\in\mathbb{Z}}$ is a stationary random process with $\mathrm{E}(\xi_1) = 0$ and $\mathrm{E}(\xi_1^4) < \infty$, and $f : [0,1]^d \to \mathbb{R}$ is an unknown regression function. Moreover, it is understood that ξ_t is independent of X_t, for any $t \in \mathbb{Z}$. Our goal is to estimate the quadratic functional Q_f (1.1) from $Z_1, \ldots, Z_n$. We consider the α-mixing dependence. This kind of dependence is particularly interesting for nonparametric regression models thanks to its applications in dynamic economic systems and financial time series (see, e.g., [22], [40] and the references therein).

Assumptions. We formulate the following assumptions.

- There exists a constant $C > 0$ such that

$$\sup_{\mathbf{x}\in[0,1]^d} |f(\mathbf{x})| \le C. \tag{4.8}$$

- There exists a constant $c > 0$ such that

$$\inf_{\mathbf{x}\in[0,1]^d} g(\mathbf{x}) \ge c. \tag{4.9}$$

- There exist two constants $a > 0$ and $b > 0$ such that the m-th α-mixing coefficient (4.1) of $(Z_t)_{t\in\mathbb{Z}}$ satisfies

$$\alpha_m \le a e^{-bm}. \tag{4.10}$$

This corresponds to the so-called strong exponentially mixing case.

Result. Proposition 2 below investigates the performances of $\hat{Q}$ (3.2) with a suitable choice of $\hat{\alpha}_{j,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},u}$ under the MAE for $f \in \mathcal{L}^s(M)$.

Proposition 2. *Let us consider the regression model framework described above under (4.8), (4.9) and (4.10). Let Q_f be (1.1), $\hat{Q}$ be (3.2) with*

$$\hat{\alpha}_{\tau,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\Phi_{\tau,\mathbf{k}}(X_i), \qquad \hat{\beta}_{j,\mathbf{k},u} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\Psi_{j,\mathbf{k},u}(X_i) \tag{4.11}$$

and j_* such that

$$\left(\frac{n}{\ln n}\right)^{1/(2d)} < 2^{j_*+1} \leq 2\left(\frac{n}{\ln n}\right)^{1/(2d)}.$$

Suppose that $f \in \mathcal{L}^s(M)$ with $M > 0$ and $s > d/2$. Then there exists a constant $C > 0$ such that

$$E(|\hat{Q} - Q_f|) \leq C\sqrt{\frac{\ln n}{n}}.$$

Note that, in comparison to the corresponding optimal rate of convergence in the minimax sense for the *i.i.d.* case i.e. $1/\sqrt{n}$, we pay an extra logarithmic term. We explain this term by the mild assumptions made on our nonparametric regression model (remark that no "Castellana-Leadbetter-type condition" (as (4.4)) is done on $(Z_t)_{t\in\mathbb{Z}}$).

Remark 6. Other types of nonparametric regression models with dependent observations can be considered. For instance, one can considered (4.7) with $X_1, \ldots, X_n$ *i.i.d.* (or deterministic) and $(\xi_t)_{t\in\mathbb{Z}}$ a α-mixing process. In this setting, using similar arguments to [29], one can also apply Theorem 1.

4.3. GARCH model

Model. Let $(Z_t)_{t\in\mathbb{Z}}$ be a strictly stationary random sequence defined on the probability space $([0,1], \mathcal{B}([0,1]), \mathrm{P})$ where

$$Z_t = X_t\xi_t, \qquad t \in \mathbb{Z}, \tag{4.12}$$

$(\xi_t)_{t\in\mathbb{Z}}$ is a strictly stationary random sequence, the density of ξ_1 is known and is denotes by g, and $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary random sequence, the density of X_1 is unknown ans is denoted by f. Moreover, it is understood that ξ_t is independent of X_t, for any $t \in \mathbb{Z}$. Our goal is to estimate the quadratic functional Q_f (1.1) from $Z_1, \ldots, Z_n$. We focus our attention on the β-mixing dependence.

The model (4.12) belongs to the family of GARCH-type models. Financial applications related to (4.12) can be found in [8].

Definitions. For any $m \in \mathbb{Z}$, we define the m-th β-mixing coefficient of $(Z_t)_{t\in\mathbb{Z}}$ by

$$\beta_m = \frac{1}{2} \sup_{((A_i)_{i\in I}, (B_i)_{i\in J}) \in \mathcal{F}^Z_{-\infty,0} \times \mathcal{F}^Z_{m,\infty}} \sum_{i\in I}\sum_{j\in J} |\mathrm{P}(A_i \cap B_j) - \mathrm{P}(A_i)\mathrm{P}(B_j)|, \tag{4.13}$$

where the supremum is taken over all finite partitions $(A_i)_{i\in I}$ and $(B_j)_{j\in J}$ of Ω, which are respectively $\mathcal{F}^Z_{-\infty,0}$ and $\mathcal{F}^Z_{m,\infty}$ measurable.

We say that $(Z_t)_{t\in\mathbb{Z}}$ is β-mixing if and only if $\lim_{m\to\infty} \beta_m = 0$.

Full details can be found in, e.g., [17], [39] and [8].

Assumptions. We formulate the following assumptions.

- There exists an integer $\nu \geq 1$ such that, for any $i \in \{1, \ldots, n\}$,

$$\xi_i = \prod_{r=1}^{\nu} U_{r,i}, \tag{4.14}$$

where $U_{1,i}, \ldots, U_{\nu,i}$ are ν *i.i.d.* random variables with $U_{1,1} \sim \mathcal{U}([0,1])$.

