Research Article

A Note on the Adaptive Estimation of a Multiplicative Separable Regression Function

Christophe Chesneau

Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France

Correspondence should be addressed to Christophe Chesneau; christophe.chesneau@gmail.com

Received 18 January 2014; Accepted 25 February 2014; Published 20 March 2014

Academic Editors: F. Ding, E. Skubalska-Rafajlowicz, and H. C. So

Copyright © 2014 Christophe Chesneau. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigate the estimation of a multiplicative separable regression function from a bidimensional nonparametric regression model with random design. We present a general estimator for this problem and study its mean integrated squared error (MISE) properties. A wavelet version of this estimator is developed. In some situations, we prove that it attains the standard unidimensional rate of convergence under the MISE over Besov balls.

1. Motivations

We consider the bidimensional nonparametric regression model with random design described as follows. Let $(Y_i, U_i, V_i)_{i \in \mathbb{Z}}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where

$$Y_i = h(U_i, V_i) + \xi_i, \quad i \in \mathbb{Z}, \quad (1)$$

$(\xi_i)_{i \in \mathbb{Z}}$ is a strictly stationary stochastic process, $(U_i, V_i)_{i \in \mathbb{Z}}$ is a strictly stationary stochastic process with support in $[0, 1]^2$, and $h : [0, 1]^2 \rightarrow \mathbb{R}$ is an unknown bivariate regression function. It is assumed that $\mathbb{E}(\xi_1) = 0$, $\mathbb{E}(\xi_1^2)$ exists, $(U_i, V_i)_{i \in \mathbb{Z}}$ are independent, $(\xi_i)_{i \in \mathbb{Z}}$ are independent, and, for any $i \in \mathbb{Z}$, (U_i, V_i) and ξ_i are independent. In this study, we focus our attention on the case where h is a multiplicative separable regression function: there exist two functions $f : [0, 1] \rightarrow \mathbb{R}$ and $g : [0, 1] \rightarrow \mathbb{R}$ such that

$$h(x, y) = f(x)g(y). \quad (2)$$

We aim to estimate h from the n random variables: $(Y_1, U_1, V_1), \dots, (Y_n, U_n, V_n)$. This problem is plausible in many practical situations as in utility, production, and cost function applications (see, e.g., Linton and Nielsen [1],

Yatchew and Bos [2], Pinske [3], Lewbel and Linton [4], and Jacho-Chávez [5]).

In this note, we provide a theoretical contribution to the subject by introducing a new general estimation method for h . A sharp upper bound for its mean integrated squared error (MISE) is proved. Then we adapt our methodology to propose an efficient and adaptive procedure. It is based on two wavelet thresholding estimators following the construction studied in Chaubey et al. [6]. It has the features to be adaptive for a wide class of unknown functions and enjoy nice MISE properties. Further details on wavelet estimators can be found in, for example, Antoniadis [7], Vidakovic [8], and Härdle et al. [9]. Despite the so-called “curse of dimensionality” coming from the bidimensionality of (1), we prove that our wavelet estimator attains the standard unidimensional rate of convergence under the MISE over Besov balls (for both the homogeneous and inhomogeneous zones). It completes asymptotic results proved by Linton and Nielsen [1] via nonadaptive kernel methods for the structured nonparametric regression model.

The paper is organized as follows. Assumptions on (1) and some notations are introduced in Section 2. Section 3 presents our general MISE result. Section 4 is devoted to our wavelet estimator and its performances in terms of rate

of convergence under the MISE over Besov balls. Technical proofs are collected in Section 5.

2. Assumptions and Notations

For any $p \geq 1$, we set

$$\mathbb{L}^p([0, 1]) = \left\{ v : [0, 1] \longrightarrow \mathbb{R}; \|v\|_p = \left(\int_0^1 |v(x)|^p dx \right)^{1/p} < \infty \right\}. \quad (3)$$

We set

$$e_o = \int_0^1 f(x) dx, \quad e_* = \int_0^1 g(x) dx, \quad (4)$$

provided that they exist.

We formulate the following assumptions.

(H1) There exists a known constant $C_1 > 0$ such that

$$\sup_{x \in [0, 1]} |f(x)| \leq C_1. \quad (5)$$

(H2) There exists a known constant $C_2 > 0$ such that

$$\sup_{x \in [0, 1]} |g(x)| \leq C_2. \quad (6)$$

(H3) The density of (U_1, V_1) , denoted by q , is known and there exists a constant $c_3 > 0$ such that

$$c_3 \leq \inf_{(x, y) \in [0, 1]^2} q(x, y). \quad (7)$$

(H4) There exists a known constant $\omega > 0$ such that

$$|e_o e_*| \geq \omega. \quad (8)$$

The assumptions (H1) and (H2), involving the boundedness of h , are standard in nonparametric regression models. The knowledge of q discussed in (H3) is restrictive but plausible in some situations, the most common case being $(U_1, V_1) \sim \mathcal{U}([0, 1]^2)$ (the uniform distribution on $[0, 1]^2$). Finally, mention that (H4) is just a technical assumption more realistic to the knowledge of e_o and e_* (depending on f and g , resp.).

