



HAL
open science

Kernel Selection in Nonparametric Regression

Hélène Halconruy, Nicolas Marie

► **To cite this version:**

Hélène Halconruy, Nicolas Marie. Kernel Selection in Nonparametric Regression. 2020. hal-02867190v1

HAL Id: hal-02867190

<https://hal.science/hal-02867190v1>

Preprint submitted on 13 Jun 2020 (v1), last revised 20 Mar 2021 (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.

KERNEL SELECTION IN NONPARAMETRIC REGRESSION

HÉLÈNE HALCONRUY* AND NICOLAS MARIE†

ABSTRACT. In the regression model $Y = b(X) + \varepsilon$, where X has a density f , this paper deals with an oracle inequality for an estimator of bf , involving a kernel in the sense of Lerasle et al. (2016), selected via the PCO method. In addition to the bandwidth selection for kernel-based estimators already studied in Lacour, Massart and Rivoirard (2017) and Comte and Marie (2020), the dimension selection for anisotropic projection estimators of f and bf is covered.

CONTENTS

1. Introduction	1
2. Risk bound	2
3. Kernel selection	4
4. Basic numerical experiments	6
Appendix A. Details on kernels sets: proofs of Propositions 2.2, 2.3 and 3.3	8
A.1. Proof of Proposition 2.2	8
A.2. Proof of Proposition 2.3	8
A.3. Proof of Proposition 3.3	9
Appendix B. Proofs of risk bounds	11
B.1. Preliminary results	11
B.2. Proof of Proposition 2.4	17
B.3. Proof of Theorem 2.5	17
B.4. Proof of Theorem 3.2	18
References	20

MSC2010: 62G05 ; 62G08.

1. INTRODUCTION

Consider $n \in \mathbb{N}^*$ independent $\mathbb{R}^d \times \mathbb{R}$ -valued ($d \in \mathbb{N}^*$) random variables $(X_1, Y_1), \dots, (X_n, Y_n)$, having the same probability distribution assumed to be absolutely continuous with respect to Lebesgue's measure, and

$$\widehat{s}_{K,\ell}(n; x) := \frac{1}{n} \sum_{i=1}^n K(X_i, x) \ell(Y_i) ; x \in \mathbb{R}^d,$$

where $\ell : \mathbb{R} \rightarrow \mathbb{R}$ is a Borel function and K is a symmetric continuous map from $\mathbb{R}^d \times \mathbb{R}^d$ into \mathbb{R} . This is an estimator of the function $s : \mathbb{R}^d \rightarrow \mathbb{R}$ defined by

$$s(x) := \mathbb{E}(\ell(Y_1) | X_1 = x) f(x) ; \forall x \in \mathbb{R}^d,$$

where f is a density of X_1 . For $\ell = 1$, $\widehat{s}_{K,\ell}(n; \cdot)$ coincides with the estimator of f studied in Lerasle et al. [11], but for $\ell \neq 1$, it covers estimators involved in nonparametric regression. Assume that for every $i \in \{1, \dots, n\}$,

$$(1) \quad Y_i = b(X_i) + \varepsilon_i$$

where ε_i is a centered random variable, independent of X_i , and $b : \mathbb{R}^d \rightarrow \mathbb{R}$ is a Borel function.

Key words and phrases. Nonparametric estimators ; Projection estimators ; Model selection ; Regression model.

- If $\ell = \text{Id}_{\mathbb{R}}$, k is a symmetric kernel and

$$(2) \quad K(x', x) = \prod_{q=1}^d \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) \text{ with } h_1, \dots, h_d > 0$$

for every $x, x' \in \mathbb{R}^d$, then $\widehat{s}_{K,\ell}(n; \cdot)$ is the numerator of Nadaraya-Watson's estimator of the regression function b . Precisely, $\widehat{s}_{K,\ell}(n; \cdot)$ is an estimator of $s = bf$. If $\ell \neq \text{Id}_{\mathbb{R}}$, then $\widehat{s}_{K,\ell}(n; \cdot)$ is the numerator of the estimator studied in Einmahl and Mason [5, 6].

- If $\ell = \text{Id}_{\mathbb{R}}$, $\mathcal{B}_{m_q} = \{\varphi_1^{m_q}, \dots, \varphi_{m_q}^{m_q}\}$ ($m_q \in \mathbb{N}^*$ and $q \in \{1, \dots, d\}$) is an orthonormal family of $\mathbb{L}^2(\mathbb{R})$ and

$$(3) \quad K(x', x) = \prod_{q=1}^d \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q) \varphi_j^{m_q}(x'_q)$$

for every $x, x' \in \mathbb{R}^d$, then $\widehat{s}_{K,\ell}(n; \cdot)$ is the projection estimator on $\mathcal{S} = \text{span}(\mathcal{B}_{m_1} \otimes \dots \otimes \mathcal{B}_{m_d})$ of $s = bf$.

Now, assume that for every $i \in \{1, \dots, n\}$, Y_i is defined by the heteroscedastic model

$$(4) \quad Y_i = \sigma(X_i)\varepsilon_i,$$

where ε_i is a centered random variable of variance 1, independent of X_i , and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}$ is a Borel function. If $\ell(x) = x^2$ for every $x \in \mathbb{R}$, then $\widehat{s}_{K,\ell}(n; \cdot)$ is an estimator of $s = \sigma^2 f$.

These ten last years, several data-driven procedures have been proposed in order to select the bandwidth of Parzen-Rosenblatt's estimator ($\ell = 1$ and K defined by (2)). First, Goldenshluger-Lepski's method, introduced in [8], which reaches the adequate bias-variance compromise, but is not completely satisfactory on the numerical side (see Comte and Rebafka [4]). More recently, in [10], Lacour, Massart and Rivoirard proposed the PCO (Penalized Comparison to Overfitting) method and proved an oracle inequality for the associated adaptative Parzen-Rosenblatt's estimator by using a concentration inequality for the U-statistics due to Houdré and Reynaud-Bouret [9]. Together with Varet, they established the numerical efficiency of the PCO method in Varet et al. [13].

Comte and Marie [3] deals with an oracle inequality and numerical experiments for an adaptative Nadaraya-Watson's estimator with a numerator and a denominator having distinct bandwidths, both selected via the PCO method. Since the output variable in a regression model has no reason to be bounded, there were significant additional difficulties, bypassed in [3], to establish an oracle inequality for the numerator's adaptative estimator. Via similar arguments, the present article deals with an oracle inequality for $\widehat{s}_{\widehat{K},\ell}(n; \cdot)$, where \widehat{K} is selected via the PCO method in the spirit of Lerasle et al. [11]. In addition to the bandwidth selection for kernel-based estimators already studied in [10, 3], it covers the dimension selection for anisotropic projection estimators of f , bf (when Y_1, \dots, Y_n are defined by Model (1)) and $\sigma^2 f$ (when Y_1, \dots, Y_n are defined by Model (4)). As for the bandwidth selection for kernel based estimators, for $d > 1$, the PCO method allows to bypass the numerical difficulties generated by the Goldenshluger-Lepski type method involved in the anisotropic model selection procedures (see Chagny [1]).

In Section 2, some examples of kernels sets are provided and a risk bound for $\widehat{s}_{K,\ell}(n; \cdot)$ is established. Section 3 deals with an oracle inequality for $\widehat{s}_{\widehat{K},\ell}(n; \cdot)$, where \widehat{K} is selected via the PCO method. Finally, Section 4 deals with a basic numerical study.

2. RISK BOUND

Throughout the paper, $s \in \mathbb{L}^2(\mathbb{R}^d)$. Let \mathcal{K}_n be a set of symmetric continuous maps from $\mathbb{R}^d \times \mathbb{R}^d$ into \mathbb{R} , of cardinal less or equal than n , fulfilling the following assumption.

Assumption 2.1. *There exists a deterministic constant $\mathfrak{m}_{\mathcal{K},\ell} > 0$, not depending on n , such that*

(1) For every $K \in \mathcal{K}_n$,

$$\sup_{x' \in \mathbb{R}^d} \|K(x', \cdot)\|_2^2 \leq \mathfrak{m}_{\mathcal{K}, \ell} n.$$

(2) For every $K \in \mathcal{K}_n$,

$$\|s_{K, \ell}\|_2^2 \leq \mathfrak{m}_{\mathcal{K}, \ell}$$

with

$$s_{K, \ell} := \mathbb{E}(\widehat{s}_{K, \ell}(n; \cdot)) = \mathbb{E}(K(X_1, \cdot)\ell(Y_1)).$$

(3) For every $K, K' \in \mathcal{K}_n$,

$$\mathbb{E}(\langle K(X_1, \cdot), K'(X_2, \cdot)\ell(Y_2) \rangle_2^2) \leq \mathfrak{m}_{\mathcal{K}, \ell} \mathfrak{s}_{K', \ell}$$

with

$$\mathfrak{s}_{K', \ell} := \mathbb{E}(\|K'(X_1, \cdot)\ell(Y_1)\|_2^2).$$

(4) For every $K \in \mathcal{K}_n$ and $\psi \in \mathbb{L}^2(\mathbb{R}^d)$,

$$\mathbb{E}(\langle K(X_1, \cdot), \psi \rangle_2^2) \leq \mathfrak{m}_{\mathcal{K}, \ell} \|\psi\|_2^2.$$

The elements of \mathcal{K}_n are called kernels. Let us provide two natural examples of kernels sets.

Proposition 2.2. Consider

$$\mathcal{K}_k(h_{\min}) := \left\{ (x', x) \mapsto \prod_{q=1}^d \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) ; h_1, \dots, h_d \in \{h_{\min}, \dots, 1\} \right\},$$

where k is a symmetric kernel (in the usual sense) and $nh_{\min}^d \geq 1$. The kernels set $\mathcal{K}_k(h_{\min})$ fulfills Assumption 2.1 and, for any $K \in \mathcal{K}_k(h_{\min})$ such that

$$K(x', x) = \prod_{q=1}^d \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) ; \forall x, x' \in \mathbb{R}^d$$

with $h_1, \dots, h_d \in \{h_{\min}, \dots, 1\}$,

$$\mathfrak{s}_{K, \ell} = \|k\|_2^{2d} \mathbb{E}(\ell(Y_1)^2) \prod_{q=1}^d \frac{1}{h_q}.$$

Proposition 2.3. Consider

$$\mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max}) := \left\{ (x', x) \mapsto \prod_{q=1}^d \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q) \varphi_j^{m_q}(x'_q) ; m_1, \dots, m_d \in \{1, \dots, m_{\max}\} \right\},$$

where $m_{\max}^d \in \{1, \dots, n\}$ and, for every $m \in \{1, \dots, n\}$, $\mathcal{B}_m = \{\varphi_1^m, \dots, \varphi_m^m\}$ is an orthonormal family of $\mathbb{L}^2(\mathbb{R})$ such that

$$\sup_{x' \in \mathbb{R}} \sum_{j=1}^m \varphi_j^m(x')^2 \leq \mathfrak{m}_{\mathcal{B}} m$$

with $\mathfrak{m}_{\mathcal{B}} > 0$ not depending on m and n , and

$$(5) \quad \mathcal{B}_m \subset \mathcal{B}_{m+1} ; \forall m \in \{1, \dots, n-1\}$$

or

$$(6) \quad \bar{\mathfrak{m}}_{\mathcal{B}} := \sup\{|\mathbb{E}(K(X_1, x))| ; K \in \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max}) \text{ and } x \in \mathbb{R}^d\} \text{ is finite and doesn't depend on } n.$$

The kernels set $\mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$ fulfills Assumption 2.1 and, for any $K \in \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$ such that

$$K(x', x) = \prod_{q=1}^d \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q) \varphi_j^{m_q}(x'_q) ; \forall x, x' \in \mathbb{R}^d$$

with $m_1, \dots, m_n \in \{1, \dots, m_{\max}\}$,

$$\mathfrak{s}_{K,\ell} \leq \mathfrak{m}_B^d \mathbb{E}(\ell(Y_1)^2) \prod_{q=1}^d m_q.$$

Remark. Note that Condition (5) (resp. (6)) is close to (resp. the same that) Condition (19) (resp. (20)) of Lerasle et al. [11], Proposition 3.2. See also Massart [12], Chapter 7 on these conditions. For instance, the trigonometric basis and Hermite's basis satisfy Condition (5). The regular histograms basis satisfy Condition (6). Indeed, by taking $\varphi_j^m = \psi_j^m := \sqrt{m} \mathbf{1}_{[(j-1)/m, j/m[}$ for every $m \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$,

$$\begin{aligned} \left| \mathbb{E} \left[\prod_{q=1}^d \sum_{j=1}^{m_q} \psi_j^{m_q}(X_{1,q}) \psi_j^{m_q}(x_q) \right] \right| &= \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \left(\prod_{q=1}^d m_q \mathbf{1}_{[(j_q-1)/m_q, j_q/m_q[}(x_q) \right) \\ &\quad \times \int_{(j_1-1)/m_1}^{j_1/m_1} \cdots \int_{(j_d-1)/m_d}^{j_d/m_d} f(x'_1, \dots, x'_d) dx'_1 \cdots dx'_d \\ &\leq \|f\|_\infty \prod_{q=1}^d \sum_{j=1}^{m_q} \mathbf{1}_{[(j-1)/m_q, j/m_q[}(x) \leq \|f\|_\infty \end{aligned}$$

for every $m_1, \dots, m_d \in \{1, \dots, n\}$ and $x \in \mathbb{R}^d$.

