



HAL
open science

A note on the adaptive estimation of a quadratic functional from dependent observations

Christophe Chesneau, Maher Kachour

► **To cite this version:**

Christophe Chesneau, Maher Kachour. A note on the adaptive estimation of a quadratic functional from dependent observations. 2012. hal-00707015v1

HAL Id: hal-00707015

<https://hal.science/hal-00707015v1>

Preprint submitted on 11 Jun 2012 (v1), last revised 9 Oct 2012 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on the adaptive estimation of a quadratic functional from dependent observations

CHRISTOPHE CHESNEAU¹ AND MAHER KACHOUR²

¹*LMNO, Université de Caen Basse-Normandie, 14032, France*

²*LMJL, Université de Nantes, 44322, France.*

June 11, 2012

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 conditions.

Key words and phrases: Quadratic functional estimation, Plug-in approach, Wavelets, Rates of convergence, Mixing dependence.

AMS 2000 Subject Classifications: 62G07, 62G20.

1 Introduction

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

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

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

When $d = 1$ and Z_1, \dots, Z_n are independent, this problem has been addressed in many papers for a wide variety of models under various settings. See e.g. Bickel and Ritov (1988),

Donoho and Nussbaum (1990), Kerkyacharian and Picard (1996), Gayraud and Tribouley (1999), Prakasa Rao (1999), Johnstone (2001a,b), Delaigle and Gijbels (2002), Laurent (2005), Cai and Low (2005, 2006), Giné and Nickl (2008) and Petsa and Sapatinas (2010). The multidimensional case has been considered by Barbedor (2006) for the density model. When $d = 1$ and Z_1, \dots, Z_n are dependent, the estimation of Q_f has been investigated by Hosseinioun *et al.* (2009) for the density model and by Butucea and Comte (2009) 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, including a wide variety of models, with possible dependent Z_1, \dots, 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 Härdle *et al.* (1998) and Vidakovic (1999) for a detailed discussion of 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, \dots, 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 conditions: 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.

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. Finally, the proofs are postponed to Section 5.

2 Preliminaries on wavelets

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.

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

- A scale function:

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

- $2^d - 1$ wavelet functions:

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

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

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

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

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

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

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

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

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

For any integer j_* such that $j_* \geq \tau$, a function $h \in \mathbb{L}_2([0, 1]^d)$ can be expressed via \mathcal{S} as wavelet series as

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

where

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

For further details about wavelet bases, we refer to Meyer (1992), Cohen *et al.* (1993) and Mallat (2009).

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

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

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

Under appropriate assumptions on s , $\mathcal{L}^s(M)$ corresponds to the so-called Besov ball $B_{2, \infty}^s(M)$. Details about such function spaces can be found in e.g. Devore and Popov (1988), Meyer (1992) and Härdle *et al.* (1998).

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

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

where

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

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

$$\hat{Q} = \sum_{\mathbf{k} \in D_\tau} \hat{\alpha}_{\tau, \mathbf{k}}^2 + \sum_{u=1}^{2^d-1} \sum_{j=\tau}^{j_0} \sum_{\mathbf{k} \in D_j} \hat{\beta}_{j, \mathbf{k}, u}^2, \quad (3.2)$$

where $\hat{\alpha}_{\tau, \mathbf{k}}$ and $\hat{\beta}_{j, \mathbf{k}}$ denote two estimators of $\alpha_{\tau, \mathbf{k}}$ and $\beta_{j, \mathbf{k}}$ respectively and j_0 denotes a positive integer.

From Z_1, \dots, Z_n , we suppose that

- we are able to construct $\hat{\alpha}_{\tau, \mathbf{k}}$ and $\hat{\beta}_{j, \mathbf{k}}$ 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 $\lim_{n \rightarrow \infty} w_n = 0$, a real number $\delta \geq 0$ and a constant $C > 0$ such that

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

- j_0 is the integer satisfying

$$w_n^{-1/(2d(1+\delta))} < 2^{j_0+1} \leq 2w_n^{-1/(2d(1+\delta))}, \quad (3.4)$$

(where w_n refers to the one of the previous point).