3. MISE Result

Theorem 1 presents an estimator for h and shows an upper bound for its MISE.

Theorem 1. One considers (1) under (H1)–(H4). One introduces the following estimator for h (2):

$$\hat{h}(x, y) = \frac{\tilde{f}(x) \tilde{g}(y)}{\tilde{e}} \mathbf{1}_{\{|\tilde{e}| \geq \omega/2\}}, \quad (9)$$

where $\tilde{f}$ denotes an arbitrary estimator for f_{e_*} in $\mathbb{L}^2([0, 1])$, $\tilde{g}$ denotes an arbitrary estimator for g_{e_o} in $\mathbb{L}^2([0, 1])$, $\mathbf{1}$ denotes the indicator function,

$$\tilde{e} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{q(U_i, V_i)}, \quad (10)$$

and ω refers to (H4).

Then there exists a constant $C > 0$ such that

$$\begin{aligned} & \mathbb{E} \left(\iint_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\mathbb{E} (\|\tilde{g} - g_{e_o}\|_2^2) + \mathbb{E} (\|\tilde{f} - f_{e_*}\|_2^2) \right. \\ & \quad \left. + \mathbb{E} (\|\tilde{g} - g_{e_o}\|_2^2 \|\tilde{f} - f_{e_*}\|_2^2) + \frac{1}{n} \right). \end{aligned} \quad (11)$$

The form of $\tilde{h}$ (9) is derived to the multiplicative separable structure of h (2) and a ratio-type normalization. Other results about such ratio-type estimators in a general statistical context can be found in Vasiliev [10].

Based on Theorem 1, $\hat{h}$ is efficient for h if and only if $\tilde{f}$ is efficient for f_{e_*} and $\tilde{g}$ is efficient for g_{e_o} in terms of MISE. Even if several methods are possible, we focus our attention on wavelet methods enjoying adaptivity for a wide class of unknown functions and having optimal properties under the MISE. For details on the interests of wavelet methods in nonparametric statistics, we refer to Antoniadis [7], Vidakovic [8], and Härdle et al. [9].

4. Adaptive Wavelet Estimation

Before introducing our wavelet estimators, let us present some basics on wavelets.

4.1. Wavelet Basis on $[0, 1]$. Let us briefly recall the construction of wavelet basis on the interval $[0, 1]$ introduced by Cohen et al. [11]. Let N be a positive integer, and let ϕ and ψ be the initial wavelets of the Daubechies orthogonal wavelets $db2N$. We set

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k). \quad (12)$$

With appropriate treatments at the boundaries, there exists an integer τ satisfying $2^\tau \geq 2N$ such that the collection $\mathcal{S} = \{\phi_{\tau,k}(\cdot), k \in \{0, \dots, 2^\tau - 1\}; \psi_{j,k}(\cdot); j \in \mathbb{N} - \{0, \dots, \tau - 1\}, k \in \{0, \dots, 2^j - 1\}\}$, is an orthonormal basis of $\mathbb{L}^2([0, 1])$.

Any $v \in \mathbb{L}^2([0, 1])$ can be expanded on $\mathcal{S}$ as

$$v(x) = \sum_{k=0}^{2^\tau-1} \alpha_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(x), \quad x \in [0, 1], \quad (13)$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the wavelet coefficients of v defined by

$$\alpha_{j,k} = \int_0^1 v(x) \phi_{j,k}(x) dx, \quad \beta_{j,k} = \int_0^1 v(x) \psi_{j,k}(x) dx. \quad (14)$$

4.2. Besov Balls. For the sake of simplicity, we consider the sequential version of Besov balls defined as follows. Let $M > 0$, $s \in (0, N)$, $p \geq 1$ and $r \geq 1$. A function v belongs to $B_{p,r}^s(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (14) satisfy

$$2^{\tau(1/2-1/p)} \left(\sum_{k=0}^{2^{\tau}-1} |\alpha_{\tau,k}|^p \right)^{1/p} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*. \quad (15)$$

In this expression, s is a smoothness parameter and p and r are norm parameters. For a particular choice of s , p , and r , $B_{p,r}^s(M)$ contains the Hölder and Sobolev balls (see, e.g., DeVore and Popov [12], Meyer [13], and Härdle et al. [9]).

4.3. Hard Thresholding Estimators. In the sequel, we consider (1) under (H1)–(H4).

We consider hard thresholding wavelet estimators for $\tilde{f}$ and $\tilde{g}$ in (9). They are based on a term-by-term selection of estimators of the wavelet coefficients of the unknown function. Those which are greater to a threshold are kept; the others are removed. This selection is the key to the adaptivity and the good performances of the hard thresholding wavelet estimators (see, e.g., Donoho et al. [14], Delyon and Juditsky [15], and Härdle et al. [9]).