The following proposition provides a suitable control of the variance of $\widehat{s}_{K,\ell}(n; \cdot)$.

Proposition 2.4. *Under Assumption 2.1.(1,2,3), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{2.4} > 0$, not depending on n , such that for every $\theta \in]0, 1[$,*

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \|\widehat{s}_{K,\ell}(n; \cdot) - s_{K,\ell}\|_2^2 - \frac{\mathfrak{s}_{K,\ell}}{n} \right\} - \frac{\theta}{n} \mathfrak{s}_{K,\ell} \right) \leq \mathfrak{c}_{2.4} \frac{\log(n)^5}{\theta n}.$$

Finally, let us state the main result of this section.

Theorem 2.5. *Under Assumption 2.1, if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathfrak{c}_{2.5}, \bar{\mathfrak{c}}_{2.5} > 0$, not depending on n , such that for every $\theta \in]0, 1[$,*

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 - (1 + \theta) \left(\|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} \right) \right\} \right) \leq \mathfrak{c}_{2.5} \frac{\log(n)^5}{\theta n}$$

and

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} - \frac{1}{1 - \theta} \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 \right\} \right) \leq \bar{\mathfrak{c}}_{2.5} \frac{\log(n)^5}{\theta(1 - \theta)n}.$$

Remark. Note that the first inequality in Theorem 2.5 gives a risk bound on the estimator $\widehat{s}_{K,\ell}(n; \cdot)$:

$$\mathbb{E}(\|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2) \leq (1 + \theta) \left(\|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} \right) + \mathfrak{c}_{2.5} \frac{\log(n)^5}{\theta n}$$

for every $\theta \in]0, 1[$. The second inequality is useful in order to establish a risk bound on the adaptive estimator defined in the next section (see Theorem 3.2).

3. KERNEL SELECTION

This section deals with a risk bound on the adaptive estimator $\widehat{s}_{\widehat{K},\ell}(n; \cdot)$, where

$$\widehat{K} \in \arg \min_{K \in \mathcal{K}_n} \{ \|\widehat{s}_{K,\ell}(n; \cdot) - \widehat{s}_{K_0,\ell}(n; \cdot)\|_2^2 + \text{pen}(K) \},$$

K_0 is an overfitting proposal for K and

$$(7) \quad \text{pen}(K) := \frac{2}{n^2} \sum_{i=1}^n \langle K(\cdot, X_i), K_0(\cdot, X_i) \rangle_2 \ell(Y_i)^2; \quad \forall K \in \mathcal{K}_n.$$

Example. For $\mathcal{K}_n = \mathcal{K}_k(h_{\min})$, one should take

$$K_0(x', x) = \frac{1}{h_{\min}^d} \prod_{q=1}^d k\left(\frac{x'_q - x_q}{h_{\min}}\right); \forall x, x' \in \mathbb{R}^d,$$

and for $\mathcal{K}_n = \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$, one should take

$$K_0(x', x) = \prod_{q=1}^d \sum_{j=1}^{m_{\max}} \varphi_j^{m_{\max}}(x_q) \varphi_j^{m_{\max}}(x'_q); \forall x, x' \in \mathbb{R}^d.$$

In the sequel, in addition to Assumption 2.1, the kernels set \mathcal{K}_n fulfills the following assumption.

Assumption 3.1. *There exists a deterministic constant $\bar{\mathbf{m}}_{\mathcal{K}, \ell} > 0$, not depending on n , such that*

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \langle K(X_1, \cdot), s_{K', \ell} \rangle_2^2 \right) \leq \bar{\mathbf{m}}_{\mathcal{K}, \ell}.$$

The following theorem provides an oracle inequality for the adaptative estimator $\widehat{s}_{\widehat{K}, \ell}(n; \cdot)$.

Theorem 3.2. *Under Assumptions 2.1 and 3.1, if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathbf{c}_{3.2} > 0$, not depending on n , such that for every $\vartheta \in]0, 1[$,*

$$\mathbb{E}(\|\widehat{s}_{\widehat{K}, \ell}(n; \cdot) - s\|_2^2) \leq (1 + \vartheta) \min_{K \in \mathcal{K}_n} \mathbb{E}(\|\widehat{s}_{K, \ell}(n; \cdot) - s\|_2^2) + \frac{\mathbf{c}_{3.2}}{\vartheta} \left(\|s_{K_0, \ell} - s\|_2^2 + \frac{\log(n)^5}{n} \right).$$

Finally, let us discuss about Assumption 3.1. Note that if s is bounded and

$$\mathbf{m}_{\mathcal{K}} := \sup\{\|K(x', \cdot)\|_1^2; K \in \mathcal{K}_n \text{ and } x' \in \mathbb{R}^d\}$$

doesn't depend on n , then \mathcal{K}_n fulfills Assumption 3.1. Indeed,

$$\begin{aligned} \mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \langle K(X_1, \cdot), s_{K', \ell} \rangle_2^2 \right) &\leq \left(\sup_{K' \in \mathcal{K}_n} \|s_{K', \ell}\|_{\infty}^2 \right) \mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \|K(X_1, \cdot)\|_1^2 \right) \\ &\leq \mathbf{m}_{\mathcal{K}} \sup \left\{ \left(\int_{-\infty}^{\infty} |K'(x', x) s(x)| dx \right)^2; K' \in \mathcal{K}_n \text{ and } x' \in \mathbb{R}^d \right\} \leq \mathbf{m}_{\mathcal{K}}^2 \|s\|_{\infty}^2. \end{aligned}$$

In the nonparametric regression framework (see Model (1)), to assume s bounded means that bf is bounded. For instance, this condition is fulfilled by the linear regression models with Gaussian inputs. The following examples focus on the condition on $\mathbf{m}_{\mathcal{K}}$.

Examples:

- (1) Consider $K \in \mathcal{K}_k(h_{\min})$. Then, there exists $h_1, \dots, h_d \in \{h_{\min}, \dots, 1\}$ such that

$$K(x', x) = \prod_{q=1}^d \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right); \forall x, x' \in \mathbb{R}^d.$$

Clearly, $\|K(x', \cdot)\|_1 = \|k\|_1^d$ for every $x' \in \mathbb{R}^d$. So, for $\mathcal{K}_n = \mathcal{K}_k(h_{\min})$, $\mathbf{m}_{\mathcal{K}} \leq \|k\|_1^{2d}$.

- (2) For $\mathcal{K}_n = \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$, the condition on $\mathbf{m}_{\mathcal{K}}$ seems harder to check in general. Let us show that it is satisfied for the regular histograms basis defined in Section 2. For every $m_1, \dots, m_d \in \{1, \dots, n\}$,

$$\left\| \prod_{q=1}^d \sum_{j=1}^{m_q} \psi_j^{m_q}(x'_q) \psi_j^{m_q}(\cdot) \right\|_1 \leq \prod_{q=1}^d \left(m_q \sum_{j=1}^{m_q} \mathbf{1}_{[(j-1)/m_q, j/m_q]}(x'_q) \int_{(j-1)/m_q}^{j/m_q} dx \right) \leq 1.$$

The following proposition shows that $\mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$ fulfills Assumption 3.1 for the trigonometric basis, even if the condition on $\mathbf{m}_{\mathcal{K}}$ is not satisfied.

Proposition 3.3. Consider $\chi_1 := \mathbf{1}_{[0,1]}$ and, for every $j \in \mathbb{N}^*$, the functions χ_{2j} and χ_{2j+1} defined on \mathbb{R} by

$$\chi_{2j}(x) := \sqrt{2} \cos(2\pi j x) \mathbf{1}_{[0,1]}(x) \text{ and } \chi_{2j+1}(x) := \sqrt{2} \sin(2\pi j x) \mathbf{1}_{[0,1]}(x) ; \forall x \in \mathbb{R}.$$

If $s \in C^2(\mathbb{R}^d)$ and $\mathcal{B}_m = \{\chi_1, \dots, \chi_m\}$ for every $m \in \{1, \dots, n\}$, then $\mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$ fulfills Assumption 3.1.

4. BASIC NUMERICAL EXPERIMENTS

Throughout this section, $d = 1$, $\ell \in \{1, \text{Id}_{\mathbb{R}}\}$ and Y_1, \dots, Y_n are defined by Model (1) with $\varepsilon_1, \dots, \varepsilon_n \rightsquigarrow \mathcal{N}(0, 1)$. Some numerical experiments on $\widehat{s}_{K,1}(n; \cdot)$ (resp. $\widehat{s}_{K, \text{Id}_{\mathbb{R}}}(n; \cdot)$) for $K \in \mathcal{K}_k(h_{\min})$ have already been done in Varet et al. [13] (resp. Comte and Marie [3]). So, this section deals with basic numerical experiments on $\widehat{s}_{K,1}(n; \cdot)$ and $\widehat{s}_{K, \text{Id}_{\mathbb{R}}}(n; \cdot)$ for $K \in \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$ and $\mathcal{B}_m = \{\psi_1^m, \dots, \psi_m^m\}$ for every $m = 1, \dots, n$.

In this case, $\widehat{K} = K_{\widehat{m}(\ell)}$ where

$$K_m(x', x) := \sum_{j=1}^m \psi_j^m(x') \psi_j^m(x) ; \forall x, x' \in \mathbb{R}, \forall m \in \mathcal{M} = \{1, \dots, m_{\max}\},$$

$\widehat{m}(\ell)$ is a solution of the minimization problem

$$\min_{m \in \mathcal{M}} \{ \|\widehat{s}_{K_m, \ell}(n; \cdot) - \widehat{s}_{K_{m_{\max}}, \ell}(n; \cdot)\|_2^2 + \text{pen}(m) \}$$

and

$$\text{pen}(m) := \frac{2}{n^2} \sum_{i=1}^n \langle K_m(\cdot, X_i), K_{m_{\max}}(\cdot, X_i) \rangle_{2\ell} (Y_i)^2 ; \forall m \in \mathcal{M}.$$

For $\ell \in \{1, \text{Id}_{\mathbb{R}}\}$, $n = 250$ and $m_{\max} = 30$, m is selected in \mathcal{M} for two basic densities and two nonlinear regression functions:

- $f = f_1$ the density of $\mathcal{E}(5)$.
- $f = f_2$ the density of $\mathcal{N}(1/2, (1/8)^2)$.
- $b(x) = b_1(x) := 10(x^2 - 1/2)$ for every $x \in [0, 1]$.
- $b(x) = b_2(x) := \cos(5\pi x)$ for every $x \in [0, 1]$.

