On the adaptive estimation of a multiplicative separable regression function

Christophe Chesneau

Noname manuscript No.
(will be inserted by the editor)

On the adaptive estimation of a multiplicative separable regression function

Christophe Chesneau

Received:

Abstract We investigate the estimation of a multiplicative separable regression function from a bi-dimensional 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.

Keywords Nonparametric regression · Multiplicative separable regression function · Wavelet methods.

2000 Mathematics Subject Classification 62G08, 62G20.

1 Motivations

We consider the bi-dimensional 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

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

$(\xi_i)_{i\in\mathbb{Z}}$ 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 $h : [0,1]^2 \to \mathbb{R}$ 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 $f : [0,1] \to \mathbb{R}$ and $g : [0,1] \to \mathbb{R}$ such that

$$h(x, y) = f(x)g(y). \tag{2}$$

Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France. E-mail: chesneau@math.unicaen.fr

We aim to estimate h from the n random variables: $(Y_1, U_1, V_1), \ldots, (Y_n, U_n, V_n)$. This problem is plausible in many practical situations as in utility, production, and cost function applications. See, e.g., Linton and Nielsen (1995), Yatchew and Bos (1997), Pinske (2000), Lewbel and Linton (2007) and Jacho-Chávez *et al.* (2010).

In this paper, we provide a theoretical contribution to the subject by introducing a new general estimation method for h. A sharp upper bound for its associated mean integrated squared error (MISE) is proved. Then we adapt our methodology to propose an efficient and adaptive wavelet procedure. It is based on two wavelet thresholding estimators having the features to be adaptive for a wide class of unknown functions and enjoying nice MISE properties. Further details on such wavelet estimators can be found in, e.g., Antoniadis (1997), Vidakovic (1999) and Härdle *et al.* (1998). Despite the so-called "curse of dimension" coming from the bi-dimensionality 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 (1995) via non adaptive 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

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

We set

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

(provided that they exist).

We formulate the following assumptions.

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

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

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

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

(H3) The density of (U_1, V_1), denoted by q, is known and there exist two constants $c_3 > 0$ and $C_3 > 0$ such that

$$c_3 \leq \inf_{(x,y)\in[0,1]^2} q(x,y), \qquad \sup_{(x,y)\in[0,1]^2} q(x,y) \leq C_3.$$

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

$$|e_o e_*| \geq \omega.$$

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 respectively).

3 MISE result

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

Theorem 1 *We consider* (1) *under* **(H1)-(H4)**. *We introduce the following estimator for h* (2)*:*

$$\hat{h}(x,y) = \frac{\tilde{f}(x)\tilde{g}(y)}{\tilde{e}} \mathbf{1}_{\{|\tilde{e}|\geq\omega/2\}}, \tag{3}$$

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

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

and ω refers to **(H4)**.

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

$$\mathbb{E}\left(\int_0^1\int_0^1 (\hat{h}(x,y) - h(x,y))^2 dxdy\right)$$
$$\leq C\left(\mathbb{E}(||\tilde{g} - ge_o||_2^2) + \mathbb{E}(||\tilde{f} - fe_*||_2^2) + \mathbb{E}(||\tilde{g} - ge_o||_2^2||\tilde{f} - fe_*||_2^2) + \frac{1}{n}\right).$$

The form of $\tilde{h}$ (3) 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 (2012).

Based on Theorem 1, $\hat{h}$ is efficient for h if and only if $\tilde{f}$ is efficient for fe_* and $\tilde{g}$ is efficient for ge_o in terms of MISE. This result motivates the investigation of 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 (1997), Vidakovic (1999) and Härdle *et al.* (1998).

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.* (1993). Let N be a positive integer, ϕ and ψ be the initial wavelets of the Daubechies orthogonal wavelets $db2N$. We set

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

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

For any integer $\ell \geq \tau$, any $v \in \mathbb{L}^2([0,1])$ can be expanded on $\mathcal{S}$ as

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

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

$$\alpha_{j,k} = \int_0^1 v(x)\phi_{j,k}(x)dx, \qquad \beta_{j,k} = \int_0^1 v(x)\psi_{j,k}(x)dx. \tag{4}$$

4.2 Besov balls

For the sake of simplicity, we consider the sequential version of Besov balls defined as follows. Let $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$. A function v belongs to $B^s_{p,r}(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (4) satisfy

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

In this expression, s is a smoothness parameter and p and r are norm parameters. For a particular choice of s, p and r, $B^s_{p,r}(M)$ contains the Hölder and Sobolev balls. See, e.g., Devore and Popov (1988), Meyer (1992) and Härdle *et al.* (1998).