- There exists a constant $C > 0$ such that

$$\sup_{x\in[0,1]} f(x) \leq C. \tag{4.15}$$

- There exists a constant $C > 0$ such that the m-th β-mixing coefficient (4.13) of $(Z_t)_{t\in\mathbb{Z}}$ satisfies

$$\sum_{m=1}^{n} \beta_m \leq C. \tag{4.16}$$

Result. Proposition 3 below evaluates the performances of $\hat{Q}$ (3.2) with a suitable choice of $\hat{\alpha}_{j,\mathbf{k}}$ and $\hat{\beta}_{j,\mathbf{k},u}$ under the MAE for $f \in \mathcal{L}^s(M)$.

PROPOSITION 3. *Let us consider the GARCH model framework described above under (4.14), (4.15) and (4.16). Let Q_f be (1.1) with $d=1$, for any integer $\ell \geq 1$ and any $h \in \mathcal{C}^\ell([0,1])$,*

$$T(h)(x) = (xh(x))', \qquad T_\ell(h)(x) = T(T_{\ell-1}(h))(x), \qquad x \in [0,1],$$

$\hat{Q}$ be (3.2) with $d=1$,

$$\hat{\alpha}_{\tau,\mathbf{k}} = \frac{1}{n}\sum_{i=1}^{n} T_\nu(\phi_{j,\mathbf{k}})(Z_i), \qquad \hat{\beta}_{j,\mathbf{k},1} = \frac{1}{n}\sum_{i=1}^{n} T_\nu(\psi_{j,\mathbf{k},1})(Z_i) \tag{4.17}$$

and j_* such that

$$n^{1/(2(1+2\nu))} < 2^{j_*+1} \leq 2n^{1/(2(1+2\nu))}.$$

Suppose that $f \in \mathcal{L}^s(M)$ with $M>0$ and $s > (1+2\nu)/2$. Then there exists a constant $C>0$ such that

$$E(|\hat{Q} - Q_f|) \leq C\frac{1}{\sqrt{n}}.$$

The proof of Proposition 3 is based on [10, Proposition 5.2] showing that the wavelet coefficients estimators (4.17) satisfy (3.3) with $w_n = 1/n$ and $\delta = 2\nu$, and Theorem 1. For this reason, the details are omitted.

5. A simulation study

In this section, we examine the finite-sample performance of the proposed wavelet estimator by a short simulation study in the context of Section 4.2.

5.1. The one dimensional case

We consider the nonparametric regression model

$$Y_i = f(X_i) + \xi_i, \qquad i \in \{1,\ldots,n\},$$

where $X_i = i/n$, $f:[0,1] \to \mathbb{R}$ is an unknown regression function and $(\xi_t)_{t\in\mathbb{Z}}$ is an AR(1)-process, i.e.,

$$\xi_t = \alpha\xi_{t-1} + \epsilon_t,$$

where $(\epsilon_t)_{t\in\mathbb{Z}}$ is a sequence of *i.i.d.* random variables drawn from a zero-mean normal distribution with variance σ_ϵ^2. Let us mention that $Y_1,\ldots,Y_n$ are dependent and $(\xi_t)_{t\in\mathbb{Z}}$ is strictly stationary and strongly mixing for $|\alpha|<1$ (see [17]). We aim to estimate Q_f (1.1) from $Y_1,\ldots,Y_n$.

Two regression functions ("Wave" and "Time Shifted Sine", initially introduced in [31]) were used (see Figure 1(a) and Figure 2(a)). They are defined by

1. Wave:

$$f_1(x) = 0.5 + 0.2\cos(4\pi x) + 0.1\cos(24\pi x).$$

2. Time Shifted Sine: first define the transformation $h(x) = (1-\cos(\pi x))/2$, then

$$f_2(x) = 0.3\sin\big(3\pi\big(h(h(h(h(x)))) + x\big)\big) + 0.5.$$

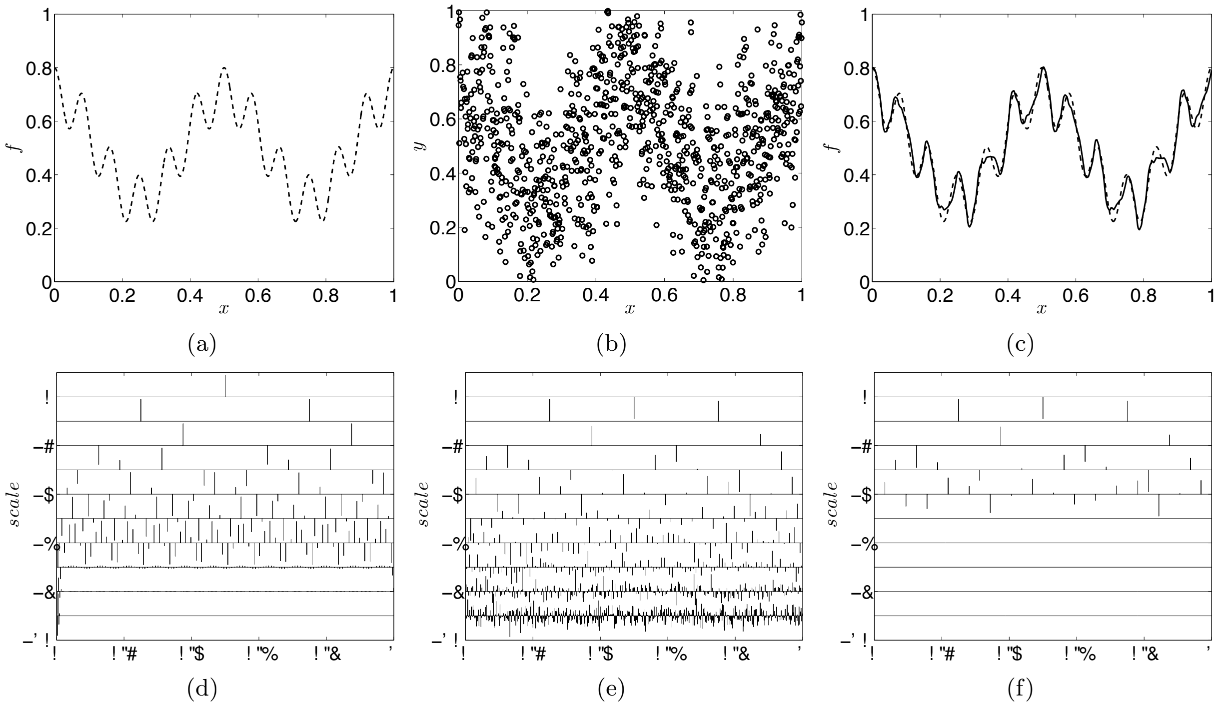

FIGURE 1. (a) Wave theoretical regression function f_1. (b) Noisy observations. (c) Typical reconstructions from 100 Monte-Carlo simulations with $n = 1024$ with the basic wavelet linear estimator (solid) and theoretical regression function f_1 (dashed). (d)–(f) Original/Noisy/Estimated wavelet coefficients from a single simulation.