On the one hand, on the four following figures, one can see the beam of all possible estimations of f and bf (i.e. for each $m \in \mathcal{M}$) at left, the PCO criteria for $\widehat{s}_{K,1}(n; \cdot)$ and $\widehat{s}_{K, \text{Id}_{\mathbb{R}}}(n; \cdot)$ for each $m \in \mathcal{M}$ at the middle, and the PCO estimations of f and bf (i.e. for $m = \widehat{m}(1)$ and $m = \widehat{m}(\text{Id}_{\mathbb{R}})$) at right:

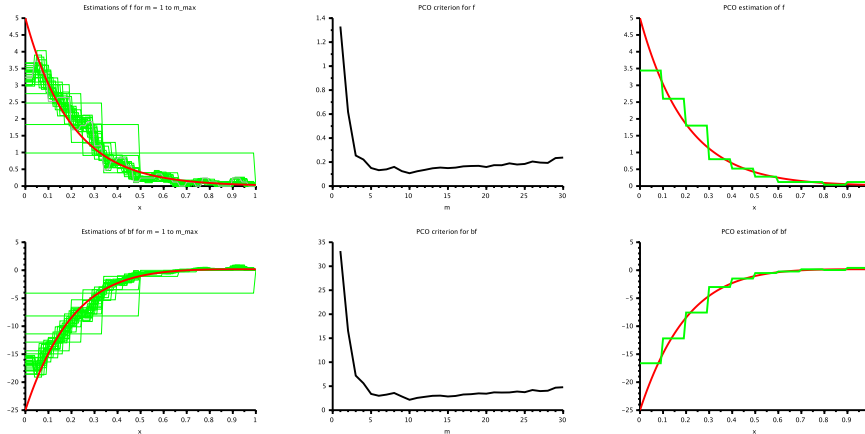


FIGURE 1. $f = f_1$, $b = b_1$, $\widehat{m}(1) = 10$ and $\widehat{m}(\text{Id}_{\mathbb{R}}) = 10$.

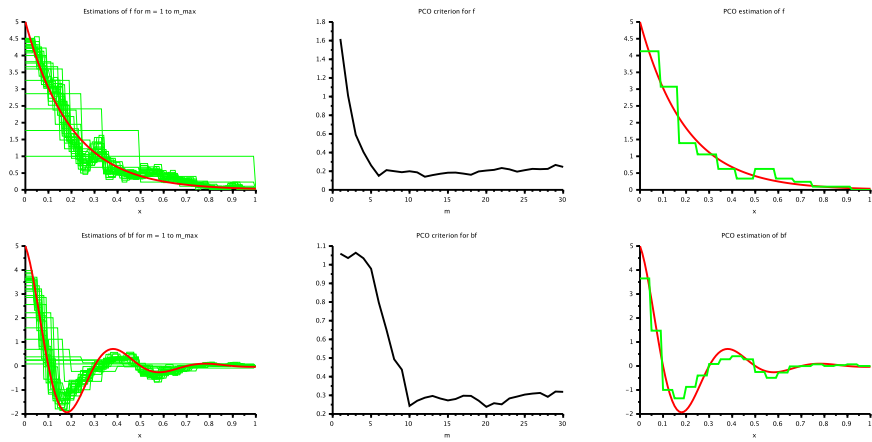


FIGURE 2. $f = f_1$, $b = b_2$, $\hat{m}(1) = 12$ and $\hat{m}(\text{Id}_{\mathbb{R}}) = 20$.

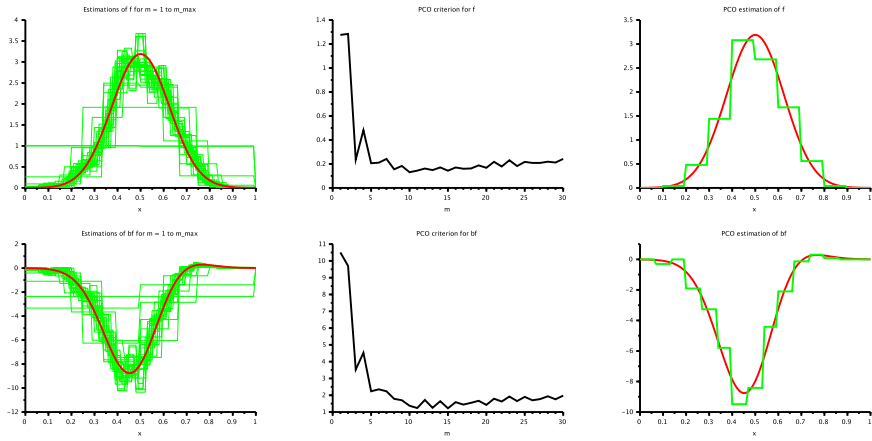


FIGURE 3. $f = f_2$, $b = b_1$, $\hat{m}(1) = 10$ and $\hat{m}(\text{Id}_{\mathbb{R}}) = 15$.

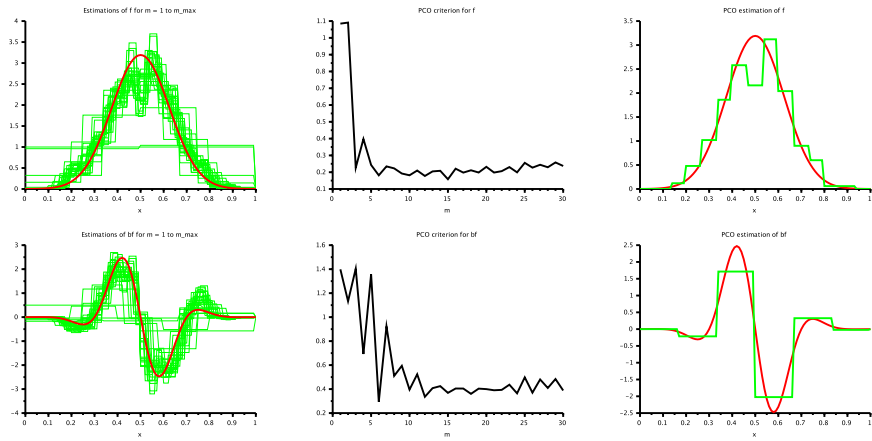
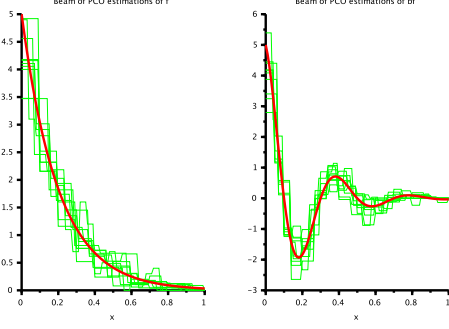
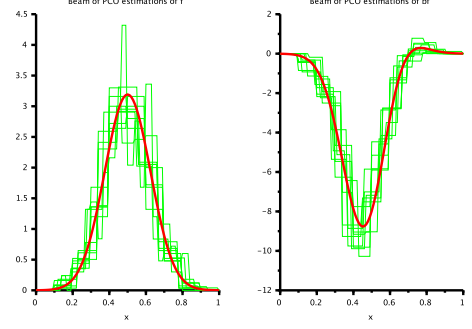


FIGURE 4. $f = f_2$, $b = b_2$, $\hat{m}(1) = 15$ and $\hat{m}(\text{Id}_{\mathbb{R}}) = 6$.

On the other hand, for $(f, b) = (f_1, b_2)$ and $(f, b) = (f_2, b_1)$, let us generate 10 datasets of $n = 250$ observations of (X_1, Y_1) and, for each of these, select $m \in \mathcal{M}$ via the PCO criterion introduced previously. On the two following figures, the beam of all PCO estimations of f (resp. bf) is plotted at left (resp. at right):

FIGURE 5. $f = f_1$ and $b = b_2$.FIGURE 6. $f = f_2$ and $b = b_1$.

APPENDIX A. DETAILS ON KERNELS SETS: PROOFS OF PROPOSITIONS 2.2, 2.3 AND 3.3

A.1. Proof of Proposition 2.2. Consider $K, K' \in \mathcal{K}_k(h_{\min})$. Then, there exist $h, h' \in \{h_{\min}, \dots, 1\}^d$ such that

$$K(x', x) = \prod_{q=1}^d \frac{1}{h_q} k\left(\frac{x'_q - x_q}{h_q}\right) \text{ and } K'(x', x) = \prod_{q=1}^d \frac{1}{h'_q} k\left(\frac{x'_q - x_q}{h'_q}\right)$$

for every $x, x' \in \mathbb{R}^d$.

(1) For every $x' \in \mathbb{R}^d$,

$$\|K(x', \cdot)\|_2^2 = \|k\|_2^{2d} \prod_{q=1}^d \frac{1}{h_q} \leq \|k\|_2^{2d} n.$$

(2) Since $s_{K,\ell} = K * s$, $\|s_{K,\ell}\|_2^2 \leq \|k\|_1^{2d} \|s\|_2^2$.

(3) First,

$$\mathfrak{s}_{K',\ell} = \|k\|_2^{2d} \mathbb{E}(\ell(Y_1)^2) \prod_{q=1}^d \frac{1}{h'_q}.$$

Then,

$$\begin{aligned} \mathbb{E}(\langle K(X_1, \cdot), K'(X_2, \cdot) \ell(Y_2) \rangle_2^2) &= \mathbb{E}((K * K')(X_1 - X_2)^2 \ell(Y_2)^2) \\ &\leq \|f\|_\infty \|K * K'\|_2^2 \mathbb{E}(\ell(Y_1)^2) \\ &\leq \|f\|_\infty \|k\|_1^{2d} \mathfrak{s}_{K',\ell}. \end{aligned}$$

(4) For every $\psi \in \mathbb{L}^2(\mathbb{R}^d)$,

$$\begin{aligned} \mathbb{E}(\langle K(X_1, \cdot), \psi \rangle_2^2) &= \mathbb{E}((K * \psi)(X_1)^2) \\ &\leq \|f\|_\infty \|K * \psi\|_2^2 \leq \|f\|_\infty \|k\|_1^{2d} \|\psi\|_2^2. \end{aligned}$$

A.2. Proof of Proposition 2.3. Consider $K, K' \in \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$. Then, there exist $m, m' \in \{1, \dots, m_{\max}\}^d$ such that

$$K(x', x) = \prod_{q=1}^d \sum_{j=1}^{m_q} \varphi_j^{m_q}(x_q) \varphi_j^{m_q}(x'_q) \text{ and } K'(x', x) = \prod_{q=1}^d \sum_{j=1}^{m'_q} \varphi_j^{m'_q}(x_q) \varphi_j^{m'_q}(x'_q)$$

for every $x, x' \in \mathbb{R}^d$.

(1) For every $x' \in \mathbb{R}^d$,

$$\begin{aligned} \|K(x', \cdot)\|_2^2 &= \prod_{q=1}^d \sum_{j, j'=1}^{m_q} \varphi_{j'}^{m_q}(x'_q) \varphi_j^{m_q}(x'_q) \int_{-\infty}^{\infty} \varphi_{j'}^{m_q}(x) \varphi_j^{m_q}(x) dx \\ &= \prod_{q=1}^d \sum_{j=1}^{m_q} \varphi_j^{m_q}(x'_q)^2 \leq \mathbf{m}_{\mathcal{B}}^d \prod_{q=1}^d m_q \leq \mathbf{m}_{\mathcal{B}}^d n. \end{aligned}$$

(2) Since

$$s_{K, \ell}(\cdot) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \langle s, \varphi_{j_1}^{m_1} \otimes \cdots \otimes \varphi_{j_d}^{m_d} \rangle_2 (\varphi_{j_1}^{m_1} \otimes \cdots \otimes \varphi_{j_d}^{m_d})(\cdot),$$

by Pythagore's theorem, $\|s_{K, \ell}\|_2^2 \leq \|s\|_2^2$.

