# Adaptive warped kernel estimators

Gaëlle Chagny

# ADAPTIVE WARPED KERNEL ESTIMATORS

GAËLLE CHAGNY[A] *

Abstract. In this work, we develop a method of adaptive nonparametric estimation, based on "warped" kernels. The aim is to estimate a real-valued function $s$ from a sample of random couples $(X, Y)$. We deal with transformed data $(\Phi(X), Y)$, with $\Phi$ a one-to-one function, to build a collection of kernel estimators. The data-driven selection of the best bandwidth is done with a method inspired by Goldenshluger and Lepski (2011). The method permits to handle various problems such as additive and multiplicative regression, conditional density estimation, hazard rate estimation based on randomly right censored data, and cumulative distribution function estimation from current-status data. The interest is threefold. First, the squared-bias/variance trade-off is automatically realized. Next, non-asymptotic risk bounds are derived. Last, the estimator is easily computed thanks to its simple expression: a short simulation study is presented.

Keywords: Adaptive estimator. Censored data. Bandwidth selection. Nonparametric estimation. Regression. Warped kernel.

AMS Subject Classification 2010: 62G05; 62G08; 62N02.

## 1. Introduction

Let $(X, Y)$ be a couple of real random variables, and $(X_i, Y_i)_{i=1,\dots,n}$ an *i.i.d.* sample drawn as $(X, Y)$. The main goal of nonparametric estimation is to recover an unknown function $s$, linked with $(X, Y)$, such as the regression function, from the data. Among the huge variety of methods that have been investigated, the use of transformed data $(F_X(X_i), Y_i)$, with $F_X$ the cumulative distribution function (c.d.f.) of $X$, has received attention in the past decades. In this context, both kernel and projection estimators have been studied in random design regression estimation (Yang 1981, Stute 1984, Kerkyacharian and Picard 2004, Pham Ngoc 2009, Chagny 2011), conditional density or c.d.f estimation (Stute 1986, Mehra et al. 2000, Chagny 2012) or for the white noise model (Chesneau 2007). However, to our knowledge, few papers focus on the problem of adaptivity of such "warped estimators". The aim of the present work is twofold: first, we want to show that a warping kernel device can be applied to various estimation problems, including survival analysis models (see examples below). Secondly, we address the problem of bandwidth selection, with the intention of providing an adaptive "warped" estimator, which satisfies nonasymptotic risk bounds.

The basic idea, which motivates the study of warped kernel estimators introduced by Yang (1981), can be first explained in the classical regression framework. Here, the target function is the conditional expectation, $s : x \mapsto \mathbb{E}[Y|X = x]$ *i.e.*

$$s(x) = \frac{1}{f_X(x)} \int_{\mathbb{R}} y f_{(X,Y)}(x, y) dy, \tag{1}$$

* Corresponding author. Email: gaelle.chagny@parisdescartes.fr

[A] *Laboratoire MAP5 (UMR CNRS 8145), Université Paris Descartes, Sorbonne Paris Cité, France.*

when a density $f_{(X,Y)}$ for the couple $(X,Y)$ exists, and where $f_X$ is the marginal density of the design $X$. Historical kernel methods were initiated by Nadaraya (1964) and Watson (1964). The famous estimator named after them is built as the ratio of a kernel estimator of the product $sf_X$ divided by a kernel estimator of the density $f_X$:

$$\tilde{s}^{NW} : x \mapsto \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i K_h(x - X_i)}{\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)},$$

where $K_h : x \mapsto K(x/h)/h$, for $h > 0$, and $K : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} K(x)dx = 1$. Adaptive estimation then requires the automatic selection of the bandwidth $h$, and the ratio form of the estimate suggests that two such parameters should be selected: one for the numerator, and one for the denominator. From the theoretical point of view, there is no reason to choose the same. Nevertheless, nonasymptotic results such as oracle-inequality are difficult to derive for an estimator defined with two different data-driven smoothing parameters. See Penskaya (1995) for a thorough study of the ratio-form estimators. Moreover, when the design $X$ is very irregular (for example when a "hole" occurs in the data), a ratio may lead to instability (see Pham Ngoc 2009). The warped kernel estimators introduced by Yang (1981) and Stute (1984) avoid the ratio-form. Indeed denote by $\hat{F}_n$ the empirical c.d.f. of the $X_i$'s and let

$$\hat{s}_h = \frac{1}{n}\sum_{i=1}^{n} Y_i K_h\left(F_X(x) - F_X(X_i)\right), \text{ or } \hat{s}_h = \frac{1}{n}\sum_{i=1}^{n} Y_i K_h\left(\hat{F}_n(x) - \hat{F}_n(X_i)\right), \tag{2}$$

depending on whether the c.d.f. $F_X$ is known or not. The following equality (see Proposition 1) holds:

$$\mathbb{E}\left[Y K_h\left(u - F_X(X)\right)\right] = K_h \star (s \circ F_X^{-1})(u),$$

where $\star$ is the convolution product and $\circ$ is the composition symbol. Thus, the first estimator of (2) can be viewed as $\hat{s}_h = \widehat{s \circ F_X^{-1}} \circ F_X$. The main advantage is that its expression involves one bandwidth $h$ only.