The primary level $\tau = 0$ and the Symmlet wavelet with 6 vanishing moments were used throughout all experiments. All simulations were carried out using Matlab.

Figure 1(c) and Figure 2(c) show the results of the basic wavelet linear estimator from 100 replications of $n = 1024$ samples, with $\sigma_\epsilon = 0.2$ and $\alpha = 0.2$. Using the empirical wavelet coefficient estimators of f_u (see Figure 1(c) and Figure 2(c)) in $\hat{Q}_u$ (3.2) (estimator for Q_{f_u} (1.1)) for any $u \in \{1, 2\}$, we obtain

$$\hat{Q}_1 \approx 0.2749, \qquad \hat{Q}_2 \approx 0.2938, \qquad MAE(\hat{Q}_1) = 0.0105, \qquad MAE(\hat{Q}_2) = 0.0106.$$

Then, the MAE of our estimation procedure is analyzed with sample size 512, 1024 and 2048. Table 1 gives the MAE calculated by taking an average of the absolute errors based on 100 replications. Furthermore, we study the influence of the variance σ_ϵ (ranging from 0.04 to 1) of the noise and of the parameter α (ranging from 0.05 to 0.5) in the AR(1) process on the estimator. Table 1 shows that increasing the variance of the noise and/or α in the AR(1) process increases the MAE. Moreover, as expected, the MAE is decreasing as the sample size increases.

5.2. The two-dimensional case

We conclude the simulation results by a two-dimensional example. We consider the (two-dimensional) nonparametric regression model

$$Y_{i,j} = f(X_{1,i}, X_{2,j}) + \xi_{i,j}, \qquad (i,j) \in \{1, \ldots, n_*\}^2,$$

where $X_{1,i} = i/n_*$, $X_{2,j} = j/n_*$, $f : [0,1]^2 \to \mathbb{R}$ is an unknown regression function and $\xi_{i,j} = \xi_{1,i} + \xi_{2,j}$, $(\xi_{1,t})_{t\in\mathbb{Z}}$ and $(\xi_{2,t})_{t\in\mathbb{Z}}$ are two independent AR(1)-processes given by

$$\xi_{u,t} = \alpha_u \xi_{u,t-1} + \epsilon_{u,t}, \quad u \in \{1, 2\},$$

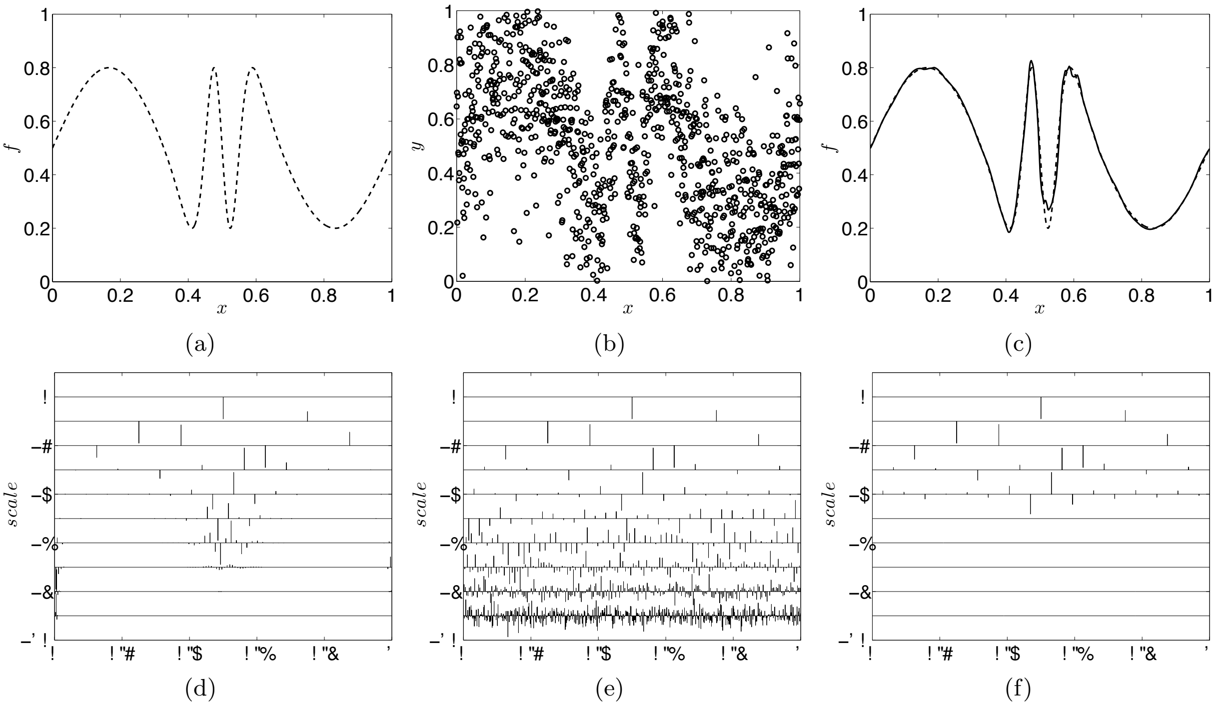

FIGURE 2. (a) Time Shifted Sine theoretical regression function f_2. (b) Noisy observations. (c) Typical reconstructions from 100 Monte-Carlo simulations with $n = 1024$ with the basic wavelet linear estimator (solid) and theoretical regression function f_2 (dashed). (d)–(f) Original/Noisy/Estimated wavelet coefficients from a single simulation.