(3) First,

$$s_{K', \ell} = \mathbb{E} \left[\ell(Y_1)^2 \prod_{q=1}^d \sum_{j=1}^{m'_q} \varphi_j^{m'_q}(X_{1,q})^2 \right] \leq \mathbf{m}_{\mathcal{B}}^d \mathbb{E}(\ell(Y_1)^2) \prod_{q=1}^d m'_q.$$

On the one hand, if $\mathcal{B}_1, \dots, \mathcal{B}_n$ satisfy Condition (5), then

$$\begin{aligned} \mathbb{E}(\langle K(X_1, \cdot), K'(X_2, \cdot) \ell(Y_2) \rangle_2^2) &= \int_{\mathbb{R}^d} \mathbb{E} \left[\left(\prod_{q=1}^d \sum_{j=1}^{m_q \wedge m'_q} \varphi_j^{m'_q}(x'_q) \varphi_j^{m_q}(X_{2,q}) \right)^2 \ell(Y_2)^2 \right] f(x') \lambda_d(dx') \\ &\leq \|f\|_{\infty} \mathbb{E} \left[\ell(Y_2)^2 \prod_{q=1}^d \sum_{j, j'=1}^{m_q \wedge m'_q} \varphi_{j'}^{m'_q}(X_{2,q}) \varphi_j^{m_q}(X_{2,q}) \int_{-\infty}^{\infty} \varphi_{j'}^{m'_q}(x') \varphi_j^{m_q}(x') dx' \right] \\ &\leq \|f\|_{\infty} s_{K', \ell}. \end{aligned}$$

On the other hand, if $\mathcal{B}_1, \dots, \mathcal{B}_n$ satisfy Condition (6), then

$$\begin{aligned} \mathbb{E}(\langle K(X_1, \cdot), K'(X_2, \cdot) \ell(Y_2) \rangle_2^2) &\leq \mathbb{E}(\|K(X_1, \cdot)\|_2^2 \|K'(X_2, \cdot)\|_2^2 \ell(Y_2)^2) \\ &= \mathbb{E}(K(X_1, X_1)) \mathbb{E}(\|K'(X_2, \cdot)\|_2^2 \ell(Y_2)^2) \leq \bar{\mathbf{m}}_{\mathcal{B}} s_{K', \ell}. \end{aligned}$$

(4) For every $\psi \in \mathbb{L}^2(\mathbb{R}^d)$,

$$\begin{aligned} \mathbb{E}(\langle K(X_1, \cdot), \psi \rangle_2^2) &= \mathbb{E} \left[\left| \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \langle \psi, \varphi_{j_1}^{m_1} \otimes \cdots \otimes \varphi_{j_d}^{m_d} \rangle_2 (\varphi_{j_1}^{m_1} \otimes \cdots \otimes \varphi_{j_d}^{m_d})(X_1) \right|^2 \right] \\ &\leq \|f\|_{\infty} \left\| \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \langle \psi, \varphi_{j_1}^{m_1} \otimes \cdots \otimes \varphi_{j_d}^{m_d} \rangle_2 (\varphi_{j_1}^{m_1} \otimes \cdots \otimes \varphi_{j_d}^{m_d})(\cdot) \right\|_2^2 \leq \|f\|_{\infty} \|\psi\|_2^2. \end{aligned}$$

A.3. Proof of Proposition 3.3. For the sake of readability, assume that $d = 1$. Consider $K, K' \in \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n}(m_{\max})$. Then, there exist $m, m' \in \{1, \dots, m_{\max}\}$ such that

$$K(x', x) = \sum_{j=1}^m \chi_j(x) \chi_j(x') \text{ and } K'(x', x) = \sum_{j=1}^{m'} \chi_j(x) \chi_j(x') ; \forall x, x' \in \mathbb{R}.$$

First, there exist $\mathbf{m}_1(m, m') \in \{0, \dots, n\}$ and $\mathbf{c}_1 > 0$, not depending on n , K and K' , such that for any $x' \in [0, 1]$,

$$\begin{aligned} |\langle K(x', \cdot), s_{K', \ell} \rangle_2| &= \left| \sum_{j=1}^{m \wedge m'} \mathbb{E}(\ell(Y_1) \chi_j(X_1)) \chi_j(x') \right| \\ &\leq \mathbf{c}_1 + 2 \left| \sum_{j=1}^{\mathbf{m}_1(m, m')} \mathbb{E}(\ell(Y_1) (\cos(2\pi j X_1) \cos(2\pi j x') + \sin(2\pi j X_1) \sin(2\pi j x'))) \mathbf{1}_{[0,1]}(X_1) \right| \\ &= \mathbf{c}_1 + 2 \left| \sum_{j=1}^{\mathbf{m}_1(m, m')} \mathbb{E}(\ell(Y_1) \cos(2\pi j(X_1 - x'))) \mathbf{1}_{[0,1]}(X_1) \right|. \end{aligned}$$

Moreover, for any $j \in \{2, \dots, \mathbf{m}_1(m, m')\}$,

$$\begin{aligned} \mathbb{E}(\ell(Y_1) \cos(2\pi j(X_1 - x'))) \mathbf{1}_{[0,1]}(X_1) &= \int_0^1 \cos(2\pi j(x - x')) s(x) dx \\ &= \frac{1}{j} \left[\frac{\sin(2\pi j(x - x'))}{2\pi} s(x) \right]_0^1 \\ &\quad + \frac{1}{j^2} \left[\frac{\cos(2\pi j(x - x'))}{4\pi^2} s'(x) \right]_0^1 - \frac{1}{j^2} \int_0^1 \frac{\cos(2\pi j(x - x'))}{4\pi^2} s''(x) dx \\ &= \frac{s(0) - s(1)}{2\pi} \cdot \frac{\alpha_j(x')}{j} + \frac{\beta_j(x')}{j^2} \end{aligned}$$

where $\alpha_j(x') := \sin(2\pi j x')$ and

$$\beta_j(x') := \frac{1}{4\pi^2} \left((s'(1) - s'(0)) \cos(2\pi j x') - \int_0^1 \cos(2\pi j(x - x')) s''(x) dx \right).$$

Then, there exists a deterministic constant $\mathbf{c}_2 > 0$, not depending on n , K , K' and x' , such that

$$(8) \quad \langle K(x', \cdot), s_{K', \ell} \rangle_2^2 \leq \mathbf{c}_2 \left[1 + \left(\sum_{j=1}^{\mathbf{m}_1(m, m')} \frac{\alpha_j(x')}{j} \right)^2 + \left(\sum_{j=1}^{\mathbf{m}_1(m, m')} \frac{\beta_j(x')}{j^2} \right)^2 \right].$$

Let us show that each term of the right-hand side of Inequality (8) are uniformly bounded in x' , m and m' . On the one hand,

$$\left| \sum_{j=1}^{\mathbf{m}_1(m, m')} \frac{\beta_j(x')}{j^2} \right| \leq \max_{j \in \{1, \dots, n\}} \|\beta_j\|_\infty \sum_{j=1}^n \frac{1}{j^2} \leq \frac{1}{24} (2\|s'\|_\infty + \|s''\|_\infty).$$

On the other hand, for every $x \in]0, \pi[$ such that $[\pi/x] + 1 \leq \mathbf{m}_1(m, m')$ (without loss of generality),

$$(9) \quad \begin{aligned} \left| \sum_{j=1}^{\mathbf{m}_1(m, m')} \frac{\sin(jx)}{j} \right| &\leq \left| \sum_{j=1}^{[\pi/x]} \frac{\sin(jx)}{j} \right| + \left| \sum_{j=[\pi/x]+1}^{\mathbf{m}_1(m, m')} \frac{\sin(jx)}{j} \right| \\ &\leq x \left[\frac{\pi}{x} \right] + \frac{2}{(1 + [\pi/x]) \sin(x/2)} \leq \pi + 2. \end{aligned}$$

Since $x \mapsto \sin(x)$ is continuous, odd and 2π -periodic, Inequality (9) holds true for every $x \in \mathbb{R}$. So,

$$\left| \sum_{j=1}^{\mathbf{m}_1(m, m')} \frac{\alpha_j(x')}{j} \right| \leq \pi + 2.$$

Therefore,

$$\mathbb{E} \left[\sup_{K, K' \in \mathcal{K}_{\mathcal{B}_1, \dots, \mathcal{B}_n(m_{\max})}} \langle K(X_1, \cdot), s_{K', \ell} \rangle_2^2 \right] \leq \mathbf{c}_2 \left(1 + (\pi + 2)^2 + \frac{1}{24^2} (2\|s'\|_\infty + \|s''\|_\infty)^2 \right).$$

APPENDIX B. PROOFS OF RISK BOUNDS

In this section, the proofs follow the same pattern as in Comte and Marie [2, 3].

B.1. Preliminary results. This subsection provides three lemmas used several times in the sequel.

Lemma B.1. *Consider*

$$U_{K, K', \ell}(n) := \sum_{i \neq j} \langle K(X_i, \cdot) \ell(Y_i) - s_{K, \ell}, K'(X_j, \cdot) \ell(Y_j) - s_{K', \ell} \rangle_2; \forall K, K' \in \mathcal{K}_n.$$

Under Assumption 2.1.(1,2,3), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathbf{c}_{B.1} > 0$, not depending on n , such that for every $\theta \in]0, 1[$,

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \left\{ \frac{|U_{K, K', \ell}(n)|}{n^2} - \frac{\theta}{n} \mathfrak{s}_{K', \ell} \right\} \right) \leq \mathbf{c}_{B.1} \frac{\log(n)^5}{\theta n}.$$

Lemma B.2. *Consider*

$$V_{K, \ell}(n) := \frac{1}{n} \sum_{i=1}^n \|K(X_i, \cdot) \ell(Y_i) - s_{K, \ell}\|_2^2; \forall K \in \mathcal{K}_n.$$

Under Assumption 2.1.(1,2), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathbf{c}_{B.2} > 0$, not depending on n , such that for every $\theta \in]0, 1[$,

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \frac{1}{n} |V_{K, \ell}(n) - \mathfrak{s}_{K, \ell}| - \frac{\theta}{n} \mathfrak{s}_{K, \ell} \right\} \right) \leq \mathbf{c}_{B.2} \frac{\log(n)^3}{\theta n}.$$

Lemma B.3. *Consider*

$$W_{K, K', \ell}(n) := \langle \widehat{s}_{K, \ell}(n; \cdot) - s_{K, \ell}, s_{K', \ell} - s \rangle_2; \forall K, K' \in \mathcal{K}_n.$$

Under Assumption 2.1.(1,2,4), if $s \in \mathbb{L}^2(\mathbb{R}^d)$ and if there exists $\alpha > 0$ such that $\mathbb{E}(\exp(\alpha|\ell(Y_1)|)) < \infty$, then there exists a deterministic constant $\mathbf{c}_{B.3} > 0$, not depending on n , such that for every $\theta \in]0, 1[$,

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \{ |W_{K, K', \ell}(n)| - \theta \|s_{K', \ell} - s\|_2^2 \} \right) \leq \mathbf{c}_{B.3} \frac{\log(n)^4}{\theta n}.$$

B.1.1. Proof of Lemma B.1. Consider $\mathfrak{m}(n) := 8 \log(n)/\alpha$. For any $K, K' \in \mathcal{K}_n$,

$$U_{K, K', \ell}(n) = U_{K, K', \ell}^1(n) + U_{K, K', \ell}^2(n) + U_{K, K', \ell}^3(n) + U_{K, K', \ell}^4(n)$$

where

$$U_{K, K', \ell}^l(n) := \sum_{i \neq j} g_{K, K', \ell}^l(n; X_i, Y_i, X_j, Y_j); \quad l = 1, 2, 3, 4$$

with, for every $(x', y), (x'', y') \in E = \mathbb{R}^d \times \mathbb{R}$,

$$\begin{aligned} g_{K, K', \ell}^1(n; x', y, x'', y') &:= \langle K(x', \cdot) \ell(y) \mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)} - s_{K, \ell}^+(n; \cdot), K'(x'', \cdot) \ell(y') \mathbf{1}_{|\ell(y')| \leq \mathfrak{m}(n)} - s_{K', \ell}^+(n; \cdot) \rangle_2, \\ g_{K, K', \ell}^2(n; x', y, x'', y') &:= \langle K(x', \cdot) \ell(y) \mathbf{1}_{|\ell(y)| > \mathfrak{m}(n)} - s_{K, \ell}^-(n; \cdot), K'(x'', \cdot) \ell(y') \mathbf{1}_{|\ell(y')| \leq \mathfrak{m}(n)} - s_{K', \ell}^+(n; \cdot) \rangle_2, \\ g_{K, K', \ell}^3(n; x', y, x'', y') &:= \langle K(x', \cdot) \ell(y) \mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)} - s_{K, \ell}^+(n; \cdot), K'(x'', \cdot) \ell(y') \mathbf{1}_{|\ell(y')| > \mathfrak{m}(n)} - s_{K', \ell}^-(n; \cdot) \rangle_2, \\ g_{K, K', \ell}^4(n; x', y, x'', y') &:= \langle K(x', \cdot) \ell(y) \mathbf{1}_{|\ell(y)| > \mathfrak{m}(n)} - s_{K, \ell}^-(n; \cdot), K'(x'', \cdot) \ell(y') \mathbf{1}_{|\ell(y')| > \mathfrak{m}(n)} - s_{K', \ell}^-(n; \cdot) \rangle_2 \end{aligned}$$

and, for every $k \in \mathcal{K}_n$,

$$s_{k, \ell}^+(n; \cdot) := \mathbb{E}(k(X_1, \cdot) \ell(Y_1) \mathbf{1}_{|\ell(Y_1)| \leq \mathfrak{m}(n)}) \quad \text{and} \quad s_{k, \ell}^-(n; \cdot) := \mathbb{E}(k(X_1, \cdot) \ell(Y_1) \mathbf{1}_{|\ell(Y_1)| > \mathfrak{m}(n)}).$$