Note that, contrary to the wavelet-based estimators constructed from a U -statistics (see e.g. Kerkyacharian and Picard (1996), Prakasa Rao (1999) and Barbedor (2006)), \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 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 3.1 below and the applications in Section 4).

3.2 Result

Theorem 3.1 below investigates the performances of \hat{Q} under the MAE for $f \in \mathcal{L}^s(M)$.

Theorem 3.1 *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*

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

Theorem 3.1 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 4.1).

Remark 3.1 *Since $\lim_{n \rightarrow \infty} w_n = 0$, Theorem 3.1 implies the consistency of \hat{Q} .*

Remark 3.2 *The construction of \hat{Q} does not depend on the smoothness parameter s ; \hat{Q} is adaptive.*

Remark 3.3 *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.*

The following section shows examples of applications of Theorem 3.1 to three non-parametric problems: the biased density model, the nonparametric regression model and a GARCH-type model, under various dependent structures.

4 Applications of Theorem 3.1

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 $([0, 1]^d, \mathcal{B}([0, 1]^d), \mathbb{P})$. The density of Z_1 is given by

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

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, \dots, Z_n .

When Z_1, \dots, Z_n are independent and $d = 1$, this problem has been studied by Petsa and Sapatinas (2010). Further details about the weighted density estimation problem can be found in e.g. Efromovich (2004), Brunel *et al.* (2009) 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

$$\mathcal{F}_{-\infty, j}^Z = \sigma(Z_k, k \leq j), \quad \mathcal{F}_{j, \infty}^Z = \sigma(Z_k, k \geq j).$$

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

$$\alpha_m = \sup_{(A, B) \in \mathcal{F}_{-\infty, 0}^Z \times \mathcal{F}_{m, \infty}^Z} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|. \quad (4.1)$$

Full details on the α -mixing dependence can be found in e.g. Rosenblatt (1956), Doukhan (1994), Carrasco and Chen (2002) and Fryzlewicz and Subba Rao (2011).

Assumptions. We formulate the following assumptions.

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

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

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

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

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

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

- 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

$$\sum_{m=1}^n m^q \alpha_m^q \leq C. \quad (4.5)$$

Result. Proposition 4.1 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 4.1 *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*

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

and j_0 such that

$$n^{1/(2d)} < 2^{j_0+1} \leq 2n^{1/(2d)}.$$

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

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

The proof of Proposition 4.1 is omitted; it is based on an adaptation of (Chesneau, 2011, 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 3.1.

Remark 4.1 *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. (Tsybakov, 2004, 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 $(\mathbb{R} \times [0, 1]^d, \mathcal{B}(\mathbb{R} \times [0, 1]^d), \mathbb{P})$ where $Z_t = (Y_t, X_t)$,

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

$(X_t)_{t \in \mathbb{Z}}$ is a stationary random process with $X_1 \sim \mathcal{U}([0, 1]^d)$, $(\xi_t)_{t \in \mathbb{Z}}$ is a stationary random process with $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}(\xi_1^4) < \infty$, and $f : [0, 1]^d \rightarrow \mathbb{R}$ 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, \dots, Z_n . We consider the α -mixing dependence.

Further details about the nonparametric regression estimation problem can be found in e.g. White and Domowitz (1984), Tsybakov (2004) and the references therein.

Assumptions. We formulate the following assumptions.

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

$$\sup_{\mathbf{x} \in [0,1]^d} f(\mathbf{x}) \leq C. \quad (4.7)$$

- 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

$$\alpha_m \leq ab^{-m}, \quad (4.8)$$

Result. Proposition 4.2 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 4.2 *Let us consider the regression model framework described above under (4.7) and (4.8). Let Q_f be (1.1), \hat{Q} be (3.2) with*

$$\hat{\alpha}_{\tau,\mathbf{k}} = \frac{1}{n} \sum_{i=1}^n Y_i \Phi_{\tau,\mathbf{k}}(X_i), \quad \hat{\beta}_{j,\mathbf{k},u} = \frac{1}{n} \sum_{i=1}^n Y_i \Psi_{j,\mathbf{k},u}(X_i) \quad (4.9)$$

and j_0 such that

$$\left(\frac{n}{\ln n} \right)^{1/(2d)} < 2^{j_0+1} \leq 2 \left(\frac{n}{\ln n} \right)^{1/(2d)}.$$

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

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

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 $(Z_t)_{t \in \mathbb{Z}}$).