To be more specific, we use the “double thresholding” wavelet technique, introduced by Delyon and Juditsky [15] then recently improved by Chaubey et al. [6]. The role of the second thresholding (appearing in the definition of the wavelet estimator for $\beta_{j,k}$) is to relax assumption on the model (see Remark 6).

Estimator $\tilde{f}$ for fe_* . We define the hard thresholding wavelet estimator $\tilde{f}$ by

$$\tilde{f}(x) = \sum_{k=0}^{2^{\tau}-1} \hat{\alpha}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa C_* \lambda_n\}} \psi_{j,k}(x), \quad (16)$$

where

$$\hat{\alpha}_{\tau,k} = \frac{1}{a_n} \sum_{i=1}^{a_n} \frac{Y_i}{q(U_i, V_i)} \phi_{\tau,k}(U_i), \quad (17)$$

where a_n is the integer part of $n/2$,

$$\hat{\beta}_{j,k} = \frac{1}{a_n} \sum_{i=1}^{a_n} W_{i,j,k} \mathbf{1}_{\{|W_{i,j,k}| \leq C_* / \lambda_n\}}, \quad (18)$$

$$W_{i,j,k} = \frac{Y_i}{q(U_i, V_i)} \psi_{j,k}(U_i),$$

where j_1 is the integer satisfying $(1/2)a_n < 2^{j_1} \leq a_n$, $\kappa = 2 + 8/3 + 2\sqrt{4 + 16/9}$, $C_* = \sqrt{(2/c_3)(C_1^2 C_2^2 + \mathbb{E}(\xi_1^2))}$, and

$$\lambda_n = \sqrt{\frac{\ln a_n}{a_n}}. \quad (19)$$

Estimator $\tilde{g}$ for ge_o . We define the hard thresholding wavelet estimator $\tilde{g}$ by

$$\tilde{g}(x) = \sum_{k=0}^{2^{\tau}-1} \hat{v}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \hat{\theta}_{j,k} \mathbf{1}_{\{|\hat{\theta}_{j,k}| \geq \kappa_* C_* \eta_n\}} \psi_{j,k}(x), \quad (20)$$

where

$$\hat{v}_{\tau,k} = \frac{1}{b_n} \sum_{i=1}^{b_n} \frac{Y_{a_n+i}}{q(U_{a_n+i}, V_{a_n+i})} \phi_{\tau,k}(V_{a_n+i}), \quad (21)$$

where a_n is the integer part of $n/2$, $b_n = n - a_n$,

$$\hat{\theta}_{j,k} = \frac{1}{b_n} \sum_{i=1}^{b_n} Z_{a_n+i,j,k} \mathbf{1}_{\{|Z_{a_n+i,j,k}| \leq C_* / \eta_n\}}, \quad (22)$$

$$Z_{a_n+i,j,k} = \frac{Y_{a_n+i}}{q(U_{a_n+i}, V_{a_n+i})} \psi_{j,k}(V_{a_n+i}),$$

Where j_2 is the integer satisfying $(1/2)b_n < 2^{j_2} \leq b_n$, $\kappa_* = 2 + 8/3 + 2\sqrt{4 + 16/9}$, $C_* = \sqrt{(2/c_3)(C_1^2 C_2^2 + \mathbb{E}(\xi_1^2))}$, and

$$\eta_n = \sqrt{\frac{\ln b_n}{b_n}}. \quad (23)$$

Estimator for h . From $\tilde{f}$ (16) and $\tilde{g}$ (20), we consider the following estimator for h (2):

$$\hat{h}(x, y) = \frac{\tilde{f}(x) \tilde{g}(y)}{\tilde{e}} \mathbf{1}_{\{|\tilde{e}| \geq \omega/2\}}, \quad (24)$$

where

$$\tilde{e} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{q(U_i, V_i)} \quad (25)$$

and ω refers to (H4).

Let us mention that $\hat{h}$ is adaptive in the sense that it does not depend on f or g in its construction.

Remark 2. Since $\tilde{f}$ is defined with $(Y_1, U_1, V_1), \dots, (Y_{a_n}, U_{a_n}, V_{a_n})$ and $\tilde{g}$ is defined with $(Y_{a_n+1}, U_{a_n+1}, V_{a_n+1}), \dots, (Y_n, U_n, V_n)$, thanks to the independence of $(Y_1, U_1, V_1), \dots, (Y_n, U_n, V_n)$, $\tilde{f}$ and $\tilde{g}$ are independent.

Remark 3. The calibration of the parameters in $\tilde{f}$ and $\tilde{g}$ is based on theoretical considerations; thus defined, $\tilde{f}$ and $\tilde{g}$ can attain a fast rate of convergence under the MISE over Besov balls (see [6], Theorem 6.1). Further details are given in the proof of Theorem 4.