On the one hand, since $\mathbb{E}(g_{K,K',\ell}^1(n; x', y, X_1, Y_1)) = 0$ for every $(x', y) \in E$, by Giné and Nickl [7], Theorem 3.4.8, there exists a universal constant $\mathfrak{m} \geq 1$ such that for any $\lambda > 0$, with probability larger than $1 - 5.4e^{-\lambda}$,

$$\frac{|U_{K,K',\ell}^1(n)|}{n^2} \leq \frac{\mathfrak{m}}{n^2} (\mathfrak{c}_{K,K',\ell}(n)\lambda^{1/2} + \mathfrak{d}_{K,K',\ell}(n)\lambda + \mathfrak{b}_{K,K',\ell}(n)\lambda^{3/2} + \mathfrak{a}_{K,K',\ell}(n)\lambda^2)$$

where the constants $\mathfrak{a}_{K,K',\ell}(n)$, $\mathfrak{b}_{K,K',\ell}(n)$, $\mathfrak{c}_{K,K',\ell}(n)$ and $\mathfrak{d}_{K,K',\ell}(n)$ are defined and controlled later. First, note that

$$(10) \quad U_{K,K',\ell}^1(n) = \sum_{i \neq j} (\varphi_{K,K',\ell}(n; X_i, Y_i, X_j, Y_j) - \psi_{K,K',\ell}(n; X_i, Y_i) - \psi_{K',K,\ell}(n; X_j, Y_j) + \mathbb{E}(\varphi_{K,K',\ell}(n; X_i, Y_i, X_j, Y_j))),$$

where

$$\varphi_{K,K',\ell}(n; x', y, x'', y'') := \langle K(x', \cdot)\ell(y)\mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)}, K'(x'', \cdot)\ell(y'')\mathbf{1}_{|\ell(y'')| \leq \mathfrak{m}(n)} \rangle_2$$

and

$$\psi_{k,k',\ell}(n; x', y) := \langle k(x', \cdot)\ell(y)\mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)}, s_{k',\ell}^+(n; \cdot) \rangle_2 = \mathbb{E}(\varphi_{k,k',\ell}(n; x', y, X_1, Y_1))$$

for every $k, k' \in \mathcal{K}_n$ and $(x', y), (x'', y') \in E$. Let us now control $\mathfrak{a}_{K,K',\ell}(n)$, $\mathfrak{b}_{K,K',\ell}(n)$, $\mathfrak{c}_{K,K',\ell}(n)$ and $\mathfrak{d}_{K,K',\ell}(n)$:

- **The constant $\mathfrak{a}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{a}_{K,K',\ell}(n) := \sup_{(x',y),(x'',y') \in E} |g_{K,K',\ell}^1(n; x', y, x'', y')|.$$

By (10), Cauchy-Schwarz's inequality and Assumption 2.1.(1),

$$\begin{aligned} \mathfrak{a}_{K,K',\ell}(n) &\leq 4 \sup_{(x',y),(x'',y') \in E} |\langle K(x', \cdot)\ell(y)\mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)}, K'(x'', \cdot)\ell(y'')\mathbf{1}_{|\ell(y'')| \leq \mathfrak{m}(n)} \rangle_2| \\ &\leq 4\mathfrak{m}(n)^2 \left(\sup_{x' \in \mathbb{R}^d} \|K(x', \cdot)\|_2 \right) \left(\sup_{x'' \in \mathbb{R}^d} \|K'(x'', \cdot)\|_2 \right) \leq 4\mathfrak{m}_{\mathcal{K},\ell} \mathfrak{m}(n)^2 n. \end{aligned}$$

So,

$$\frac{1}{n^2} \mathfrak{a}_{K,K',\ell}(n) \lambda^2 \leq \frac{4}{n} \mathfrak{m}_{\mathcal{K},\ell} \mathfrak{m}(n)^2 \lambda^2.$$

- **The constant $\mathfrak{b}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{b}_{K,K',\ell}(n)^2 := n \sup_{(x',y) \in E} \mathbb{E}(g_{K,K',\ell}^1(n; x', y, X_1, Y_1)^2).$$

By (10), Jensen's inequality, Cauchy-Schwarz's inequality and Assumption 2.1.(1),

$$\begin{aligned} \mathfrak{b}_{K,K',\ell}(n)^2 &\leq 16n \sup_{(x',y) \in E} \mathbb{E}(\langle K(x', \cdot)\ell(y)\mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)}, K'(X_1, \cdot)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)| \leq \mathfrak{m}(n)} \rangle_2^2) \\ &\leq 16n\mathfrak{m}(n)^2 \sup_{x' \in \mathbb{R}^d} \|K(x', \cdot)\|_2^2 \mathbb{E}(\|K'(X_1, \cdot)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)| \leq \mathfrak{m}(n)}\|_2^2) \leq 16\mathfrak{m}_{\mathcal{K},\ell} n^2 \mathfrak{m}(n)^2 \mathfrak{s}_{K',\ell}. \end{aligned}$$

So, for any $\theta \in]0, 1[$,

$$\begin{aligned} \frac{1}{n^2} \mathfrak{b}_{K,K',\ell}(n) \lambda^{3/2} &\leq 2 \left(\frac{3\mathfrak{m}}{\theta} \right)^{1/2} \frac{2}{n^{1/2}} \mathfrak{m}_{\mathcal{K},\ell}^{1/2} \mathfrak{m}(n) \lambda^{3/2} \times \left(\frac{\theta}{3\mathfrak{m}} \right)^{1/2} \frac{1}{n^{1/2}} \mathfrak{s}_{K',\ell}^{1/2} \\ &\leq \frac{\theta}{3\mathfrak{m}n} \mathfrak{s}_{K',\ell} + \frac{12\mathfrak{m}\lambda^3}{\theta n} \mathfrak{m}_{\mathcal{K},\ell} \mathfrak{m}(n)^2. \end{aligned}$$

- **The constant $\mathfrak{c}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{c}_{K,K',\ell}(n)^2 := n^2 \mathbb{E}(g_{K,K',\ell}^1(n; X_1, Y_1, X_2, Y_2)^2).$$

By (10), Jensen's inequality and Assumption 2.1.(3),

$$\begin{aligned} \mathfrak{c}_{K,K',\ell}(n)^2 &\leq 16n^2 \mathbb{E}(\langle K(X_1, \cdot)\ell(Y_1)\mathbf{1}_{|\ell(Y_1)| \leq \mathfrak{m}(n)}, K'(X_2, \cdot)\ell(Y_2)\mathbf{1}_{|\ell(Y_2)| \leq \mathfrak{m}(n)} \rangle_2^2) \\ &\leq 16n^2 \mathfrak{m}(n)^2 \mathbb{E}(\langle K(X_1, \cdot), K'(X_2, \cdot)\ell(Y_2) \rangle_2^2) \leq 16\mathfrak{m}_{\mathcal{K},\ell} n^2 \mathfrak{m}(n)^2 \mathfrak{s}_{K',\ell}. \end{aligned}$$

So,

$$\frac{1}{n^2} \mathbf{c}_{K,K',\ell}(n) \lambda^{1/2} \leq \frac{\theta}{3mn} \mathfrak{s}_{K',\ell} + \frac{12m\lambda}{\theta n} \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2.$$

- **The constant $\mathfrak{d}_{K,K',\ell}(n)$.** Consider

$$\mathfrak{d}_{K,K',\ell}(n) := \sup_{(a,b) \in \mathcal{A}} \mathbb{E} \left[\sum_{i < j} a_i(X_i, Y_i) b_j(X_j, Y_j) g_{K,K',\ell}^1(n; X_i, Y_i, X_j, Y_j) \right],$$

where

$$\mathcal{A} := \left\{ (a, b) : \sum_{i=1}^{n-1} \mathbb{E}(a_i(X_i, Y_i)^2) \leq 1 \text{ and } \sum_{j=2}^n \mathbb{E}(b_j(X_j, Y_j)^2) \leq 1 \right\}.$$

By (10), Jensen's inequality, Cauchy-Schwarz's inequality and Assumption 2.1.(3),

$$\begin{aligned} \mathfrak{d}_{K,K',\ell}(n) &\leq 4 \sup_{(a,b) \in \mathcal{A}} \mathbb{E} \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n |a_i(X_i, Y_i) b_j(X_j, Y_j) \varphi_{K,K',\ell}(n; X_i, Y_i, X_j, Y_j)| \right] \\ &\leq 4nm(n) \mathbb{E}(\langle K(X_1, \cdot), K'(X_2, \cdot) \ell(Y_2) \rangle_2^2)^{1/2} \leq 4m_{\mathcal{K},\ell}^{1/2} nm(n) \mathfrak{s}_{K',\ell}^{1/2}. \end{aligned}$$

So,

$$\frac{1}{n^2} \mathfrak{d}_{K,K',\ell}(n) \lambda \leq \frac{\theta}{3mn} \mathfrak{s}_{K',\ell} + \frac{12m\lambda^2}{\theta n} \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2.$$

Then, since $m \geq 1$ and $\lambda > 0$, with probability larger than $1 - 5.4e^{-\lambda}$,

$$\frac{|U_{K,K',\ell}^1(n)|}{n^2} \leq \frac{\theta}{n} \mathfrak{s}_{K',\ell} + \frac{40m^2}{\theta n} \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2 (1 + \lambda)^3.$$

So, with probability larger than $1 - 5.4|\mathcal{K}_n|e^{-\lambda}$,

$$S_{\mathcal{K},\ell}(n, \theta) := \sup_{K, K' \in \mathcal{K}_n} \left\{ \frac{|U_{K,K',\ell}^1(n)|}{n^2} - \frac{\theta}{n} \mathfrak{s}_{K',\ell} \right\} \leq \frac{40m^2}{\theta n} \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2 (1 + \lambda)^3.$$

For every $t \in \mathbb{R}_+$, consider

$$\lambda_{\mathcal{K},\ell}(n, \theta, t) := -1 + \left(\frac{t}{\mathbf{m}_{\mathcal{K},\ell}(n, \theta)} \right)^{1/3} \text{ with } \mathbf{m}_{\mathcal{K},\ell}(n, \theta) = \frac{40m^2}{\theta n} \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2.$$

Then, for any $T > 0$,

$$\begin{aligned} \mathbb{E}(S_{\mathcal{K},\ell}(n, \theta)) &\leq T + \int_T^\infty \mathbb{P}(S_{\mathcal{K},\ell}(n, \theta) \geq (1 + \lambda_{\mathcal{K},\ell}(n, \theta, t))^3 \mathbf{m}_{\mathcal{K},\ell}(n, \theta)) dt \\ &\leq 2T + 5.4\mathbf{c}_1 |\mathcal{K}_n| \mathbf{m}_{\mathcal{K},\ell}(n, \theta) \exp\left(-\frac{T^{1/3}}{2\mathbf{m}_{\mathcal{K},\ell}(n, \theta)^{1/3}}\right) \text{ with } \mathbf{c}_1 = \int_0^\infty e^{1-r^{1/3}/2} dr. \end{aligned}$$

Moreover,

$$\mathbf{m}_{\mathcal{K},\ell}(n, \theta) \leq \mathbf{c}_2 \frac{\log(n)^2}{\theta n} \text{ with } \mathbf{c}_2 = \frac{40 \cdot 8^2 m^2}{\alpha^2} \mathbf{m}_{\mathcal{K},\ell}.$$

So, by taking

$$T = 2^3 \mathbf{c}_2 \frac{\log(n)^5}{\theta n},$$

and since $|\mathcal{K}_n| \leq n$,

$$\mathbb{E}(S_{\mathcal{K},\ell}(n, \theta)) \leq 2^4 \mathbf{c}_2 \frac{\log(n)^5}{\theta n} + 5.4\mathbf{c}_1 \mathbf{m}_{\mathcal{K},\ell}(n, \theta) \frac{|\mathcal{K}_n|}{n} \leq (2^4 + 5.4\mathbf{c}_1) \mathbf{c}_2 \frac{\log(n)^5}{\theta n}.$$