4.3 GARCH model

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

$$Z_t = X_t \xi_t, \quad t \in \mathbb{Z}, \quad (4.10)$$

$(\xi_t)_{t \in \mathbb{Z}}$ is a strictly stationary random sequence, the density of ξ_1 is known and is denoted by g , and $(X_t)_{t \in \mathbb{Z}}$ is a strictly stationary random sequence, the density of X_1 is unknown and 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, \dots, Z_n . We focus our attention on the β -mixing dependence.

The model (4.10) belongs to the family of GARCH-type models. Financial applications related to (4.10) can be found in Carrasco and Chen (2002).

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

$$\beta_m = \frac{1}{2} \sup_{((A_i)_{i \in I}, (B_j)_{j \in J}) \in \mathcal{F}_{-\infty, 0}^Z \times \mathcal{F}_{m, \infty}^Z} \sum_{i \in I} \sum_{j \in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|, \quad (4.11)$$

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 $\mathcal{F}_{-\infty, 0}^Z$ and $\mathcal{F}_{m, \infty}^Z$ measurable.

Full details can be found in e.g. Doukhan (1994), Viennet (1997) and Carrasco and Chen (2002).

Assumptions. We formulate the following assumptions.

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

$$\xi_i = \prod_{r=1}^{\nu} U_{r,i}, \quad (4.12)$$

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

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

$$\sup_{x \in [0, 1]} f(x) \leq C. \quad (4.13)$$

- There exists a constant $C > 0$ such that the m -th β -mixing coefficient (4.11) of $(Z_t)_{t \in \mathbb{Z}}$ satisfies

$$\sum_{m=1}^n \beta_m \leq C. \quad (4.14)$$

Result. Proposition 4.3 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 4.3 *Let us consider the GARCH model framework described above under (4.12), (4.13) and (4.14). Let Q_f be (1.1) with $d = 1$, for any integer $\ell \geq 1$ and any $h \in \mathcal{C}^\ell([0, 1])$,*

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

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

$$\hat{\alpha}_{\tau,\mathbf{k}} = \frac{1}{n} \sum_{i=1}^n T_\nu(\phi_{j,\mathbf{k}})(Z_i), \quad \hat{\beta}_{j,\mathbf{k},1} = \frac{1}{n} \sum_{i=1}^n T_\nu(\psi_{j,\mathbf{k},1})(Z_i). \quad (4.15)$$

and j_0 such that

$$n^{1/(2(1+2\nu))} < 2^{j_0+1} \leq 2n^{1/(2(1+2\nu))}.$$

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

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

The proof of Proposition 4.3 is omitted; it is based on (Chesneau and Doosti, 2012, Proposition 5.2) showing that the wavelet coefficients estimators (4.15) satisfy (3.3) with $w_n = 1/n$ and $\delta = 2\nu$, and Theorem 3.1.

5 Proofs

In this section, C denotes any constant that does not depend on j , \mathbf{k} and n . Its value may change from one term to another and may depend on ϕ and ψ .

Proof of Theorem 3.1. The triangular inequality yields

$$\mathbb{E}(|\hat{Q} - Q_f|) \leq A_1 + A_2 + A_3, \quad (5.1)$$

where

$$A_1 = \sum_{\mathbf{k} \in D_\tau} \mathbb{E}(|\hat{\alpha}_{\tau,\mathbf{k}}^2 - \alpha_{\tau,\mathbf{k}}^2|), \quad A_2 = \sum_{u=1}^{2^d-1} \sum_{j=\tau}^{j_0} \sum_{\mathbf{k} \in D_j} \mathbb{E}(|\hat{\beta}_{j,\mathbf{k},u}^2 - \beta_{j,\mathbf{k},u}^2|)$$

and

$$A_3 = \sum_{u=1}^{2^d-1} \sum_{j=j_0+1}^{\infty} \sum_{\mathbf{k} \in D_j} \beta_{j,\mathbf{k},u}^2.$$

Let us bound A_1 , A_2 and A_3 .