4.3 Hard thresholding estimators

In the sequel, we consider (1) under **(H1)-(H4)**.

We focus our attention on wavelet hard thresholding estimators for $\tilde{f}$ and $\tilde{g}$ in (3). 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 other are removed. This selection is the key to the adaptivity and the good performances of the hard wavelet estimators. See, e.g., Donoho *et al.* (1996), Delyon and Juditsky (1996) and Härdle *et al.* (1998).

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

$$\tilde{f}(x) = \sum_{k=0}^{2^\tau - 1} \hat{\alpha}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbf{1}_{\left\{|\hat{\beta}_{j,k}| \geq \kappa C_* \lambda_n\right\}} \psi_{j,k}(x), \qquad (5)$$

where

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

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

$$\hat{\beta}_{j,k} = \frac{1}{a_n} \sum_{i=1}^{a_n} W_{i,j,k} \mathbf{1}_{\{|W_{i,j,k}| \leq C_*/\lambda_n\}}, \qquad W_{i,j,k} = \frac{Y_i}{q(U_i, V_i)} \psi_{j,k}(U_i),$$

j_1 is the integer satisfying

$$\frac{1}{2} \frac{a_n}{\ln a_n} < 2^{j_1} \leq \frac{a_n}{\ln a_n},$$

κ is a large enough constant, $C_* = \sqrt{2(C_3/c_3^2)(C_1^2 C_2^2 + \mathbb{E}(\xi_1^2))}$ and

$$\lambda_n = \sqrt{\frac{\ln a_n}{a_n}}.$$

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

$$\tilde{g}(x) = \sum_{k=0}^{2^\tau - 1} \hat{\upsilon}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \hat{\theta}_{j,k} \mathbf{1}_{\left\{|\hat{\theta}_{j,k}| \geq \kappa_* C_* \eta_n\right\}} \psi_{j,k}(x), \qquad (6)$$

where

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

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

$$\hat{\theta}_{j,k} = \frac{1}{b_n} \sum_{i=1}^{b_n} Z_{a_n+i,j,k} \mathbf{1}_{\{|Z_{a_n+i,j,k}| \leq C_*/\eta_n\}}, \qquad Z_{a_n+i,j,k} = \frac{Y_{a_n+i}}{q(U_{a_n+i}, V_{a_n+i})} \psi_{j,k}(V_{a_n+i}),$$

j_2 is the integer satisfying

$$\frac{1}{2}\frac{b_n}{\ln b_n} < 2^{j_2} \leq \frac{b_n}{\ln b_n},$$

κ_* is a large enough constant, $C_* = \sqrt{2(C_3/c_3^2)(C_1^2C_2^2 + \mathbb{E}(\xi_1^2))}$ and

$$\eta_n = \sqrt{\frac{\ln b_n}{b_n}}.$$

Estimator for h: From $\tilde{f}$ (5) and $\tilde{g}$ (6), we consider the following estimator for h (2):

$$\hat{h}(x, y) = \frac{\tilde{f}(x)\tilde{g}(y)}{\tilde{e}}\mathbf{1}_{\{|\tilde{e}|\geq \omega/2\}}, \tag{7}$$

where

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

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

Remark 2 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 (Chaubey *et al.* 2013, Theorem 6.1). Further details are given in the proof of Theorem 2.

4.4 Rate of convergence

Theorem 2 investigates the rate of convergence attains by $\hat{h}$ under the MISE over Besov balls.

Theorem 2 *We consider* (1) *under* **(H1)-(H4)**. *Let* $\hat{h}$ *be* (7) *and* h *be* (2). *Suppose that*

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

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

$$\mathbb{E}\left(\int_0^1 \int_0^1 (\hat{h}(x,y) - h(x,y))^2 dxdy\right) \leq C \left(\frac{\ln n}{n}\right)^{2s_*/(2s_*+1)},$$

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 $B^{s_*}_{p,r}(M)$. See, e.g., Tsybakov (2004) and Härdle *et al.* (1998). In this sense, Theorem 2 proves that our estimator escapes to the so-called "curse of dimension". Such a result is not possible with the standard bi-dimensional hard thresholding estimator attaining the rate of convergence $(\ln n/n)^{2s/(2s+d)}$ with $d = 2$ under the MISE over bi-dimensional Besov balls defined with s as smoothness parameter. See Delyon and Juditsky (1996).

Theorem 2 completes asymptotic results proved by Linton and Nielsen (1995) investigating this problem for the structured nonparametric regression model via another estimation method based on non adaptive kernels.