On the other hand, by Assumption 2.1.(1), Cauchy-Schwarz's inequality and Markov's inequality,

$$\begin{aligned} \mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} |g_{K, K', \ell}^2(n; X_1, Y_1, X_2, Y_2)| \right) &\leq 4\mathfrak{m}(n) \sum_{K, K' \in \mathcal{K}_n} \mathbb{E}(|\ell(Y_1)| \mathbf{1}_{|\ell(Y_1)| > \mathfrak{m}(n)} | \langle K(X_1, \cdot), K'(X_2, \cdot) \rangle_2 |) \\ &\leq 4\mathfrak{m}(n) \mathfrak{m}_{\mathcal{K}, \ell} n |\mathcal{K}_n|^2 \mathbb{E}(\ell(Y_1)^2)^{1/2} \mathbb{P}(|\ell(Y_1)| > \mathfrak{m}(n))^{1/2} \leq \mathfrak{c}_3 \frac{\log(n)}{n} \end{aligned}$$

with

$$\mathfrak{c}_3 = \frac{32}{\alpha} \mathfrak{m}_{\mathcal{K}, \ell} \mathbb{E}(\ell(Y_1)^2)^{1/2} \mathbb{E}(\exp(\alpha|\ell(Y_1)|))^{1/2}.$$

So,

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \frac{|U_{K, K', \ell}^2(n)|}{n^2} \right) \leq \mathfrak{c}_3 \frac{\log(n)}{n}$$

and, symmetrically,

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \frac{|U_{K, K', \ell}^3(n)|}{n^2} \right) \leq \mathfrak{c}_3 \frac{\log(n)}{n}.$$

By Assumption 2.1.(1), Cauchy-Schwarz's inequality and Markov's inequality,

$$\begin{aligned} \mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} |g_{K, K', \ell}^4(n; X_1, Y_1, X_2, Y_2)| \right) &\leq 4 \sum_{K, K' \in \mathcal{K}_n} \mathbb{E}(|\ell(Y_1)\ell(Y_2)| \mathbf{1}_{|\ell(Y_1)|, |\ell(Y_2)| > \mathfrak{m}(n)} | \langle K(X_1, \cdot), K'(X_2, \cdot) \rangle_2 |) \\ &\leq 4\mathfrak{m}_{\mathcal{K}, \ell} n |\mathcal{K}_n|^2 \mathbb{E}(\ell(Y_1)^2) \mathbb{P}(|\ell(Y_1)| > \mathfrak{m}(n)) \leq \frac{\mathfrak{c}_4}{n^5} \end{aligned}$$

with

$$\mathfrak{c}_4 = 4\mathfrak{m}_{\mathcal{K}, \ell} \mathbb{E}(\ell(Y_1)^2) \mathbb{E}(\exp(\alpha|\ell(Y_1)|)).$$

So,

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \frac{|U_{K, K', \ell}^4(n)|}{n^2} \right) \leq \frac{\mathfrak{c}_4}{n^5}.$$

Therefore,

$$\mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \left\{ \frac{|U_{K, K', \ell}(n)|}{n^2} - \frac{\theta}{n} \mathfrak{s}_{K', \ell} \right\} \right) \leq (2^4 + 5.4\mathfrak{c}_1) \mathfrak{c}_2 \frac{\log(n)^5}{\theta n} + 2\mathfrak{c}_3 \frac{\log(n)}{n} + \frac{\mathfrak{c}_4}{n^5}.$$

B.1.2. *Proof of Lemma B.2.* First, the two following results are used several times in the sequel:

$$\begin{aligned} \|s_{K, \ell}\|_2^2 &\leq \mathbb{E}(\ell(Y_1)^2) \int_{\mathbb{R}^d} f(x') \int_{\mathbb{R}^d} K(x', x)^2 \lambda_d(dx) \lambda_d(dx') \\ (11) \quad &\leq \mathbb{E}(\ell(Y_1)^2) \mathfrak{m}_{\mathcal{K}, \ell} n \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(V_{K, \ell}(n)) &= \mathbb{E}(\|K(X_1, \cdot)\ell(Y_1) - s_{K, \ell}\|_2^2) \\ (12) \quad &= \mathbb{E}(\|K(X_1, \cdot)\ell(Y_1)\|_2^2) + \|s_{K, \ell}\|_2^2 - 2 \int_{\mathbb{R}^d} s_{K, \ell}(x) \mathbb{E}(K(X_1, x)\ell(Y_1)) \lambda_d(dx) = \mathfrak{s}_{K, \ell} - \|s_{K, \ell}\|_2^2. \end{aligned}$$

Consider $\mathfrak{m}(n) := 2 \log(n)/\alpha$ and

$$v_{K, \ell}(n) := V_{K, \ell}(n) - \mathbb{E}(V_{K, \ell}(n)) = v_{K, \ell}^1(n) + v_{K, \ell}^2(n),$$

where

$$v_{K, \ell}^j(n) = \frac{1}{n} \sum_{i=1}^n (g_{K, \ell}^j(n; X_i, Y_i) - \mathbb{E}(g_{K, \ell}^j(n; X_i, Y_i))) ; j = 1, 2$$

with, for every $(x', y) \in E$,

$$g_{K, \ell}^1(n; x', y) := \|K(x', \cdot)\ell(y) - s_{K, \ell}\|_2^2 \mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)}$$

and

$$g_{K,\ell}^2(n; x', y) := \|K(x', \cdot)\ell(y) - s_{K,\ell}\|_2^2 \mathbf{1}_{|\ell(y)| > \mathbf{m}(n)}.$$

On the one hand, by Bernstein's inequality, for any $\lambda > 0$, with probability larger than $1 - 2e^{-\lambda}$,

$$|v_{K,\ell}^1(n)| \leq \sqrt{\frac{2\lambda}{n} \mathbf{v}_{K,\ell}(n)} + \frac{\lambda}{n} \mathbf{c}_{K,\ell}(n)$$

where

$$\mathbf{c}_{K,\ell}(n) = \frac{\|g_{K,\ell}^1(n; \cdot)\|_\infty}{3} \text{ and } \mathbf{v}_{K,\ell}(n) = \mathbb{E}(g_{K,\ell}^1(n; X_1, Y_1)^2).$$

Moreover,

$$\begin{aligned} \mathbf{c}_{K,\ell}(n) &= \frac{1}{3} \sup_{(x', y) \in E} \|K(x', \cdot)\ell(y) - s_{K,\ell}\|_2^2 \mathbf{1}_{|\ell(y)| \leq \mathbf{m}(n)} \\ &\leq \frac{2}{3} \left(\mathbf{m}(n)^2 \sup_{x' \in \mathbb{R}^d} \|K(x', \cdot)\|_2^2 + \|s_{K,\ell}\|_2^2 \right) \leq \frac{2}{3} (\mathbf{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} n \end{aligned}$$

by Inequality (11), and

$$\begin{aligned} \mathbf{v}_{K,\ell}(n) &\leq \|g_{K,\ell}^1(n; \cdot)\|_\infty \mathbb{E}(V_{K,\ell}(n)) \\ &\leq 2(\mathbf{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} n (\mathbf{s}_{K,\ell} - \|s_{K,\ell}\|_2^2) \end{aligned}$$

by Inequality (11) and Equality (12). Then, for any $\theta \in]0, 1[$,

$$\begin{aligned} |v_{K,\ell}^1(n)| &\leq 2\sqrt{\lambda(\mathbf{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} (\mathbf{s}_{K,\ell} - \|s_{K,\ell}\|_2^2)} + \frac{2\lambda}{3} (\mathbf{m}(n)^2 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} \\ &\leq \theta \mathbf{s}_{K,\ell} + \frac{5\lambda}{3\theta} (1 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2 \end{aligned}$$

with probability larger than $1 - 2e^{-\lambda}$. So, with probability larger than $1 - 2|\mathcal{K}_n|e^{-\lambda}$,

$$S_{\mathcal{K},\ell}(n, \theta) := \sup_{K \in \mathcal{K}_n} \left\{ \frac{|v_{K,\ell}^1(n)|}{n} - \frac{\theta}{n} \mathbf{s}_{K,\ell} \right\} \leq \frac{5\lambda}{3\theta n} (1 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2.$$

For every $t \in \mathbb{R}_+$, consider

$$\lambda_{\mathcal{K},\ell}(n, \theta, t) := \frac{t}{\mathbf{m}_{\mathcal{K},\ell}(n, \theta)} \text{ with } \mathbf{m}_{\mathcal{K},\ell}(n, \theta) = \frac{5}{3\theta n} (1 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2.$$

Then, for any $T > 0$,

$$\begin{aligned} \mathbb{E}(S_{\mathcal{K},\ell}(n, \theta)) &\leq T + \int_T^\infty \mathbb{P}(S_{\mathcal{K},\ell}(n, \theta) \geq \lambda_{\mathcal{K},\ell}(n, \theta, t) \mathbf{m}_{\mathcal{K},\ell}(n, \theta)) dt \\ &\leq 2T + 2\mathbf{c}_1 |\mathcal{K}_n| \mathbf{m}_{\mathcal{K},\ell}(n, \theta) \exp\left(-\frac{T}{2\mathbf{m}_{\mathcal{K},\ell}(n, \theta)}\right) \text{ with } \mathbf{c}_1 = \int_0^\infty e^{-r/2} dr = 2. \end{aligned}$$

Moreover,

$$\mathbf{m}_{\mathcal{K},\ell}(n, \theta) \leq \mathbf{c}_2 \frac{\log(n)^2}{\theta n} \text{ with } \mathbf{c}_2 = \frac{10}{3\alpha^2} (1 + \mathbb{E}(\ell(Y_1)^2)) \mathbf{m}_{\mathcal{K},\ell}.$$

So, by taking

$$T = 2\mathbf{c}_2 \frac{\log(n)^3}{\theta n},$$

and since $|\mathcal{K}_n| \leq n$,

$$\mathbb{E}(S_{\mathcal{K},\ell}(n, \theta)) \leq 4\mathbf{c}_2 \frac{\log(n)^3}{\theta n} + 4\mathbf{m}_{\mathcal{K},\ell}(n, \theta) \frac{|\mathcal{K}_n|}{n} \leq 8\mathbf{c}_2 \frac{\log(n)^3}{\theta n}.$$

On the other hand, by Inequality (11) and Markov's inequality,

$$\begin{aligned} \mathbb{E} \left[\sup_{K \in \mathcal{K}_n} \frac{|v_{K,\ell}^2(n)|}{n} \right] &\leq \frac{2}{n} \mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \|K(X_1, \cdot)\ell(Y_1) - s_{K,\ell}\|_2^2 \mathbf{1}_{|\ell(Y_1)| > \mathfrak{m}(n)} \right) \\ &\leq \frac{4}{n} \mathbb{E} \left[\left[\ell(Y_1)^2 \sup_{K \in \mathcal{K}_n} \|K(X_1, \cdot)\|_2^2 + \sup_{K \in \mathcal{K}_n} \|s_{K,\ell}\|_2^2 \right]^2 \right]^{1/2} \mathbb{P}(|\ell(Y_1)| > \mathfrak{m}(n))^{1/2} \leq \frac{\mathfrak{c}_3}{n} \end{aligned}$$

with

$$\mathfrak{c}_3 = 8\mathfrak{m}_{\mathcal{K},\ell} \mathbb{E}(\ell(Y_1)^4)^{1/2} \mathbb{E}(\exp(\alpha|\ell(Y_1)|))^{1/2}.$$

Therefore,

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \frac{|v_{K,\ell}(n)|}{n} - \frac{\theta}{n} \mathfrak{s}_{K,\ell} \right\} \right) \leq 8\mathfrak{c}_2 \frac{\log(n)^3}{\theta n} + \frac{\mathfrak{c}_3}{n}$$

and, by Equality (12), the definition of $v_{K,\ell}(n)$ and Assumption 2.1.(2),

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \frac{1}{n} |V_{K,\ell}(n) - \mathfrak{s}_{K,\ell}| - \frac{\theta}{n} \mathfrak{s}_{K,\ell} \right\} \right) \leq 8\mathfrak{c}_2 \frac{\log(n)^3}{\theta n} + \frac{\mathfrak{c}_3 + \mathfrak{m}_{\mathcal{K},\ell}}{n}.$$