4.4. Rate of Convergence. Theorem 4 investigates the rate of convergence attains by $\hat{h}$ under the MISE over Besov balls.

Theorem 4. We consider (1) under (H1)–(H4). Let $\hat{h}$ be (24) and let h be (2). Suppose that

- (i) $f \in B_{p_1, r_1}^{s_1}(M_1)$ with $M_1 > 0$, $r_1 \geq 1$, either $\{p_1 \geq 2$ and $s_1 \in (0, N)\}$ or $\{p_1 \in [1, 2)$ and $s_1 \in (1/p_1, N)\}$,
- (ii) $g \in B_{p_2, r_2}^{s_2}(M_2)$ with $M_2 > 0$, $r_2 \geq 1$, either $\{p_2 \geq 2$ and $s_2 \in (0, N)\}$ or $\{p_2 \in [1, 2)$ and $s_2 \in (1/p_2, N)\}$.

Then there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\iint_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \leq C \left(\frac{\ln n}{n} \right)^{2s_*/(2s_*+1)}, \quad (26)$$

where $s_* = \min(s_1, s_2)$.

The rate of convergence $(\ln n/n)^{2s_*/(2s_*+1)}$ is the near optimal one in the minimax sense for the unidimensional regression model with random design under the MISE over Besov balls $B_{p, r}^{s, *}(M)$ (see, e.g., Tsybakov [16], and Härdle et al. [9]). Thus Theorem 4 proves that our estimator escapes to the so-called “curse of dimensionality.” Such a result is not possible with the standard bidimensional hard thresholding wavelet estimator attaining the rate of convergence $(\ln n/n)^{2s/(2s+d)}$ with $d = 2$ under the MISE over bidimensional Besov balls defined with s as smoothness parameter (see Delyon and Juditsky [15]).

Theorem 4 completes asymptotic results proved by Linton and Nielsen [1] investigating this problem for the structured nonparametric regression model via another estimation method based on nonadaptive kernels.

Remark 5. In Theorem 4, we take into account both the homogeneous zone of Besov balls, that is, $\{p_1 \geq 2$ and $s_1 \in (0, N)\}$, and the inhomogeneous zone, that is, $\{p_1 \in [1, 2)$ and $s_1 \in (1/p_1, N)\}$, for the case $f \in B_{p_1, r_1}^{s_1}(M_1)$ and the same for $g \in B_{p_2, r_2}^{s_2}(M_2)$. This has the advantage to cover a very rich class of unknown regression functions h .

Remark 6. Note that Theorem 4 does not require the knowledge of the distribution of ξ_1 ; $\{\mathbb{E}(\xi_1) = 0$ and the existence of $\mathbb{E}(\xi_1^2)\}$ is enough.

Remark 7. Let us mention that the phenomenon of curse of dimensionality has also been studied via wavelet methods by Neumann [17] but for the multidimensional Gaussian white noise model and with different approaches based on anisotropic frameworks.

Remark 8. Our study can be extended to the multidimensional case considered by Yatchew and Bos [2], that is, $f : [0, 1]^{q_1} \rightarrow \mathbb{R}$ and $g : [0, 1]^{q_2} \rightarrow \mathbb{R}$; q_1 and q_2 denoting two positive integers. In this case, adapting our framework to the multidimensional case (q_1 dimensional Besov balls, q_1 dimensional (tensorial) wavelet basis, q_1 dimensional hard thresholding wavelet estimator, ... see, e.g., Delyon and

Juditsky [15]), one can prove that (9) attains the rate of convergence $(\ln n/n)^{2s_*/(2s_*+q_*)}$, where $s_* = \min(s_1, s_2)$ and $q_* = \max(q_1, q_2)$.

5. Proofs

In this section, for the sake of simplicity, C denotes a generic constant; its value may change from one term to another.

Proof of Theorem 1. Observe that

$$\begin{aligned} \hat{h}(x, y) - h(x, y) &= \frac{\tilde{f}(x) \tilde{g}(y)}{\tilde{e}} \mathbf{1}_{\{|\tilde{e}| \geq \omega/2\}} - f(x) g(y) \\ &= \frac{1}{\tilde{e}} (\tilde{f}(x) \tilde{g}(y) - f(x) g(y) \tilde{e}) \mathbf{1}_{\{|\tilde{e}| \geq \omega/2\}} \\ &\quad - f(x) g(y) \mathbf{1}_{\{|\tilde{e}| < \omega/2\}}. \end{aligned} \quad (27)$$