Remark 3 In Theorem 2, we take into account both the homogeneous zone of Besov balls, i.e., $\{p_1 \geq 2$ and $s_1 > 0\}$, and the inhomogeneous zone, i.e., $\{p_1 \in [1, 2)$ and $s_1 > 1/p_1\}$, for the case $f \in B^{s_1}_{p_1,r_1}(M_1)$, and the same for $g \in B^{s_2}_{p_2,r_2}(M_2)$. This has the advantage to cover a very rich class of unknown regression functions h.

Remark 4 Note that Theorem 2 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 5 Our study can be extended to the multidimensional case considered by Yatchew and Bos (1997), i.e., $f : [0, 1]^{q_1} \to \mathbb{R}$ and $g : [0, 1]^{q_2} \to \mathbb{R}$, q_1 and q_2 denoting two positive integer. In this case, adapting our framework to the multidimensional case (q_1 dimensional Besov balls, q_1 dimensional (tensorial) wavelet basis, q_1 dimensional wavelet hard thresholding estimator,. . . see, for instance, Delyon and Juditsky (1996)), one can prove that (3) 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

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

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

$$\begin{aligned} |\hat{h}(x,y) - h(x,y)| &\leq \frac{2}{\omega} |\tilde{f}(x)\tilde{g}(y) - f(x)g(y)\tilde{e}| + |f(x)||g(y)| \mathbf{1}_{\{|\tilde{e}| < \omega/2\}} \\ &\leq C\left(|\tilde{f}(x)\tilde{g}(y) - f(x)g(y)\tilde{e}| + \mathbf{1}_{\{|\tilde{e} - e_* e_o| \geq \omega/2\}} \right) \\ &\leq C\left(|\tilde{f}(x)\tilde{g}(y) - f(x)g(y)\tilde{e}| + |\tilde{e} - e_* e_o| \right). \end{aligned} \tag{8}$$

On the other hand, we have the decomposition

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

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

$$\begin{aligned} &|\tilde{f}(x)\tilde{g}(y) - f(x)g(y)\tilde{e}| \\ &\leq C\big(|\tilde{g}(y) - g(y)e_o| + |\tilde{f}(x) - f(x)e_*| + |\tilde{g}(y) - g(y)e_o||\tilde{f}(x) - f(x)e_*| \\ &+ |\tilde{e} - e_* e_o|\big). \end{aligned} \tag{9}$$

Putting (8) and (9) together, we obtain

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

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

$$\begin{aligned} &\mathbb{E}\left(\int_0^1 \int_0^1 (\hat{h}(x,y) - h(x,y))^2 dxdy \right) \\ &\leq C\big(\mathbb{E}(||\tilde{g} - ge_o||_2^2) + \mathbb{E}(||\tilde{f} - fe_*||_2^2) + \mathbb{E}(||\tilde{g} - ge_o||_2^2 ||\tilde{f} - fe_*||_2^2) \\ &+ \mathbb{E}((\tilde{e} - e_* e_o)^2)\big). \end{aligned} \tag{10}$$

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

$$\begin{aligned} \mathbb{E}(\tilde{e}) &= \mathbb{E}\left(\frac{Y_1}{q(U_1,V_1)} \right) = \mathbb{E}\left(\frac{h(U_1,V_1)}{q(U_1,V_1)} \right) + \mathbb{E}(\xi_1)\mathbb{E}\left(\frac{1}{q(U_1,V_1)} \right) \\ &= \int_0^1 \int_0^1 \frac{f(x)g(y)}{q(x,y)} q(x,y) dxdy = \left(\int_0^1 f(x)dx \right) \left(\int_0^1 g(y)dy \right) = e_* e_o. \end{aligned} \tag{11}$$

Then, using similar arguments to (11), $(a+b)^2 \leq 2(a^2+b^2)$, $(a,b) \in \mathbb{R}^2$, **(H1)**, **(H2)**, **(H3)** and $\mathbb{E}(\xi_1^2) < \infty$, we have

$$\begin{aligned} \mathbb{E}((\tilde{e} - e_* e_o)^2) = \mathbb{V}(\tilde{e}) &= \frac{1}{n}\mathbb{V}\left(\frac{Y_1}{q(U_1,V_1)}\right) \leq \frac{1}{n}\mathbb{E}\left(\left(\frac{Y_1}{q(U_1,V_1)}\right)^2\right) \\ &\leq \frac{2}{n}\mathbb{E}\left(\frac{(h(U_1,V_1))^2 + \xi_1^2}{(q(U_1,V_1))^2}\right) \leq \frac{2}{c_3^2}\left(C_1^2 C_2^2 + \mathbb{E}(\xi_1^2)\right)\frac{1}{n} \\ &= C\frac{1}{n}. \end{aligned} \tag{12}$$