TABLE 1. 100× mean MAE values from 100 replications of sample sizes 512, 1024 and 2048.

$\alpha = 0.05$

	$\sigma_\epsilon = 0.04$			$\sigma_\epsilon = 0.2$			$\sigma_\epsilon = 1$		
n	512	1024	2048	512	1024	2048	512	1024	2048
MAE($\hat{Q}_1$)	0.263	0.166	0.097	0.954	0.825	0.477	14.223	14.066	7.224
MAE($\hat{Q}_2$)	0.217	0.135	0.077	1.486	0.984	0.820	14.291	13.892	7.203

$\alpha = 0.2$

	$\sigma_\epsilon = 0.04$			$\sigma_\epsilon = 0.2$			$\sigma_\epsilon = 1$		
n	512	1024	2048	512	1024	2048	512	1024	2048
MAE($\hat{Q}_1$)	0.314	0.197	0.115	1.185	1.055	0.596	19.653	19.472	10.081
MAE($\hat{Q}_2$)	0.260	0.163	0.092	1.211	1.061	0.574	19.770	19.266	10.056

$\alpha = 0.5$

	$\sigma_\epsilon = 0.04$			$\sigma_\epsilon = 0.2$			$\sigma_\epsilon = 1$		
n	512	1024	2048	512	1024	2048	512	1024	2048
MAE($\hat{Q}_1$)	1.367	0.843	0.516	8.201	7.976	5.374	46.041	45.576	24.852
MAE($\hat{Q}_2$)	1.198	0.781	0.451	8.157	8.079	5.371	46.237	45.237	24.817

$(\epsilon_{1,t})_{t\in\mathbb{Z}}$ and $(\epsilon_{2,t})_{t\in\mathbb{Z}}$ are two sequences of *i.i.d.* random variables drawn from a zero-mean normal distribution with variance $\sigma^2_{\epsilon_1}$ and $\sigma^2_{\epsilon_2}$ respectively. We aim to estimate Q_f (1.1) from the $n = n_*^2$ random variables $Y_{1,1}, \ldots, Y_{n_*,n_*}$.

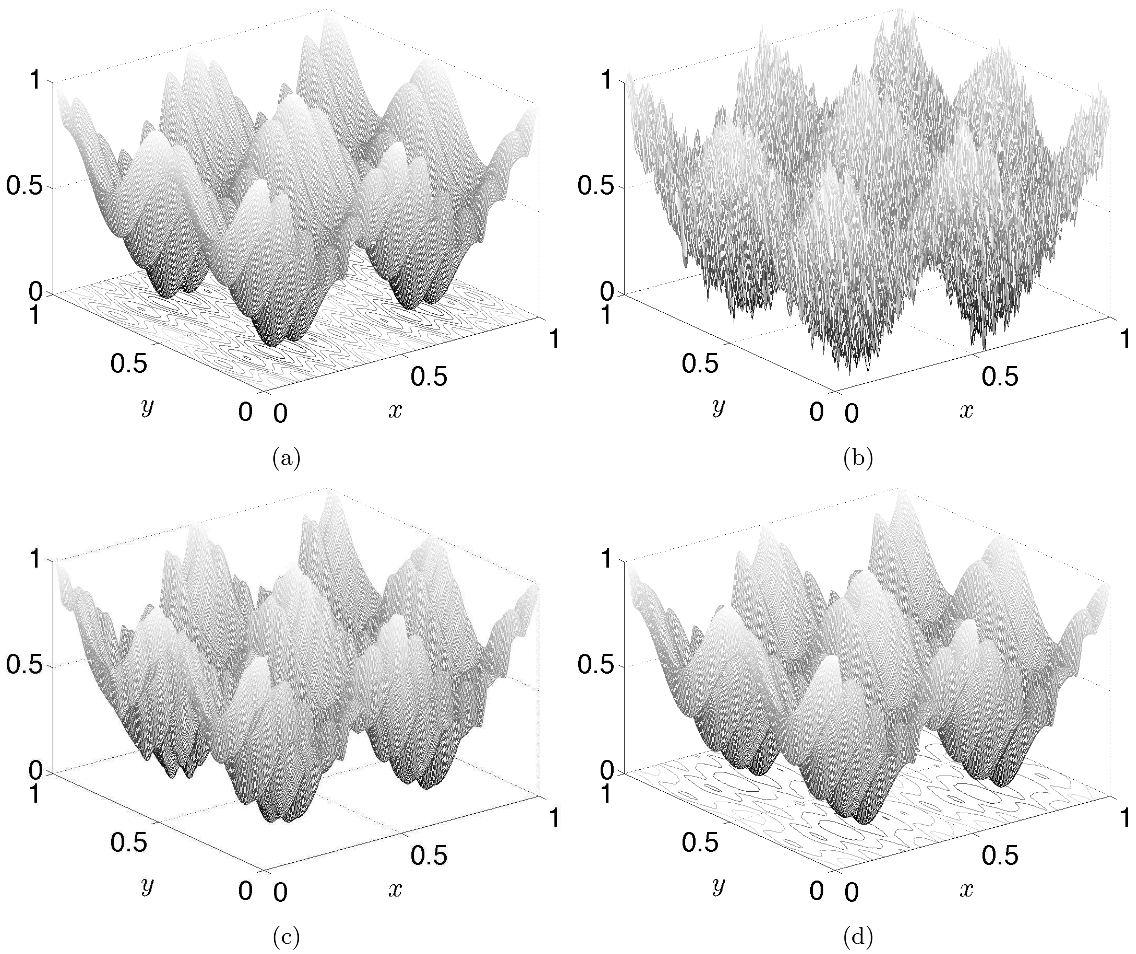

FIGURE 3. (a) Theoretical regression function f_1. (b) Noisy observations. Typical reconstructions (c) from a single simulation and (d) from 100 Monte-Carlo simulations with $n = 256^2$ with the basic wavelet linear estimator.