Upper bound for A_1 . We have

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

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

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

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

$$A_1 \leq C\sqrt{w_n}. \quad (5.2)$$

Upper bound for A_2 . Again, we can write

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

It follows from the triangular inequality, the Cauchy-Schwarz inequality and (3.3) that

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

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

$$\begin{aligned} A_2 &\leq C \left(w_n \sum_{j=\tau}^{j_0} 2^{j(1+\delta)d} + \sqrt{w_n} \sum_{u=1}^{2^d-1} \sum_{j=\tau}^{j_0} 2^{j\delta d/2} \sum_{\mathbf{k} \in D_j} |\beta_{j,\mathbf{k},u}| \right) \\ &\leq C \left(w_n \sum_{j=\tau}^{j_0} 2^{j(1+\delta)d} + \sqrt{w_n} \sum_{j=\tau}^{j_0} 2^{j(1+\delta)d/2} \sqrt{\sum_{u=1}^{2^d-1} \sum_{\mathbf{k} \in D_j} \beta_{j,\mathbf{k},u}^2} \right) \\ &\leq C \left(w_n 2^{j_0(1+\delta)d} + \sqrt{w_n} \sum_{j=\tau}^{j_0} 2^{-j(s-(1+\delta)d/2)} \right) \leq C\sqrt{w_n}. \end{aligned} \quad (5.3)$$

Upper bound for A_2 . The assumption $f \in \mathcal{L}^s(M)$ with $s > (1 + \delta)d/2$ and (3.4) yield

$$A_3 \leq C \sum_{j=j_0+1}^{\infty} 2^{-2js} \leq C2^{-2j_0s} \leq C2^{-j_0(1+\delta)d} \leq C\sqrt{w_n}. \quad (5.4)$$

Putting (5.1), (5.2), (5.3) and (5.4) together, we obtain

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

Theorem 3.1 is proved. \square

Proof of Proposition 4.2. First of all, in order to apply Theorem 3.1, let us prove that the wavelet coefficient estimators (4.9) satisfy the assumption (3.3).

Observe that, thanks to the independence between ξ_1 and X_1 and $X_1 \sim \mathcal{U}([0, 1]^d)$, $\mathbb{E}(\hat{\beta}_{j,\mathbf{k},u}) = \beta_{j,\mathbf{k},u}$. Since $(Z_t)_{t \in \mathbb{Z}}$ is a stationary process, a standard covariance decomposition yields

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

where

$$T_1 = \frac{1}{n} \mathbb{V}(Y_1 \Psi_{j,\mathbf{k},u}(X_1)), \quad T_2 = \frac{2}{n} \sum_{m=1}^{n-1} |\text{Cov}(Y_{m+1} \Psi_{j,\mathbf{k},u}(X_{m+1}), Y_1 \Psi_{j,\mathbf{k},u}(X_1))|.$$

In order to bound T_1 and T_2 , we shall need the following moments result. Using again the independence between ξ_1 and X_1 , $\mathbb{E}(\xi_1^4) < \infty$, $X_1 \sim \mathcal{U}([0, 1]^d)$, (4.7), 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 $\nu \in \{2, 4\}$,

$$\begin{aligned} \mathbb{E}((Y_1 \Psi_{j,\mathbf{k},u}(X_1))^\nu) &\leq C(C_1^\nu + \mathbb{E}(\xi_1^\nu)) \mathbb{E}((\Psi_{j,\mathbf{k},u}(X_1))^\nu) = C \int_{[0,1]^d} (\Psi_{j,\mathbf{k},u}(\mathbf{x}))^\nu d\mathbf{x} \\ &= C 2^{jd(\nu-2)/2} \int_{[0,1]^d} (\Psi(\mathbf{x}))^\nu d\mathbf{x} \leq C 2^{jd(\nu-2)/2}. \end{aligned} \quad (5.5)$$

It follows from (5.5) with $\nu = 2$ that

$$T_1 \leq \frac{1}{n} \mathbb{E}((Y_1 \Psi_{j,\mathbf{k},u}(X_1))^2) \leq C \frac{1}{n}.$$