Equations (10) and (12) yield the desired inequality:

$$\begin{aligned} &\mathbb{E}\left(\int_0^1\int_0^1 (\hat{h}(x,y) - h(x,y))^2 dxdy\right) \\ &\leq C\left(\mathbb{E}(||\tilde{g} - ge_o||_2^2) + \mathbb{E}(||\tilde{f} - fe_*||_2^2) + \mathbb{E}(||\tilde{g} - ge_o||_2^2||\tilde{f} - fe_*||_2^2) + \frac{1}{n}\right). \end{aligned}$$

□

Proof of Theorem 2. We aim to apply Theorem 1 by investigating the rate of convergence attains 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 $k \in \{0, \ldots, 2^j - 1\}$,

– using similar arguments to (11), we obtain

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

– using similar arguments to (12) and $||\gamma_{j,k}||_2^2 = 1$, we have

$$\sum_{i=1}^{a_n} \mathbb{E}\left(\left(\frac{Y_i}{q(U_i,V_i)}\gamma_{j,k}(U_i)\right)^2\right)$$
$$= \mathbb{E}\left(\left(\frac{Y_1}{q(U_1,V_1)}\gamma_{j,k}(U_1)\right)^2\right) a_n$$
$$\leq 2\mathbb{E}\left(\frac{(h(U_1,V_1))^2+\xi_1^2}{(q(U_1,V_1))^2}(\gamma_{j,k}(U_1))^2\right) a_n$$
$$\leq \frac{2}{c_3^2}\left(C_1^2C_2^2+\mathbb{E}(\xi_1^2)\right)\mathbb{E}\left((\gamma_{j,k}(U_1))^2\right) a_n$$
$$= \frac{2}{c_3^2}\left(C_1^2C_2^2+\mathbb{E}(\xi_1^2)\right)\int_0^1 (\gamma_{j,k}(x))^2\left(\int_0^1 q(x,y)dy\right) dx a_n$$
$$\leq \frac{2C_3}{c_3^2}\left(C_1^2C_2^2+\mathbb{E}(\xi_1^2)\right)||\gamma_{j,k}||_2^2 a_n = C_*^2 a_n,$$

with $C_*^2 = 2(C_3/c_3^2)(C_1^2C_2^2+\mathbb{E}(\xi_1^2))$.

Applying (Chaubey *et al.* 2013, Theorem 6.1) (see Appendix) with "$n = \mu_n = \upsilon_n = a_n$", "$\delta = 0$", "$\theta_\gamma = C_*$", and $f \in B^{s_1}_{p_1,r_1}(M_1)$ (so $fe_* \in B^{s_1}_{p_1,r_1}(M_1e_*)$) with $M_1 > 0, r_1 \geq 1$, either $\{p_1 \geq 2$ and $s_1 > 0\}$ or $\{p_1 \in [1,2)$ and $s_1 > 1/p_1\}$, we prove the existence of a constant $C > 0$ such that

$$\mathbb{E}(||\tilde{f} - fe_*||_2^2) \leq C\left(\frac{\ln a_n}{a_n}\right)^{2s_1/(2s_1+1)} \leq C\left(\frac{\ln n}{n}\right)^{2s_1/(2s_1+1)}, \qquad (13)$$

for n 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 $k \in \{0,\ldots,2^j-1\}$,

– we show that

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

– we show that

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

with always $C_*^2 = 2(C_3/c_3^2)(C_1^2C_2^2+\mathbb{E}(\xi_1^2))$.

Applying again (Chaubey *et al.* 2013, Theorem 6.1) (see Appendix) with "$n = \mu_n = \upsilon_n = b_n$", "$\delta = 0$", "$\theta_\gamma = C_*$" and $g \in B^{s_2}_{p_2,r_2}(M_2)$ with $M_2 > 0$, $r_2 \geq 1$, either $\{p_2 \geq 2$ and $s_2 > 0\}$ or $\{p_2 \in [1,2)$ and $s_2 > 1/p_2\}$, we prove the existence of a constant $C > 0$ such that

$$\mathbb{E}(||\tilde{g} - ge_o||_2^2) \leq C\left(\frac{\ln b_n}{b_n}\right)^{2s_2/(2s_2+1)} \leq C\left(\frac{\ln n}{n}\right)^{2s_2/(2s_2+1)}, \qquad (14)$$

for n large enough.