Two regression functions were used. They are defined by

1.
$$f_1(x, y) = 0.5 + 0.2\cos(4\pi x) + 0.1\cos(24\pi x) + 0.2\cos(4\pi y).$$

2. First define the transformation $h(x) = (1 - \cos(\pi x))/2$, then
$$f_2(x, y) = 0.3\sin\left(3\pi\left(h(h(h(h(x)))) + x\right)\right) + 0.1\cos(6\pi y) + 0.5.$$

Figure 3(d) and Figure 4(d) give an example of reconstruction with the basic wavelet linear estimator from 100 replications of $n = 256^2$ samples, with $\sigma_{\epsilon_1} = \sigma_{\epsilon_2} = 0.2$ and $\alpha_1 = \alpha_2 = 0.2$.

In Table 2 the MAE of the estimation procedure in the two-dimensional case is analyzed. As in the unidimensional case, it is obvious that simultaneously increasing the variances $\sigma^2_{\epsilon_1}$ and $\sigma^2_{\epsilon_2}$ of the noises of the two AR(1) processes increases the MAE and the MAE decreases as the sample size n increases. Moreover, we can see that increasing the two parameters α_1 and α_2 also increases the MAE but in a significantly lower fashion. However in association with very high level of noise (i.e., $\sigma_{\epsilon_1} = \sigma_{\epsilon_2} = 1$), the quadratic functional become rather difficult to estimate.

6. Proofs

In this section, C denotes any constant that does not depend on j, $\mathbf{k}$ and n. Its value may change

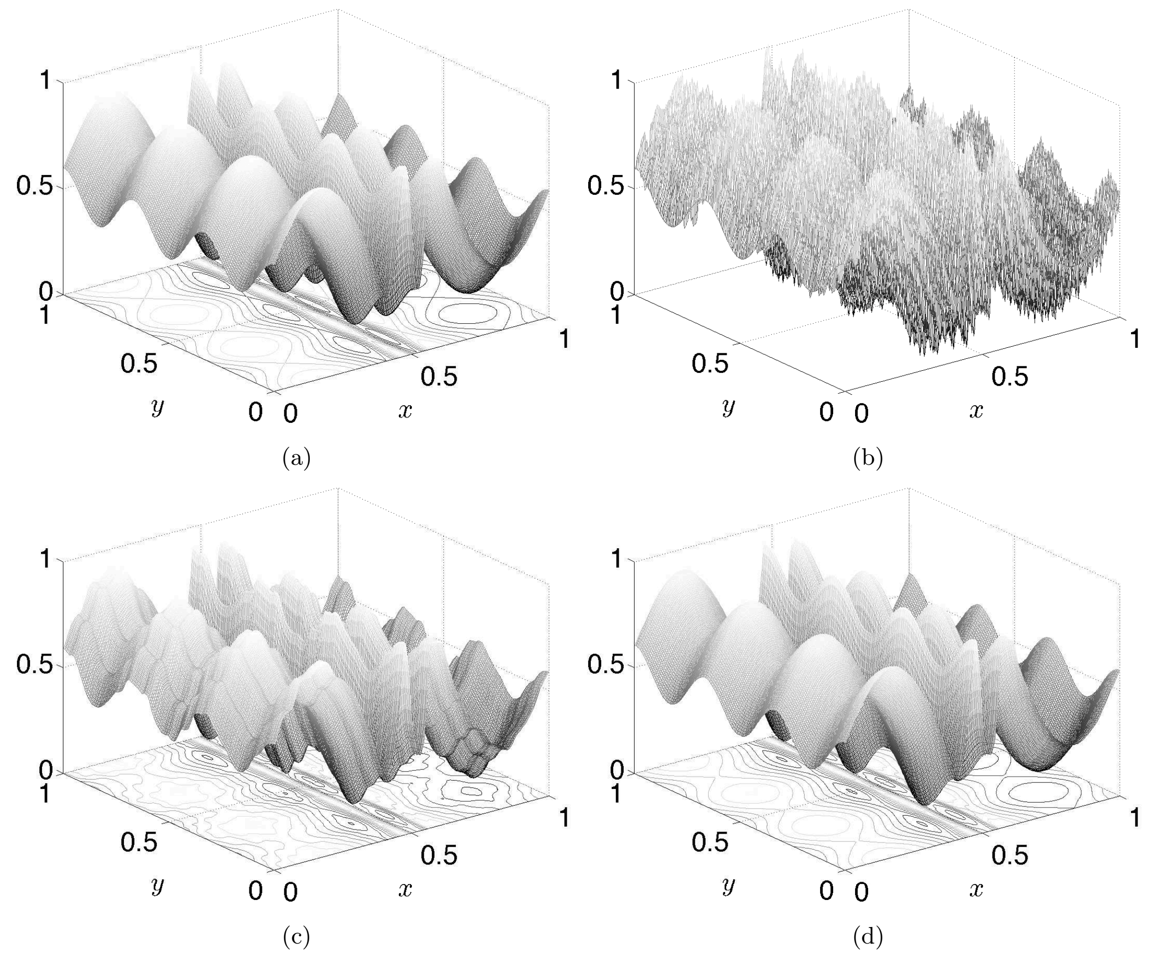

FIGURE 4. (a) Theoretical regression function f_2. (b) Noisy observations. Typical reconstructions (c) from a single simulation and (d) from 100 Monte-Carlo simulations with $n = 256^2$ with the basic wavelet linear estimator.

from one term to another and may depend on ϕ and ψ.