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

$$T_2 = T_{2,1} + T_{2,2}, \quad (5.6)$$

where

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

and

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

The Cauchy-Schwarz inequality and (5.5) with $\nu = 2$ yield

$$|\text{Cov}(Y_{m+1} \Psi_{j,\mathbf{k},u}(X_{m+1}), Y_1 \Psi_{j,\mathbf{k},u}(X_1))| \leq \mathbb{E}((Y_1 \Psi_{j,\mathbf{k},u}(X_1))^2) \leq C.$$

Hence

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

By the Davydov inequality (see Davydov (1970)), (4.8), again (5.5) with $\nu = 4$ and $2^{jd} \leq n$, we obtain

$$\begin{aligned} T_{2,2} &\leq C \frac{1}{n} \sqrt{\mathbb{E}((Y_1 \Psi_{j,\mathbf{k},u}(X_1))^4)} \sum_{m=[c \ln n]+1}^{n-1} \sqrt{\alpha_m} \\ &\leq C \frac{1}{n} 2^{jd/2} \sum_{m=[c \ln n]+1}^{\infty} b^{-m/2} \leq C \frac{1}{n} \sqrt{nb}^{-c \ln n/2} \leq C \frac{1}{n}. \end{aligned}$$

Hence

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

Combining the inequalities above, we obtain

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

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 3.1 yields the desired result. \square

References

- Barbedor, P. (2006). Analyse en composantes indépendantes par ondelettes, thèse Université Paris VII, <http://tel.archives-ouvertes.fr/tel-00119428>.
- Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya Serie A*, 50, 381-393.
- Brunel, E., Comte, F. and Guilloux, A. (2009). Nonparametric density estimation in presence of bias and censoring. *Test*, 18, 1, 166-194.

- Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications, *Bernoulli*, 15, 1, 69-98.
- Cai, T. and Low, M. (2005). Non-quadratic estimators of a quadratic functional. *The Annals of Statistics*, 33, 2930–2956.
- Cai, T. and Low, M. (2006). Optimal adaptive estimation of a quadratic functional. *The Annals of Statistics*, 34, 2298–2325.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.
- 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.
- 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.
- 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.
- 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.
- Davydov, Y. (1970). The invariance principle for stationary processes. *Theor. Probab. Appl.*, 15, 3, 498-509.
- 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.
- DeVore, R. and Popov, V. (1988). Interpolation of Besov spaces. *Trans. Amer. Math. Soc.*, 305, 397-414.
- Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Journal of Complexity*, 6, 290-323.
- Doukhan, P. (1994). *Mixing. Properties and Examples*. Lecture Notes in Statistics 85. Springer Verlag, New York.
- Efromovich, S. (2004). Density estimation for biased data. *Annals of Statistics*, 32, (3), 1137-1161.

- Fryzlewicz, P. and Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes, *Bernoulli*, Volume 17, Number 1 (2011), 320-346.
- 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.
- Giné, E. and Nickl, R. (2008). A Simple Adaptive Estimator of the Integrated Square of a Density. *Bernoulli*, 14, 47-61.
- Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York 129, Springer Verlag.
- 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.
- 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.
- Johnstone, I. M. (2001b). Thresholding for weighted chi-square. *Statistica Sinica*, 11, 691-704.
- Kerkycharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using haar wavelets. *The Annals of Statistics*, 24 (2), 485-507.
- Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM PS*, 9, 1-18.
- Mallat, S. (2009). A wavelet tour of signal processing. Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.
- Meyer, Y. (1992). Wavelets and Operators. *Cambridge University Press*, Cambridge.
- Petsa, A. and Sapatinas, T. (2010). Adaptive quadratic functional estimation of a weighted density by model selection. *Statistics*, 44, 571-585.
- Prakasa Rao. B.L.S. (1999). Estimation of the integrated squared density derivatives by wavelets. *Bull. Inform. Cyb.*, 31, 47-65.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Ac. Sc. USA*, 42, 43-47.

- Shao, Q.-M. (1995). Maximal inequality for partial sums of ρ -mixing sequences. *Ann. Probab.*, 23, 948-965.
- Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*, Springer.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc., New York, 384 pp.
- Viennet, G. (1997). Inequalities for absolutely regular processes: application to density estimation. *Probab. Theory Related Fields*, 107, 467-492.
- White, H. and Domowitz, I. (1984). Nonlinear Regression with Dependent Observations. *Econometrica*, 52, 143-162.