Therefore, using the triangular inequality, the Markov inequality, (H1), (H2), (H4), $\{|\tilde{e}| < \omega/2\} \cap \{|e_* e_o| \geq \omega\} \subseteq \{|\tilde{e} - e_* e_o| \geq \omega/2\}$, and again the Markov inequality, we get

$$\begin{aligned} |\hat{h}(x, y) - h(x, y)| &\leq \frac{2}{\omega} |\tilde{f}(x) \tilde{g}(y) - f(x) g(y) \tilde{e}| + |f(x)| |g(y)| \mathbf{1}_{\{|\tilde{e}| < \omega/2\}} \\ &\leq C (|\tilde{f}(x) \tilde{g}(y) - f(x) g(y) \tilde{e}| + \mathbf{1}_{\{|\tilde{e} - e_* e_o| \geq \omega/2\}}) \\ &\leq C (|\tilde{f}(x) \tilde{g}(y) - f(x) g(y) \tilde{e}| + |\tilde{e} - e_* e_o|). \end{aligned} \quad (28)$$

On the other hand, we have the decomposition

$$\begin{aligned} \tilde{f}(x) \tilde{g}(y) - f(x) g(y) \tilde{e} &= f(x) e_* (\tilde{g}(y) - g(y) e_o) + g(y) e_o (\tilde{f}(x) - f(x) e_*) \\ &\quad + (\tilde{g}(y) - g(y) e_o) (\tilde{f}(x) - f(x) e_*) \\ &\quad + f(x) g(y) (e_* e_o - \tilde{e}). \end{aligned} \quad (29)$$

Owing to the triangular inequality, (H1) and (H2), we have

$$\begin{aligned} |\tilde{f}(x) \tilde{g}(y) - f(x) g(y) \tilde{e}| &\leq C (|\tilde{g}(y) - g(y) e_o| + |\tilde{f}(x) - f(x) e_*| \\ &\quad + |\tilde{g}(y) - g(y) e_o| |\tilde{f}(x) - f(x) e_*| + |\tilde{e} - e_* e_o|). \end{aligned} \quad (30)$$

Putting (28) and (30) together, we obtain

$$\begin{aligned} |\hat{h}(x, y) - h(x, y)| &\leq C (|\tilde{g}(y) - g(y) e_o| + |\tilde{f}(x) - f(x) e_*| \\ &\quad + |\tilde{g}(y) - g(y) e_o| |\tilde{f}(x) - f(x) e_*| + |\tilde{e} - e_* e_o|). \end{aligned} \quad (31)$$

Therefore, by the elementary inequality: $(a + b + c + d)^2 \leq 8(a^2 + b^2 + c^2 + d^2)$, $(a, b, c, d) \in \mathbb{R}^4$, an integration over $[0, 1]^2$ and taking the expectation, it comes

$$\begin{aligned} & \mathbb{E} \left(\iint_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\mathbb{E} (\|\tilde{g} - ge_o\|_2^2) + \mathbb{E} (\|\tilde{f} - fe_*\|_2^2) \right. \\ & \quad \left. + \mathbb{E} (\|\tilde{g} - ge_o\|_2^2 \|\tilde{f} - fe_*\|_2^2) + \mathbb{E} ((\tilde{e} - e_* e_o)^2) \right). \end{aligned} \quad (32)$$

Now observe that, owing to the independence of $(U_i, V_i)_{i \in \mathbb{Z}}$, the independence between (U_1, V_1) and ξ_1 , and $\mathbb{E}(\xi_1) = 0$, we obtain

$$\begin{aligned} \mathbb{E}(\tilde{e}) &= \mathbb{E} \left(\frac{Y_1}{q(U_1, V_1)} \right) \\ &= \mathbb{E} \left(\frac{h(U_1, V_1)}{q(U_1, V_1)} \right) + \mathbb{E}(\xi_1) \mathbb{E} \left(\frac{1}{q(U_1, V_1)} \right) \\ &= \iint_0^1 \frac{f(x)g(y)}{q(x, y)} q(x, y) dx dy \\ &= \left(\int_0^1 f(x) dx \right) \left(\int_0^1 g(y) dy \right) = e_* e_o. \end{aligned} \quad (33)$$

Then, using similar arguments to (33), $(a + b)^2 \leq 2(a^2 + b^2)$, $(a, b) \in \mathbb{R}^2$, (H1), (H2), (H3), and $\mathbb{E}(\xi_1^2) < \infty$, we have

$$\begin{aligned} \mathbb{E}((\tilde{e} - e_* e_o)^2) &= \mathbb{V}(\tilde{e}) \\ &= \frac{1}{n} \mathbb{V} \left(\frac{Y_1}{q(U_1, V_1)} \right) \\ &\leq \frac{1}{n} \mathbb{E} \left(\left(\frac{Y_1}{q(U_1, V_1)} \right)^2 \right) \\ &\leq \frac{2}{n} \mathbb{E} \left(\frac{(h(U_1, V_1))^2 + \xi_1^2}{(q(U_1, V_1))^2} \right) \\ &\leq \frac{2}{c_3^2} (C_1^2 C_2^2 + \mathbb{E}(\xi_1^2)) \frac{1}{n} = C \frac{1}{n}. \end{aligned} \quad (34)$$

Equations (32) and (34) yield the desired inequality:

$$\begin{aligned} & \mathbb{E} \left(\iint_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ & \leq C \left(\mathbb{E} (\|\tilde{g} - ge_o\|_2^2) + \mathbb{E} (\|\tilde{f} - fe_*\|_2^2) \right. \\ & \quad \left. + \mathbb{E} (\|\tilde{g} - ge_o\|_2^2 \|\tilde{f} - fe_*\|_2^2) + \frac{1}{n} \right). \end{aligned} \quad (35)$$

□

Proof of Theorem 4. We aim to apply Theorem 1 by investigating the rate of convergence attained by $\tilde{f}$ and $\tilde{g}$ under the MISE over Besov balls.