Using the independence between $\tilde{f}$ and $\tilde{g}$ (see Remark 1), it follows from (13) and (14) that

$$\mathbb{E}(||\tilde{g} - ge_o||_2^2||\tilde{f} - fe_*||_2^2) = \mathbb{E}(||\tilde{g} - ge_o||_2^2)\mathbb{E}(||\tilde{f} - fe_*||_2^2) \leq C\left(\frac{\ln n}{n}\right)^{4s_1s_2/(2s_1+1)(2s_2+1)}. \quad (15)$$

Owing to Theorem 1, (13), (14) and (15), we get

$$\mathbb{E}\left(\int_0^1\int_0^1 (\hat{h}(x,y) - h(x,y))^2 dxdy\right)$$
$$\leq C\left(\mathbb{E}(||\tilde{g} - ge_o||_2^2) + \mathbb{E}(||\tilde{f} - fe_*||_2^2) + \mathbb{E}(||\tilde{g} - ge_o||_2^2||\tilde{f} - fe_*||_2^2) + \frac{1}{n}\right)$$
$$\leq C\left(\left(\frac{\ln n}{n}\right)^{2s_2/(2s_2+1)} + \left(\frac{\ln n}{n}\right)^{2s_1/(2s_1+1)} + \left(\frac{\ln n}{n}\right)^{4s_1s_2/(2s_1+1)(2s_2+1)} + \frac{1}{n}\right)$$
$$\leq C\left(\frac{\ln n}{n}\right)^{2s_*/(2s_*+1)},$$

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

Theorem 2 is proved.

□

Appendix

Let us now present in details (Chaubey *et al.* 2013, Theorem 6.1) used two times in the proof of Theorem 2.

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

$$\widehat{f}_H(x) = \sum_{k=0}^{2^\tau-1} \widehat{\alpha}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \widehat{\beta}_{j,k}\mathbf{1}_{\left\{|\widehat{\beta}_{j,k}|\geq\kappa\vartheta_j\right\}}\psi_{j,k}(x), \quad (16)$$

where

$$\widehat{\alpha}_{j,k} = \frac{1}{\upsilon_n}\sum_{i=1}^n q_i(\phi_{j,k}, W_i),$$

$$\widehat{\beta}_{j,k} = \frac{1}{\upsilon_n}\sum_{i=1}^n q_i(\psi_{j,k}, W_i)\mathbf{1}_{\{|q_i(\psi_{j,k},W_i)|\leq\varsigma_j\}},$$

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

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

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

Here, we suppose that there exist

- *n functions $q_1, \ldots, q_n$ with $q_i : \mathbb{L}^2([0,1]) \times \mathbb{R} \to \mathbb{R}$ for any $i \in \{1, \ldots, n\}$,*
- *two sequences of real numbers $(\upsilon_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty} \upsilon_n = \infty$ and $\lim_{n\to\infty} \mu_n = \infty$*

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

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

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

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

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

Let $\hat{f}_H$ be (16) under (A1) and (A2). Suppose that $f \in B^s_{p,r}(M)$ with $r \geq 1$, $\{p \geq 2$ and $s \in (0, N)\}$ or $\{p \in [1, 2)$ and $s \in ((2\delta + 1)/p, N)\}$. Then there exists a constant $C > 0$ such that

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

□

References

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society Series B*, 6, 97-144.

Chaubey, Y.P., Chesneau, C. and Doosti, H. (2013). Adaptive wavelet estimation of a density from mixtures under multiplicative censoring, in Revision to *Statistics*.

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.

Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, 3, 215–228.

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

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *The Annals of Statistics*, 24, 508–539.

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

Jacho-Chávez, D., Lewbel, A. and Linton, O. (2010). Identification And Nonparametric Estimation Of A Transformed Additively Separable Model, *Journal of Econometrics*, 156(2), 392-407.

Lewbel, A. and Linton, O. (2007). Nonparametric Matching and Efficient Estimators of Homothetically Separable Functions, *Econometrica*, 75, 1209-1228.

Linton, O. B. and Nielsen, J.P. (1995). A Kernel Model of Estimating Structured Nonparametric Regression Based on Marginal Integration, *Biometrika*, 82, 93-100.

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

Pinske, J. (2000). Feasible Multivariate Nonparametric Regression Estimation Using Weak Separability. Preprint, University of British Columbia, Canada

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

Vasiliev, V.A. (2012). One investigation method of a ratios type estimators. 16th IFAC Symposium on System Identification, Brussels, Belgium, July 11-13, pages 1-6, 2012 (in progress).

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

Yatchew, A. and Bos, L. (1997). Nonparametric Least Squares Estimation and Testing of Economic Models, *Journal of Quantitative Economics*, 13, 81-131.