Proof of Theorem 1. It follows from (3.1), (3.2) and the triangular inequality that

$$\mathrm{E}(|\hat{Q} - Q_f|) \le A_1 + A_2 + A_3, \tag{6.1}$$

where

$$A_1 = \sum_{\mathbf{k}\in D_\tau} \mathrm{E}(|\hat{\alpha}^2_{\tau,\mathbf{k}} - \alpha^2_{\tau,\mathbf{k}}|), \qquad A_2 = \sum_{u=1}^{2^d-1} \sum_{j=\tau}^{j_*} \sum_{\mathbf{k}\in D_j} \mathrm{E}(|\hat{\beta}^2_{j,\mathbf{k},u} - \beta^2_{j,\mathbf{k},u}|)$$

and

$$A_3 = \sum_{u=1}^{2^d-1} \sum_{j=j_*+1}^{\infty} \sum_{\mathbf{k}\in D_j} \beta^2_{j,\mathbf{k},u}.$$

Let us now bound A_1, A_2 and A_3.
Upper bound for A_1. We have

$$\hat{\alpha}^2_{\tau,\mathbf{k}} - \alpha^2_{\tau,\mathbf{k}} = (\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})^2 + 2\alpha_{\tau,\mathbf{k}}(\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}}).$$

TABLE 2. 100× mean MAE values from 100 replications of sample sizes 128^2, 256^2 and 512^2.

$\alpha_1 = \alpha_2 = 0.05$

	$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=0.04$			$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=0.2$			$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=1$		
n	128^2	256^2	512^2	128^2	256^2	512^2	128^2	256^2	512^2
MAE($\hat{Q}_1$)	0.650	0.188	0.046	0.610	0.347	0.220	38.975	27.934	19.810
MAE($\hat{Q}_2$)	0.903	0.140	0.024	0.821	0.343	0.212	38.452	28.002	19.588

$\alpha_1 = \alpha_2 = 0.2$

	$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=0.04$			$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=0.2$			$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=1$		
n	128^2	256^2	512^2	128^2	256^2	512^2	128^2	256^2	512^2
MAE($\hat{Q}_1$)	0.650	0.187	0.046	0.647	0.407	0.261	52.782	38.704	27.476
MAE($\hat{Q}_2$)	0.903	0.139	0.024	0.830	0.406	0.254	52.243	38.777	27.204

$\alpha_1 = \alpha_2 = 0.5$

	$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=0.04$			$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=0.2$			$\sigma_{\epsilon_1}=\sigma_{\epsilon_2}=1$		
n	128^2	256^2	512^2	128^2	256^2	512^2	128^2	256^2	512^2
MAE($\hat{Q}_1$)	0.653	0.180	0.048	1.007	0.836	0.525	113.486	90.852	65.491
MAE($\hat{Q}_2$)	0.905	0.133	0.031	1.075	0.853	0.522	112.903	90.967	65.005

Owing to the triangular inequality, the Cauchy-Schwarz inequality and (3.3), we obtain

$$\begin{aligned} \mathrm{E}(|\hat{\alpha}^2_{\tau,\mathbf{k}} - \alpha^2_{\tau,\mathbf{k}}|) &\le \mathrm{E}((\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})^2) + 2|\alpha_{\tau,\mathbf{k}}|\sqrt{\mathrm{E}((\hat{\alpha}_{\tau,\mathbf{k}} - \alpha_{\tau,\mathbf{k}})^2)} \\ &\le C(w_n + \sqrt{w_n}) \le C\sqrt{w_n}. \end{aligned}$$

Therefore, since $Card(D_\tau)$ is constant,

$$A_1 \le C\sqrt{w_n}. \tag{6.2}$$

Upper bound for A_2. Again, we can write

$$\hat{\beta}^2_{j,\mathbf{k},u} - \beta^2_{j,\mathbf{k},u} = (\hat{\beta}_{j,\mathbf{k},u} - \beta_{j,\mathbf{k},u})^2 + 2\beta_{j,\mathbf{k},u}(\hat{\beta}_{j,\mathbf{k},u} - \beta_{j,\mathbf{k},u}).$$

The triangular inequality, the Cauchy-Schwarz inequality and (3.3) lead to

$$\begin{aligned} \mathrm{E}(|\hat{\beta}^2_{j,\mathbf{k},u} - \beta^2_{j,\mathbf{k},u}|) &\le \mathrm{E}((\hat{\beta}_{j,\mathbf{k},u} - \beta_{j,\mathbf{k},u})^2) + 2|\beta_{j,\mathbf{k},u}|\sqrt{\mathrm{E}((\hat{\beta}_{j,\mathbf{k},u} - \beta_{j,\mathbf{k},u})^2)} \\ &\le C(2^{j\delta d} w_n + |\beta_{j,\mathbf{k},u}| 2^{j\delta d/2}\sqrt{w_n}). \end{aligned}$$

Using the Cauchy-schwarz inequality, $Card(D_j) = 2^{jd}$, $f \in \mathcal{L}^s(M)$ with $s > (1+\delta)d/2$ and (3.4), we obtain

$$\begin{aligned} A_2 &\le C\left(w_n \sum_{j=\tau}^{j_*} 2^{j(1+\delta)d} + \sqrt{w_n}\sum_{u=1}^{2^d-1}\sum_{j=\tau}^{j_*} 2^{j\delta d/2}\sum_{\mathbf{k}\in D_j}|\beta_{j,\mathbf{k},u}|\right) \\ &\le C\left(w_n \sum_{j=\tau}^{j_*} 2^{j(1+\delta)d} + \sqrt{w_n}\sum_{j=\tau}^{j_*} 2^{j(1+\delta)d/2}\sqrt{\sum_{u=1}^{2^d-1}\sum_{\mathbf{k}\in D_j}\beta^2_{j,\mathbf{k},u}}\right) \\ &\le C\left(w_n 2^{j_*(1+\delta)d} + \sqrt{w_n}\sum_{j=\tau}^{\infty} 2^{-j(s-(1+\delta)d/2)}\right) \le C\sqrt{w_n}. \end{aligned} \tag{6.3}$$