B.1.3. *Proof of Lemma B.3.* Consider $\mathfrak{m}(n) = 12 \log(n)/\alpha$. For any $K, K' \in \mathcal{K}_n$,

$$W_{K,K',\ell}(n) = W_{K,K',\ell}^1(n) + W_{K,K',\ell}^2(n)$$

where

$$W_{K,K',\ell}^j(n) := \frac{1}{n} \sum_{i=1}^n (g_{K,K',\ell}^j(n; X_i, Y_i) - \mathbb{E}(g_{K,K',\ell}^j(n; X_i, Y_i))) ; j = 1, 2$$

with, for every $(x', y) \in E$,

$$g_{K,K',\ell}^1(n; x', y) := \langle K(x', \cdot)\ell(y), s_{K',\ell} - s \rangle_2 \mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)}$$

and

$$g_{K,K',\ell}^2(n; x', y) := \langle K(x', \cdot)\ell(y), s_{K',\ell} - s \rangle_2 \mathbf{1}_{|\ell(y)| > \mathfrak{m}(n)}.$$

On the one hand, by Bernstein's inequality, for any $\lambda > 0$, with probability larger than $1 - 2e^{-\lambda}$,

$$|W_{K,K',\ell}^1(n)| \leq \sqrt{\frac{2\lambda}{n} \mathfrak{v}_{K,K',\ell}(n)} + \frac{\lambda}{n} \mathfrak{c}_{K,K',\ell}(n)$$

where

$$\mathfrak{c}_{K,K',\ell}(n) = \frac{\|g_{K,K',\ell}^1(n; \cdot)\|_\infty}{3} \text{ and } \mathfrak{v}_{K,K',\ell}(n) = \mathbb{E}(g_{K,K',\ell}^1(n; X_1, Y_1)^2).$$

Moreover,

$$\begin{aligned} \mathfrak{c}_{K,K',\ell}(n) &= \frac{1}{3} \sup_{(x', y) \in E} |\langle K(x', \cdot)\ell(y), s_{K',\ell} - s \rangle_2| \mathbf{1}_{|\ell(y)| \leq \mathfrak{m}(n)} \\ &\leq \frac{1}{3} \mathfrak{m}(n) \|s_{K',\ell} - s\|_2 \sup_{x' \in \mathbb{R}^d} \|K(x', \cdot)\|_2 \leq \frac{1}{3} \mathfrak{m}_{\mathcal{K},\ell}^{1/2} n^{1/2} \mathfrak{m}(n) \|s_{K',\ell} - s\|_2 \end{aligned}$$

by Assumption 2.1.(1), and

$$\mathfrak{v}_{K,\ell}(n) \leq \mathbb{E}(\langle K(X_1, \cdot)\ell(Y_1), s_{K',\ell} - s \rangle_2^2 \mathbf{1}_{|\ell(Y_1)| \leq \mathfrak{m}(n)}) \leq \mathfrak{m}(n)^2 \mathfrak{m}_{\mathcal{K},\ell} \|s_{K',\ell} - s\|_2^2$$

by Assumption 2.1.(4). Then, since $\lambda > 0$, for any $\theta \in]0, 1[$,

$$\begin{aligned} |W_{K,K',\ell}^1(n)| &\leq \sqrt{\frac{2\lambda}{n} \mathfrak{m}(n)^2 \mathfrak{m}_{\mathcal{K},\ell} \|s_{K',\ell} - s\|_2^2} + \frac{\lambda}{3n^{1/2}} \mathfrak{m}_{\mathcal{K},\ell}^{1/2} \mathfrak{m}(n) \|s_{K',\ell} - s\|_2 \\ &\leq \theta \|s_{K',\ell} - s\|_2^2 + \frac{\mathfrak{m}_{\mathcal{K},\ell}}{2\theta n} \mathfrak{m}(n)^2 (1 + \lambda)^2 \end{aligned}$$

with probability larger than $1 - 2e^{-\lambda}$. So, with probability larger than $1 - 2|\mathcal{K}_n|e^{-\lambda}$,

$$S_{\mathcal{K},\ell}(n, \theta) := \sup_{K, K' \in \mathcal{K}_n} \{|W_{K,K',\ell}^1(n)| - \theta \|s_{K',\ell} - s\|_2^2\} \leq \frac{\mathfrak{m}_{\mathcal{K},\ell}}{2\theta n} \mathfrak{m}(n)^2 (1 + \lambda)^2.$$

For every $t \in \mathbb{R}_+$, consider

$$\lambda_{\mathcal{K},\ell}(n, \theta, t) := -1 + \left(\frac{t}{\mathbf{m}_{\mathcal{K},\ell}(n, \theta)} \right)^{1/2} \quad \text{with } \mathbf{m}_{\mathcal{K},\ell}(n, \theta) = \frac{\mathbf{m}_{\mathcal{K},\ell} \mathbf{m}(n)^2}{2\theta n}.$$

Then, for any $T > 0$,

$$\begin{aligned} \mathbb{E}(S_{\mathcal{K},\ell}(n, \theta)) &\leq T + \int_T^\infty \mathbb{P}(S_{\mathcal{K},\ell}(n, \theta) \geq (1 + \lambda_{\mathcal{K},\ell}(n, \theta, t))^2 \mathbf{m}_{\mathcal{K},\ell}(n, \theta)) dt \\ &\leq 2T + 2\mathbf{c}_1 |\mathcal{K}_n| \mathbf{m}_{\mathcal{K},\ell}(n, \theta) \exp\left(-\frac{T^{1/2}}{2\mathbf{m}_{\mathcal{K},\ell}(n, \theta)^{1/2}}\right) \quad \text{with } \mathbf{c}_1 = \int_0^\infty e^{1-r^{1/2}/2} dr. \end{aligned}$$

Moreover,

$$\mathbf{m}_{\mathcal{K},\ell}(n, \theta) \leq \mathbf{c}_2 \frac{\log(n)^2}{\theta n} \quad \text{with } \mathbf{c}_2 = \frac{12^2}{2\alpha^2} \mathbf{m}_{\mathcal{K},\ell}.$$

So, by taking

$$T = 2^2 \mathbf{c}_2 \frac{\log(n)^4}{\theta n},$$

and since $|\mathcal{K}_n| \leq n$,

$$\mathbb{E}(S_{\mathcal{K},\ell}(n, \theta)) \leq 2^3 \mathbf{c}_2 \frac{\log(n)^4}{\theta n} + 2\mathbf{c}_1 \mathbf{m}_{\mathcal{K},\ell}(n, \theta) \frac{|\mathcal{K}_n|}{n} \leq (2^3 + 2\mathbf{c}_1) \mathbf{c}_2 \frac{\log(n)^4}{\theta n}.$$

On the other hand, by Assumption 2.1.(2,4), Cauchy-Schwarz's inequality and Markov's inequality,

$$\begin{aligned} \mathbb{E}\left(\sup_{K, K' \in \mathcal{K}_n} |W_{K, K', \ell}^2(n)|\right) &\leq 2\mathbb{E}(\ell(Y_1)^2 \mathbf{1}_{|\ell(Y_1)| > \mathbf{m}(n)})^{1/2} \sum_{K, K' \in \mathcal{K}_n} \mathbb{E}(\langle K(X_1, \cdot), s_{K', \ell} - s \rangle_2^2)^{1/2} \\ &\leq 2\mathbf{m}_{\mathcal{K},\ell}^{1/2} \|s_{K', \ell} - s\|_2 \mathbb{E}(\ell(Y_1)^4)^{1/4} |\mathcal{K}_n|^2 \mathbb{P}(|\ell(Y_1)| > \mathbf{m}(n))^{1/4} \leq \frac{\mathbf{c}_3}{n} \end{aligned}$$

with

$$\mathbf{c}_3 = 2\mathbf{m}_{\mathcal{K},\ell}^{1/2} (\mathbf{m}_{\mathcal{K},\ell}^{1/2} + \|s\|_2) \mathbb{E}(\ell(Y_1)^4)^{1/4} \mathbb{E}(\exp(\alpha|\ell(Y_1)|))^{1/4}.$$

Therefore,

$$\mathbb{E}\left(\sup_{K, K' \in \mathcal{K}_n} \{|W_{K, K', \ell}(n)| - \theta \|s_{K', \ell} - s\|_2^2\}\right) \leq (2^3 + 2\mathbf{c}_1) \mathbf{c}_2 \frac{\log(n)^4}{\theta n} + \frac{\mathbf{c}_3}{n} \leq \mathbf{c}_4 \frac{\log(n)^4}{\theta n}$$

with $\mathbf{c}_4 = (2^3 + 2\mathbf{c}_1) \mathbf{c}_2 + \mathbf{c}_3$.

B.2. Proof of Proposition 2.4. For any $K \in \mathcal{K}_n$,

$$(13) \quad \|\widehat{s}_{K,\ell}(n; \cdot) - s_{K,\ell}\|_2^2 = \frac{U_{K,\ell}(n)}{n^2} + \frac{V_{K,\ell}(n)}{n}$$

with $U_{K,\ell}(n) = U_{K,K,\ell}(n)$ and $V_{K,\ell}(n) = V_{K,K,\ell}(n)$. Then, by Lemmas B.1 and B.2,

$$\mathbb{E}\left(\sup_{K \in \mathcal{K}_n} \left\{ \|\widehat{s}_{K,\ell}(n; \cdot) - s_{K,\ell}\|_2^2 - \frac{\mathfrak{s}_{K,\ell}}{n} \left| -\frac{\theta}{n} \mathfrak{s}_{K,\ell} \right. \right\}\right) \leq \mathbf{c}_{2.4} \frac{\log(n)^5}{\theta n}$$

with $\mathbf{c}_{2.4} = \mathbf{c}_{B.1} + \mathbf{c}_{B.2}$.

B.3. Proof of Theorem 2.5. On the one hand, for every $K \in \mathcal{K}_n$,

$$\|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 - (1 + \theta) \left(\|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} \right)$$

can be written

$$\|\widehat{s}_{K,\ell}(n; \cdot) - s_{K,\ell}\|_2^2 - (1 + \theta) \frac{\mathfrak{s}_{K,\ell}}{n} + W_{K,K,\ell}(n) - \theta \|s_{K,\ell} - s\|_2^2.$$

Then, by Proposition 2.4 and Lemma B.3,

$$\mathbb{E}\left(\sup_{K \in \mathcal{K}_n} \left\{ \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 - (1 + \theta) \left(\|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} \right) \right\}\right) \leq \mathbf{c}_{2.5} \frac{\log(n)^5}{\theta n}$$

with $\mathbf{c}_{2.5} = \mathbf{c}_{2.4} + \mathbf{c}_{B.3}$. On the other hand, for any $K \in \mathcal{K}_n$,

$$\|s_{K,\ell} - s\|_2^2 = \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 - \|\widehat{s}_{K,\ell}(n; \cdot) - s_{K,\ell}\|_2^2 - W_{K,\ell}(n).$$

Then,

$$(1 - \theta) \left(\|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} \right) - \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 \leq |W_{K,\ell}(n)| - \theta \|s_{K,\ell} - s\|_2^2 + \Lambda_{K,\ell}(n) - \theta \frac{\mathfrak{s}_{K,\ell}}{n}$$

where

$$\Lambda_{K,\ell}(n) := \left| \|\widehat{s}_{K,\ell} - s_{K,\ell}\|_2^2 - \frac{\mathfrak{s}_{K,\ell}}{n} \right|.$$

By Equalities (13) and (12),

$$\Lambda_{K,\ell}(n) = \left| \frac{U_{K,\ell}(n)}{n^2} + \frac{v_{K,\ell}(n)}{n} - \frac{\|s_{K,\ell}\|_2^2}{n} \right|.$$

By Lemmas B.2 and B.1, there exists a deterministic constant $\mathbf{c}_1 > 0$, not depending n and θ , such that

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \Lambda_{K,\ell}(n) - \theta \frac{\mathfrak{s}_{K,\ell}}{n} \right\} \right) \leq \mathbf{c}_1 \frac{\log(n)^5}{\theta n}.$$

By Lemma B.3,

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \{ |W_{K,\ell}(n)| - \theta \|s_{K,\ell} - s\|_2^2 \} \right) \leq \mathbf{c}_{B.3} \frac{\log(n)^3}{\theta n}.$$

Therefore,

$$\mathbb{E} \left(\sup_{K \in \mathcal{K}_n} \left\{ \|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} - \frac{1}{1 - \theta} \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 \right\} \right) \leq \bar{\mathbf{c}}_{2.5} \frac{\log(n)^5}{\theta(1 - \theta)n}$$

with $\bar{\mathbf{c}}_{2.5} = \mathbf{c}_{B.3} + \mathbf{c}_1$.