First of all, remark that, for $\gamma \in \{\phi, \psi\}$, any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$.

(i) Using similar arguments to (33), we obtain

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{a_n} \sum_{i=1}^{a_n} \frac{Y_i}{q(U_i, V_i)} \gamma_{j,k}(U_i) \right) \\ &= \mathbb{E} \left(\frac{Y_1}{q(U_1, V_1)} \gamma_{j,k}(U_1) \right) \\ &= \mathbb{E} \left(\frac{h(U_1, V_1)}{q(U_1, V_1)} \gamma_{j,k}(U_1) \right) + \mathbb{E}(\xi_1) \mathbb{E} \left(\frac{\gamma_{j,k}(U_1)}{q(U_1, V_1)} \right) \\ &= \iint_0^1 \frac{f(x)g(y)}{q(x, y)} \gamma_{j,k}(x) q(x, y) dx dy \\ &= \left(\int_0^1 f(x) \gamma_{j,k}(x) dx \right) \left(\int_0^1 g(y) dy \right) \\ &= \int_0^1 (f(x) e_*) \gamma_{j,k}(x) dx. \end{aligned} \quad (36)$$

(ii) Using similar arguments to (34) and $\|\gamma_{j,k}\|_2^2 = 1$, we have

$$\begin{aligned} & \sum_{i=1}^{a_n} \mathbb{E} \left(\left(\frac{Y_i}{q(U_i, V_i)} \gamma_{j,k}(U_i) \right)^2 \right) \\ &= \mathbb{E} \left(\left(\frac{Y_1}{q(U_1, V_1)} \gamma_{j,k}(U_1) \right)^2 \right) a_n \\ &\leq 2 \mathbb{E} \left(\frac{(h(U_1, V_1))^2 + \xi_1^2}{(q(U_1, V_1))^2} (\gamma_{j,k}(U_1))^2 \right) a_n \\ &\leq \frac{2}{c_3} (C_1^2 C_2^2 + \mathbb{E}(\xi_1^2)) \mathbb{E} \left(\frac{(\gamma_{j,k}(U_1))^2}{q(U_1, V_1)} \right) a_n \\ &= \frac{2}{c_3} (C_1^2 C_2^2 + \mathbb{E}(\xi_1^2)) \iint_0^1 \frac{(\gamma_{j,k}(x))^2}{q(x, y)} q(x, y) dx dy a_n \\ &= \frac{2}{c_3} (C_1^2 C_2^2 + \mathbb{E}(\xi_1^2)) \|\gamma_{j,k}\|_2^2 a_n = C_*^2 a_n, \end{aligned} \quad (37)$$

with $C_*^2 = (2/c_3)(C_1^2 C_2^2 + \mathbb{E}(\xi_1^2))$.

Applying [6, Theorem 6.1] (see the Appendix) with $n = \mu_n = \nu_n = a_n$, $\delta = 0$, $\theta_\gamma = C_*$, $W_i = (Y_i, U_i, V_i)$,

$$q_i(\gamma, (y, x, w)) = \frac{\gamma}{q(x, w)} \gamma(x) \quad (38)$$

and $f \in B_{p_1, r_1}^{s_1}(M_1)$ (so $fe_* \in B_{p_1, r_1}^{s_1}(M_1 e_*)$) with $M_1 > 0$, $r_1 \geq 1$, either $\{p_1 \geq 2 \text{ and } s_1 \in (0, N)\}$ or $\{p_1 \in [1, 2) \text{ and } s_1 \in (1/p_1, N)\}$, we prove the existence of a constant $C > 0$ such that

$$\mathbb{E}(\|\tilde{f} - fe_*\|_2^2) \leq C \left(\frac{\ln a_n}{a_n} \right)^{2s_1/(2s_1+1)} \leq C \left(\frac{\ln n}{n} \right)^{2s_1/(2s_1+1)}, \quad (39)$$

when n is large enough.

The MISE of $\tilde{g}$ can be investigated in a similar way: for $\gamma \in \{\phi, \psi\}$, any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$.