Upper bound for A_3. The assumption $f \in \mathcal{L}^s(M)$ with $s > (1+\delta)d/2$ and (3.4) yield

$$A_3 \le C \sum_{j=j_*+1}^{\infty} 2^{-2js} \le C 2^{-2j_* s} \le C 2^{-j_*(1+\delta)d} \le C\sqrt{w_n}. \tag{6.4}$$

Putting (6.1), (6.2), (6.3) and (6.4) together, we obtain

$$\mathrm{E}(|\hat{Q} - Q_f|) \le C\sqrt{w_n}.$$

Theorem 1 is proved. □

Proof of Proposition 2. First of all, in order to apply Theorem 1, let us prove that the wavelet coefficient estimators (4.11) satisfy the assumption (3.3).

Observe that, thanks to the independence between ξ_1 and X_1 and $\mathrm{E}(\xi_1) = 0$, we have

$$\begin{aligned}\mathrm{E}(\hat{\beta}_{j,\mathbf{k},u}) &= \mathrm{E}\left(\frac{f(X_1)}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right) = \int_{[0,1]^d} \frac{f(\mathbf{x})}{g(\mathbf{x})}\Psi_{j,\mathbf{k},u}(\mathbf{x}) g(\mathbf{x}) d\mathbf{x} \\ &= \int_{[0,1]^d} f(\mathbf{x})\Psi_{j,\mathbf{k},u}(\mathbf{x}) d\mathbf{x} = \beta_{j,\mathbf{k},u}.\end{aligned}$$

Therefore, since $(Z_t)_{t\in\mathbb{Z}}$ is a stationary process, a standard covariance decomposition yields

$$\mathrm{E}\left((\hat{\beta}_{j,\mathbf{k},u} - \beta_{j,\mathbf{k},u})^2\right) = \frac{1}{n^2}\mathrm{V}\left(\sum_{i=1}^{n} \frac{Y_i}{g(X_i)}\Psi_{j,\mathbf{k},u}(X_i)\right) \le T_1 + T_2,$$

where

$$T_1 = \frac{1}{n}\mathrm{V}\left(\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)$$

and

$$T_2 = \frac{2}{n}\sum_{m=1}^{n-1}\left|\mathrm{Cov}\left(\frac{Y_{m+1}}{g(X_{m+1})}\Psi_{j,\mathbf{k},u}(X_{m+1}), \frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)\right|.$$

In order to bound T_1 and T_2, we will need the following moments result. Using again the independence between ξ_1 and X_1, $\mathrm{E}(\xi_1^4) < \infty$, (4.8), (4.9), applying the change of variables $\mathbf{y} = 2^j\mathbf{x} - \mathbf{k}$ and using the fact that Ψ is compactly supported, we have for any $\nu \in \{2,4\}$,

$$\begin{aligned}&\mathrm{E}\left(\left(\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)^{\nu}\right) \le C\left(\frac{C^{\nu} + \mathrm{E}(\xi_1^{\nu})}{c^{v-1}}\right)\mathrm{E}\left(\frac{1}{g(X_1)}\left(\Psi_{j,\mathbf{k},u}(X_1)\right)^{\nu}\right) \\ &= C\int_{[0,1]^d}\frac{1}{g(\mathbf{x})}(\Psi_{j,\mathbf{k},u}(\mathbf{x}))^{\nu} g(\mathbf{x}) d\mathbf{x} = C\int_{[0,1]^d}(\Psi_{j,\mathbf{k},u}(\mathbf{x}))^{\nu} d\mathbf{x} \\ &= C 2^{jd(\nu-2)/2}\int_{supp(\Psi)}(\Psi_u(\mathbf{x}))^{\nu} d\mathbf{x} \le C 2^{jd(\nu-2)/2}.\end{aligned} \tag{6.5}$$

It follows from (6.5) with $\nu = 2$ that

$$T_1 \le \frac{1}{n}\mathrm{E}\left(\left(\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)^2\right) \le C\frac{1}{n}.$$

Let us now study the upper bound for T_2. Let $[r\ln n]$ be the integer part of $r\ln n$ where $r = 1/b$. We have

$$T_2 = T_{2,1} + T_{2,2}, \tag{6.6}$$

where

$$T_{2,1}=\frac{2}{n}\sum_{m=1}^{[r\ln n]}\left|\mathrm{Cov}\left(\frac{Y_{m+1}}{g(X_{m+1})}\Psi_{j,\mathbf{k},u}(X_{m+1}),\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)\right|$$

and

$$T_{2,2}=\frac{2}{n}\sum_{m=[r\ln n]+1}^{n-1}\left|\mathrm{Cov}\left(\frac{Y_{m+1}}{g(X_{m+1})}\Psi_{j,\mathbf{k},u}(X_{m+1}),\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)\right|.$$

The Cauchy-Schwarz inequality and (6.5) with $\nu=2$ yield

$$\left|\mathrm{Cov}\left(\frac{Y_{m+1}}{g(X_{m+1})}\Psi_{j,\mathbf{k},u}(X_{m+1}),\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)\right|\le \mathrm{E}\left(\left(\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)^2\right)\le C.$$

Hence

$$T_{2,1}\le C\frac{\ln n}{n}.$$

By the Davydov inequality (see [13]), (4.10), again (6.5) with $\nu=4$ and $2^{jd}\le n$, we obtain

$$\begin{aligned}T_{2,2}&\le 10a^{1/2}\frac{1}{n}\sqrt{\mathrm{E}\left(\left(\frac{Y_1}{g(X_1)}\Psi_{j,\mathbf{k},u}(X_1)\right)^4\right)}\sum_{m=[r\ln n]+1}^{n-1}e^{-bm/2}\\&\le C\frac{1}{n}2^{jd/2}e^{-br\ln n/2}\le Cn^{-(1+br)/2}=C\frac{1}{n}.\end{aligned}$$