B.4. Proof of Theorem 3.2. The proof of Theorem 3.2 is dissected in three steps.

Step 1. This first step is devoted to provide a suitable decomposition of

$$\|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2.$$

First,

$$\|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2 = \|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - \widehat{s}_{K_0,\ell}(n; \cdot)\|_2^2 + \|\widehat{s}_{K_0,\ell}(n; \cdot) - s\|_2^2 - 2\langle \widehat{s}_{K_0,\ell}(n; \cdot) - \widehat{s}_{\widehat{K},\ell}(n; \cdot), \widehat{s}_{K_0,\ell}(n; \cdot) - s \rangle_2$$

From (7), it follows that for any $K \in \mathcal{K}_n$,

$$\begin{aligned} \|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2 &\leq \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 + \text{pen}(K) - \text{pen}(\widehat{K}) + \|\widehat{s}_{K_0,\ell}(n; \cdot) - s\|_2^2 \\ &\quad - 2\langle \widehat{s}_{K_0,\ell}(n; \cdot) - \widehat{s}_{\widehat{K},\ell}(n; \cdot), \widehat{s}_{K_0,\ell}(n; \cdot) - s \rangle_2 \\ (14) \quad &= \|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2 + \psi_n(K) - \psi_n(\widehat{K}) \end{aligned}$$

where

$$\psi_n(K) := 2\langle \widehat{s}_{K,\ell}(n; \cdot) - s, \widehat{s}_{K_0,\ell}(n; \cdot) - s \rangle_2 - \text{pen}(K).$$

Let's complete the decomposition of $\|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2$ by writing

$$\psi_n(K) = 2(\psi_{1,n}(K) + \psi_{2,n}(K) + \psi_{3,n}(K)),$$

where

$$\psi_{1,n}(K) := \frac{U_{K,K_0,\ell}(n)}{n^2},$$

$$\psi_{2,n}(K) := -\frac{1}{n^2} \left(\sum_{i=1}^n \ell(Y_i) \langle K_0(X_i, \cdot), s_{K,\ell} \rangle_2 + \sum_{i=1}^n \ell(Y_i) \langle K(X_i, \cdot), s_{K_0,\ell} \rangle_2 \right) + \frac{1}{n} \langle s_{K_0,\ell}, s_{K,\ell} \rangle_2 \text{ and}$$

$$\psi_{3,n}(K) := W_{K,K_0,\ell}(n) + W_{K_0,K,\ell}(n) + \langle s_{K,\ell} - s, s_{K_0,\ell} - s \rangle_2.$$

Step 2. In this step, we give controls of the quantities

$$\mathbb{E}(\psi_{i,n}(K)) \text{ and } \mathbb{E}(\psi_{i,n}(\widehat{K})) ; i = 1, 2, 3.$$

- By Lemma B.1, for any $\theta \in]0, 1[$,

$$\mathbb{E}(|\psi_{1,n}(K)|) \leq \frac{\theta}{n} \mathfrak{s}_{K,\ell} + \mathfrak{c}_{B.1} \frac{\log(n)^5}{\theta n}$$

and

$$\mathbb{E}(|\psi_{1,n}(\widehat{K})|) \leq \frac{\theta}{n} \mathbb{E}(\mathfrak{s}_{\widehat{K},\ell}) + \mathfrak{c}_{B.1} \frac{\log(n)^5}{\theta n}.$$

- On the one hand, for any $K, K' \in \mathcal{K}_n$, consider

$$\Psi_{2,n}(K, K') := \frac{1}{n} \sum_{i=1}^n \ell(Y_i) \langle K(X_i, \cdot), s_{K',\ell} \rangle_2.$$

Then, by Assumption 3.1,

$$\begin{aligned} \mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} |\Psi_{2,n}(K, K')| \right) &\leq \mathbb{E}(\ell(Y_1)^2)^{1/2} \mathbb{E} \left(\sup_{K, K' \in \mathcal{K}_n} \langle K(X_1, \cdot), s_{K',\ell} \rangle_2^2 \right)^{1/2} \\ &\leq \overline{\mathfrak{m}}_{\mathcal{K},\ell}^{1/2} \mathbb{E}(\ell(Y_1)^2)^{1/2}. \end{aligned}$$

On the other hand, by Assumption 2.1.(2),

$$|\langle s_{K,\ell}, s_{K_0,\ell} \rangle_2| \leq \mathfrak{m}_{\mathcal{K},\ell}.$$

Then, there exists a deterministic constant $\mathfrak{c}_1 > 0$, not depending on n and K , such that

$$\mathbb{E}(|\psi_{2,n}(K)|) \leq \frac{\mathfrak{c}_1}{n} \text{ and } \mathbb{E}(|\psi_{2,n}(\widehat{K})|) \leq \frac{\mathfrak{c}_1}{n}.$$

- By Lemma B.3,

$$\begin{aligned} \mathbb{E}(|\psi_{3,n}(K)|) &\leq \frac{\theta}{4} (\|s_{K,\ell} - s\|_2^2 + \|s_{K_0,\ell} - s\|_2^2) + 8\mathfrak{c}_{B.3} \frac{\log(n)^4}{\theta n} \\ &\quad + \left(\frac{\theta}{2}\right)^{1/2} \|s_{K,\ell} - s\|_2 \times \left(\frac{2}{\theta}\right)^{1/2} \|s_{K_0,\ell} - s\|_2 \\ &\leq \frac{\theta}{2} \|s_{K,\ell} - s\|_2^2 + \left(\frac{\theta}{4} + \frac{1}{\theta}\right) \|s_{K_0,\ell} - s\|_2^2 + 8\mathfrak{c}_{B.3} \frac{\log(n)^4}{\theta n} \end{aligned}$$

and

$$\mathbb{E}(|\psi_{3,n}(\widehat{K})|) \leq \frac{\theta}{2} \mathbb{E}(\|s_{\widehat{K},\ell} - s\|_2^2) + \left(\frac{\theta}{4} + \frac{1}{\theta}\right) \|s_{K_0,\ell} - s\|_2^2 + 8\mathfrak{c}_{B.3} \frac{\log(n)^4}{\theta n}.$$

Step 3. By the previous step, there exists a deterministic constant $\mathfrak{c}_2 > 0$, not depending on n , θ , K and K_0 , such that

$$\mathbb{E}(|\psi_n(K)|) \leq \theta \left(\|s_{K,\ell} - s\|_2^2 + \frac{\mathfrak{s}_{K,\ell}}{n} \right) + \left(\frac{\theta}{2} + \frac{2}{\theta} \right) \|s_{K_0,\ell} - s\|_2^2 + \mathfrak{c}_2 \frac{\log(n)^5}{\theta n}$$

and

$$\mathbb{E}(|\psi_n(\widehat{K})|) \leq \theta \mathbb{E} \left(\|s_{\widehat{K},\ell} - s\|_2^2 + \frac{\mathfrak{s}_{\widehat{K},\ell}}{n} \right) + \left(\frac{\theta}{2} + \frac{2}{\theta} \right) \|s_{K_0,\ell} - s\|_2^2 + \mathfrak{c}_2 \frac{\log(n)^5}{\theta n}.$$

Then, by Theorem 2.5,

$$\mathbb{E}(|\psi_n(K)|) \leq \frac{\theta}{1-\theta} \mathbb{E}(\|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2) + \left(\frac{\theta}{2} + \frac{2}{\theta} \right) \|s_{K_0,\ell} - s\|_2^2 + \left(\frac{\mathfrak{c}_2}{\theta} + \frac{\mathfrak{c}_{2.5}}{1-\theta} \right) \frac{\log(n)^5}{n}$$

and

$$\mathbb{E}(|\psi_n(\widehat{K})|) \leq \frac{\theta}{1-\theta} \mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2) + \left(\frac{\theta}{2} + \frac{2}{\theta} \right) \|s_{K_0,\ell} - s\|_2^2 + \left(\frac{\mathfrak{c}_2}{\theta} + \frac{\mathfrak{c}_{2.5}}{1-\theta} \right) \frac{\log(n)^5}{n}.$$

By decomposition (14), there exist two deterministic constants $\mathfrak{c}_3, \mathfrak{c}_4 > 0$, not depending on n, θ, K and K_0 , such that

$$\begin{aligned} \mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2) &\leq \mathbb{E}(\|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2) + \mathbb{E}(|\psi_n(K)|) + \mathbb{E}(|\psi_n(\widehat{K})|) \\ &\leq \left(1 + \frac{\theta}{1-\theta}\right) \mathbb{E}(\|\widehat{s}_{K,\ell}(n; \cdot) - s\|_2^2) + \frac{\theta}{1-\theta} \mathbb{E}(\|\widehat{s}_{\widehat{K},\ell}(n; \cdot) - s\|_2^2) \\ &\quad + \frac{\mathfrak{c}_3}{\theta} \|s_{K_0,\ell} - s\|_2^2 + \frac{\mathfrak{c}_4}{\theta(1-\theta)} \cdot \frac{\log(n)^5}{n}. \end{aligned}$$

This concludes the proof.

Acknowledgments. Many thanks to Fabienne Comte for her careful reading and advices.

REFERENCES

- [1] G. Chagny. *Warped Bases for Conditional Density Estimation*. Math. Methods Statist. 22, 253-282, 2013.
- [2] F. Comte and N. Marie. *Bandwidth Selection for the Wolverton-Wagner Estimator*. Journal of Statistical Planning and Inference 207, 198-214, 2020.
- [3] F. Comte and N. Marie. *On a Nadaraya-Watson Estimator with Two Bandwidths*. Submitted, 2020.
- [4] F. Comte and T. Rebafka. *Nonparametric Weighted Estimators for Biased Data*. Journal of Statistical Planning and Inference 174, 104-128, 2016.
- [5] U. Einmahl and D.M. Mason. *An Empirical Process Approach to the Uniform Consistency of Kernel-Type Function Estimators*. Journal of Theoretical Probability 13, 1-37, 2000.
- [6] U. Einmahl and D.M. Mason. *Uniform in Bandwidth Consistency of Kernel-Type Function Estimators*. Annals of Statistics 33, 1380-1403, 2005.
- [7] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge university press, 2015.
- [8] A. Goldenshluger and O. Lepski. *Bandwidth Selection in Kernel Density Estimation: Oracle Inequalities and Adaptive Minimax Optimality*. The Annals of Statistics 39, 1608-1632, 2011.
- [9] C. Houdré and P. Reynaud-Bouret. *Exponential Inequalities, with Constants, for U-Statistics of Order Two*. Stochastic Inequalities and Applications, vol. 56 of Progr. Proba., 55-69, Birkhauser, 2003.
- [10] C. Lacour, P. Massart and V. Rivoirard. *Estimator Selection: a New Method with Applications to Kernel Density Estimation*. Sankhya A 79, 2, 298-335, 2017.
- [11] M. Lerasle, N.M. Magalhaes and P. Reynaud-Bouret. *Optimal Kernel Selection for Density Estimation*. High dimensional probabilities VII: The Cargese Volume, vol. 71 of Prog. Proba., 435-460, Birkhauser, 2016.
- [12] P. Massart. *Concentration Inequalities and Model Selection*. Lecture Notes in Mathematics 1896, Springer, 2007.
- [13] S. Varet, C. Lacour, P. Massart and V. Rivoirard. *Numerical Performance of Penalized Comparison to Overfitting for Multivariate Density Estimation*. Preprint, 2020.

*LTCI, TÉLÉCOM PARIS, PALAISEAU, FRANCE
Email address: helene.halconruy@telecom-paris.fr

†LABORATOIRE MODAL'X, UNIVERSITÉ PARIS NANTERRE, NANTERRE, FRANCE
Email address: nmarie@parisnanterre.fr

*, †ESME SUDRIA, PARIS, FRANCE
Email address: helene.halconruy@esme.fr
Email address: nicolas.marie@esme.fr