(i) We show that

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{b_n} \sum_{i=1}^{b_n} \frac{Y_{a_n+i}}{q(U_{a_n+i}, V_{a_n+i})} \gamma_{j,k}(V_{a_n+i}) \right) \\ &= \int_0^1 (g(x) e_o) \gamma_{j,k}(x) dx. \end{aligned} \quad (40)$$

(ii) We show that

$$\sum_{i=1}^{b_n} \mathbb{E} \left(\left(\frac{Y_{a_n+i}}{q(U_{a_n+i}, V_{a_n+i})} \gamma_{j,k}(V_{a_n+i}) \right)^2 \right) \leq C_*^2 b_n, \quad (41)$$

with always $C_*^2 = (2/c_3)(C_1^2 C_2^2 + \mathbb{E}(\xi_1^2))$.

Applying again [6, Theorem 6.1] (see the Appendix) with $n = \mu_n = v_n = b_n$, $\delta = 0$, $\theta_j = C_*$, $W_i = (Y_i, U_i, V_i)$,

$$q_i(\gamma, (y, x, w)) = \frac{\gamma}{q(x, w)} \gamma(w) \quad (42)$$

and $g \in B_{p_2, r_2}^{s_2}(M_2)$ with $M_2 > 0$, $r_2 \geq 1$, either $\{p_2 \geq 2 \text{ and } s_2 \in (0, N)\}$ or $\{p_2 \in [1, 2) \text{ and } s_2 \in (1/p_2, N)\}$; we prove the existence of a constant $C > 0$ such that

$$\mathbb{E}(\|\tilde{g} - ge_o\|_2^2) \leq C \left(\frac{\ln b_n}{b_n} \right)^{2s_2/(2s_2+1)} \leq C \left(\frac{\ln n}{n} \right)^{2s_2/(2s_2+1)}, \quad (43)$$

when n is large enough.

Using the independence between $\tilde{f}$ and $\tilde{g}$ (see Remark 2), it follows from (39) and (43) that

$$\begin{aligned} \mathbb{E}(\|\tilde{g} - ge_o\|_2^2 \|\tilde{f} - fe_*\|_2^2) &= \mathbb{E}(\|\tilde{g} - ge_o\|_2^2) \mathbb{E}(\|\tilde{f} - fe_*\|_2^2) \\ &\leq C \left(\frac{\ln n}{n} \right)^{4s_1 s_2 / (2s_1+1)(2s_2+1)}. \end{aligned} \quad (44)$$

Owing to Theorem 1, (39), (43) and (44), we get

$$\begin{aligned} & \mathbb{E} \left(\iint_0^1 (\hat{h}(x, y) - h(x, y))^2 dx dy \right) \\ &\leq C \left(\mathbb{E}(\|\tilde{g} - ge_o\|_2^2) + \mathbb{E}(\|\tilde{f} - fe_*\|_2^2) \right) \\ &\quad + \mathbb{E}(\|\tilde{g} - ge_o\|_2^2 \|\tilde{f} - fe_*\|_2^2) + \frac{1}{n} \\ &\leq C \left(\left(\frac{\ln n}{n} \right)^{2s_2/(2s_2+1)} + \left(\frac{\ln n}{n} \right)^{2s_1/(2s_1+1)} \right. \\ &\quad \left. + \left(\frac{\ln n}{n} \right)^{4s_1 s_2 / (2s_1+1)(2s_2+1)} + \frac{1}{n} \right) \\ &\leq C \left(\frac{\ln n}{n} \right)^{2s_*/(2s_*+1)}, \end{aligned} \quad (45)$$

with $s_* = \min(s_1, s_2)$.

Theorem 4 is proved. $\square$

Appendix

Let us now present in detail [6, Theorem 6.1] which is used two times in the proof of Theorem 4.

We consider a general form of the hard thresholding wavelet estimator denoted by $\hat{f}_H$ for estimating an unknown function $f \in \mathbb{L}^2([0, 1])$ from n independent random variables $W_1, \dots, W_n$:

$$\hat{f}_H(x) = \sum_{k=0}^{2^\tau-1} \hat{\alpha}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa \vartheta_j\}} \psi_{j,k}(x), \quad (A.1)$$

where

$$\begin{aligned} \hat{\alpha}_{j,k} &= \frac{1}{v_n} \sum_{i=1}^n q_i(\phi_{j,k}, W_i), \\ \hat{\beta}_{j,k} &= \frac{1}{v_n} \sum_{i=1}^n q_i(\psi_{j,k}, W_i) \mathbf{1}_{\{|q_i(\psi_{j,k}, W_i)| \leq \varsigma_j\}}, \end{aligned} \quad (A.2)$$

$$\varsigma_j = \theta_\psi 2^{\delta j} \frac{v_n}{\sqrt{\mu_n \ln \mu_n}}, \quad \vartheta_j = \theta_\psi 2^{\delta j} \sqrt{\frac{\ln \mu_n}{\mu_n}},$$