Hence

$$T_2\le C\frac{\ln n}{n}.$$

Combining the inequalities above, we obtain

$$\mathrm{E}\left((\hat{\beta}_{j,\mathbf{k},u}-\beta_{j,\mathbf{k},u})^2\right)\le C\frac{\ln n}{n}.$$

This inequality holds for $\hat{\alpha}_{j,\mathbf{k}}$ instead of $\hat{\beta}_{j,\mathbf{k},u}$ and $\alpha_{j,\mathbf{k}}$ instead of $\beta_{j,\mathbf{k},u}$. Therefore the assumption (3.3) is satisfied with $w_n=\ln n/n$ and $\delta=0$. Theorem 1 yields the desired result. □

Acknowledgments: We thank the referee for insightful comments that helped us improve the paper significantly.

References

[1] Barbedor, P. (2006). Analyse en composantes indépendantes par ondelettes, thèse Université Paris VII, *http://tel.archives-ouvertes.fr/tel-00119428.*

[2] Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya Serie A*, 50, 381-393.

[3] Brunel, E., Comte, F. and Guilloux, A. (2009). Nonparametric density estimation in presence of bias and censoring. *Test*, 18, 1, 166-194.

[4] Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications, *Bernoulli*, 15, 1, 69-98.

[5] Butucea, C. and Meziani, K. (2011). Quadratic functional estimation in inverse problems, *Statistical Methodology*, 8, 1, 31-41.

[6] Cai, T. and Low, M. (2005). Non-quadratic estimators of a quadratic functional. *The Annals of Statistics*, 33, 2930–2956.

[7] Cai, T. and Low, M. (2006). Optimal adaptive estimation of a quadratic functional. *The Annals of Statistics*, 34, 2298–2325.

[8] Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.

[9] Chesneau, C. (2011). Adaptive wavelet estimation of a biased density for strongly mixing sequences, *International Journal of Mathematics and Mathematical Sciences*, Volume 2011, Article ID 604150, 21 pages.

[10] Chesneau, C. and Doosti, H. (2012). Wavelet linear density estimation for a GARCH model under various dependence structures, *Journal of the Iranian Statistical Society*, 11, 1, 1-21.

[11] Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 24, 1, 54–81.

[12] Cox, D. (1969). Some sampling problems in technology. In New Developments in Survey Sampling (N. L. Johnson and H. Smith, Jr., eds.). Wiley, New York, 506-527.

[13] Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.

[14] Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample, *Journal of the Royal Statistical Society*, B, 64, 869-886.

[15] DeVore, R. and Popov, V. (1988). Interpolation of Besov spaces. *Trans. Amer. Math. Soc.*, 305, 397-414.

[16] Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Journal of Complexity*, 6, 290-323.

[17] Doukhan, P. (1994). Mixing. Properties and Examples. Lecture Notes in Statistics 85. Springer Verlag, New York.

[18] Efromovich, S. (2004). Density estimation for biased data. *Annals of Statistics*, 32, (3), 1137-1161.

[19] Fryzlewicz, P. and Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes, *Bernoulli*, Volume 17, Number 1 (2011), 320-346.

[20] Gayraud, G. and Tribouley, K. (1999). Wavelet methods to estimate an integrated quadratic functional: Adaptivity and asymptotic law. *Statistics and Probability Letters*, 44, 109-122.

[21] Giné, E. and Nickl, R. (2008). A Simple Adaptive Estimator of the Integrated Square of a Density. *Bernoulli*, 14, 47-61.

[22] Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press.

[23] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York 129, *Springer Verlag*.

[24] Hosseinioun, N., Doosti, H. and Niroumand, H.A. (2009). Wavelet-based estimators of the integrated squared density derivatives for mixing sequences. *Pakistan Journal of Statistics*, 25(3), 341-350.

[25] Johnstone, I. M. (2001a). Chi-square oracle inequalities. In *State of the art in probability and statistics* (Leiden, 1999), Lecture Notes - Monograph Series, volume 36, 399-418. Institute of Mathematical Statistics, Beachwood, OH.

[26] Johnstone, I. M. (2001b). Thresholding for weighted chi-square. *Statistica Sinica*, 11, 691-704.

[27] Kerkyacharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using haar wavelets. *The Annals of Statistics*, 24 (2), 485-507.

[28] Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM PS*, 9, 1-18.

[29] Liang, H. (2011). Asymptotic normality of wavelet estimator in heteroscedastic model with α-mixing error. *J Syst Sci Complex*, 24, (4), 725-737.

[30] Mallat, S. (2009). A wavelet tour of signal processing. Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.

[31] Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H. and Patil, P. (1998). Exact risk analysis of wavelet regression. *J. Comput. Graph. Statist*, 7, 278-309.

[32] Meyer, Y. (1992). Wavelets and Operators. *Cambridge University Press*, Cambridge.

[33] Petsa, A. and Sapatinas, T. (2010). Adaptive quadratic functional estimation of a weighted density by model selection. *Statistics*, 44, 571-585.

[34] Prakasa Rao. B.L.S. (1999). Estimation of the integrated squared density derivatives by wavelets. *Bull. Inform. Cyb.*, 31, 47-65.

[35] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Ac. Sc. USA*, 42, 43-47.

[36] Shao, Q.-M. (1995). Maximal inequality for partial sums of ρ-mixing sequences. *Ann. Probab.*, 23, 948-965.

[37] Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*, Springer.

[38] Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* John Wiley & Sons, Inc., New York, 384 pp.

[39] Viennet, G. (1997). Inequalities for absolutely regular processes: application to density estimation. *Probab. Theory Related Fields*, 107, 467-492.

[40] White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52, 143-162.

1 306010 100266 >