$\kappa \geq 2 + 8/3 + 2\sqrt{4 + 16/9}$ and j_1 is the integer satisfying

$$\frac{1}{2} \mu_n^{1/(2\delta+1)} < 2^{j_1} \leq \mu_n^{1/(2\delta+1)}. \quad (A.3)$$

Here, we suppose that there exist

- (i) n functions $q_1, \dots, q_n$ with $q_i : \mathbb{L}^2([0, 1]) \times W_i(\Omega) \rightarrow \mathbb{C}$ for any $i \in \{1, \dots, n\}$,
- (ii) two sequences of real numbers $(v_n)_{n \in \mathbb{N}}$ and $(\mu_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \rightarrow \infty} v_n = \infty$ and $\lim_{n \rightarrow \infty} \mu_n = \infty$,

such that, for $\gamma \in \{\phi, \psi\}$,

(A1) any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$,

$$\mathbb{E} \left(\frac{1}{v_n} \sum_{i=1}^n q_i(\gamma_{j,k}, W_i) \right) = \int_0^1 f(x) \gamma_{j,k}(x) dx. \quad (\text{A.4})$$

(A2) there exist two constants, $\theta_\gamma > 0$ and $\delta \geq 0$, such that, for any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$,

$$\sum_{i=1}^n \mathbb{E} \left(|q_i(\gamma_{j,k}, W_i)|^2 \right) \leq \theta_\gamma^2 2^{2\delta j} \frac{v_n^2}{\mu_n}. \quad (\text{A.5})$$

Let $\hat{f}_H$ be (A.1) under (A1) and (A2). Suppose that $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 2 \text{ and } s \in (0, N)\}$ or $\{p \in [1, 2) \text{ and } s \in ((2\delta + 1)/p, N)\}$. Then there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\|\hat{f}_H - f\|_2^2 \right) \leq C \left(\frac{\ln \mu_n}{\mu_n} \right)^{2s/(2s+2\delta+1)}. \quad (\text{A.6})$$

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- [1] O. B. Linton and J. P. Nielsen, "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrika*, vol. 82, no. 1, pp. 93–100, 1995.
- [2] A. Yatchew and L. Bos, "Nonparametric least squares estimation and testing of economic models," *Journal of Quantitative Economics*, vol. 13, pp. 81–131, 1997.
- [3] J. Pinske, *Feasible Multivariate Nonparametric Regression Estimation Using Weak Separability*, University of British Columbia, Vancouver, Canada, 2000.
- [4] A. Lewbel and O. Linton, "Nonparametric matching and efficient estimators of homothetically separable functions," *Econometrica*, vol. 75, no. 4, pp. 1209–1227, 2007.
- [5] D. Jacho-Chávez, A. Lewbel, and O. Linton, "Identification and nonparametric estimation of a transformed additively separable model," *Journal of Econometrics*, vol. 156, no. 2, pp. 392–407, 2010.
- [6] Y. P. Chaubey, C. Chesneau, and H. Doosti, "Adaptive wavelet estimation of a density from mixtures under multiplicative censoring," 2014, <http://hal.archives-ouvertes.fr/hal-00918069>.
- [7] A. Antoniadis, "Wavelets in statistics: a review (with discussion)," *Journal of the Italian Statistical Society B*, vol. 6, no. 2, pp. 97–144, 1997.
- [8] B. Vidakovic, *Statistical Modeling by Wavelets*, John Wiley & Sons, New York, NY, USA, 1999.
- [9] W. Härdle, G. Kerkycharian, D. Picard, and A. Tsybakov, *Wavelet, Approximation and Statistical Applications*, vol. 129 of *Lectures Notes in Statistics*, Springer, New York, NY, USA, 1998.
- [10] V. A. Vasiliev, "One investigation method of a ratios type estimators," in *Proceedings of the 16th IFAC Symposium on System Identification*, pp. 1–6, Brussels, Belgium, July 2012.
- [11] A. Cohen, I. Daubechies, and P. Vial, "Wavelets on the interval and fast wavelet transforms," *Applied and Computational Harmonic Analysis*, vol. 1, no. 1, pp. 54–81, 1993.
- [12] R. DeVore and V. Popov, "Interpolation of Besov spaces," *Transactions of the American Mathematical Society*, vol. 305, pp. 397–414, 1988.
- [13] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, Cambridge, UK, 1992.
- [14] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Density estimation by wavelet thresholding," *Annals of Statistics*, vol. 24, no. 2, pp. 508–539, 1996.
- [15] B. Delyon and A. Juditsky, "On minimax wavelet estimators," *Applied and Computational Harmonic Analysis*, vol. 3, no. 3, pp. 215–228, 1996.
- [16] A. B. Tsybakov, *Introduction à L'Estimation Non-Paramétrique*, Springer, New York, NY, USA, 2004.
- [17] M. H. Neumann, "Multivariate wavelet thresholding in anisotropic function spaces," *Statistica Sinica*, vol. 10, no. 2, pp. 399–431, 2000.