In this paper, we generalize the warping strategy to various functional estimation problems: as a first extension of (1), we propose to recover functions $s$ of the form

$$s(x) = \frac{1}{\phi(x)} \int \theta(y) f_{(X,Y)}(x,y)dy, \tag{3}$$

for $\theta : \mathbb{R} \to \mathbb{R}$, and $\phi : \mathbb{R} \to \mathbb{R}_+\backslash\{0\}$. In this case, the warping device brings into play the transformation $(\Phi(X), Y)$ of the data, with $\Phi' = \phi$. The form (3) covers the additive regression model described above, by setting $\Phi = F_X$, and $\theta(y) = y$. But it also permits to deal with the simplified heteroskedastic model $Y = \sqrt{s(X)}\varepsilon$, where $\varepsilon$ is an unobserved noise. In this case, $\Phi = F_X$, and $\theta(y) = y^2$.

In several examples however, the couple $(X,Y)$ does not admit a density, but $X$ admits a marginal density. Then (3) can be extended and the target function $s$ takes the form:

$$s(x) = \frac{f_X(x)}{\phi(x)} \mathbb{E}[\theta(Y)|X = x]. \tag{4}$$

This allows to handle two classical settings in survival analysis: the interval censoring case 1, and right censored data. In the interval censoring model, case 1, the target function is $s(x) = \mathbb{P}(Z \leq x)$, where $Z$ is a survival time, which is not observed, and we only know a current status at the observed time $X$ of examination. We also know $Y = \mathbf{1}_{Z \leq X}$, which indicates whether $Z$ occurs before $X$ or not. We refer to Jewell and van der Laan (2004) for a review of the estimation methods in this setting, and more recently to Ma and Kosorok (2006), Brunel and Comte (2009) or Plancade (2011) for investigations including adaptivity. In right-censored

data, the function of interest at time $x$ is the hazard rate function, that is the risk of death at time $x$, given that the patient is alive until $x$. This model has been studied by Tanner and Wong (1983), Müller and Wang (1994) and Patil (1993), among all. Adaptive results are available for projection-type estimators (see Brunel and Comte 2005, 2008, Reynaud-Bouret 2006 or Akakpo and Durot 2010), but to our knowledge not for kernel estimators.

The paper is organized as follows. We present in Section 2 the estimation method, detail the examples illustrating the relevance of the introduction of a general target function $s$ defined by (4). We also study the global risk of the warped kernel estimators with fixed bandwidth. Section 3 is devoted to adaptive estimation: we define a data-driven choice of the bandwidth, inspired by Goldenshluger and Lepski (2011) which allows to derive nonasymptotic results for the adaptive estimators. Oracle-type inequalities are provided for the M.I.S.E., and convergence rates are deduced under regularity assumptions. Sections 2 and 3 deal with the case of known deformation $\Phi$. Section 4 discusses briefly the case of unknown $\Phi$, for which details are given in Appendix 1 (Section 7). In Section 5, the method is illustrated through numerical simulations. Proofs are gathered in Section 6. Finally, a more general proof of the main result is given in Appendix 2 (Section 8), under slightly different assumptions.

## 2. Estimation method

2.1. **Warped kernel strategy.** Consider a sample $(X_i, Y_i)_{i=1,\dots,n}$ of *i.i.d.* random couples with values in $A \times B$, where $A$ is an open interval of $\mathbb{R}$ and $B$ a Borel subset of $\mathbb{R}$. We assume that $X_i$ has a marginal density $f_X$ and we aim at recovering a function $s : A \to \mathbb{R}$ linked with the distribution of $(X_i, Y_i)$. To estimate $s$, we replace the explanatory variable $X_i$ by $\Phi(X_i)$, where $\Phi : A \to \Phi(A) \subset \mathbb{R}$ is one-to-one and absolutely continuous. The data $(\Phi(X_i), Y_i)_{i=1,\dots,n}$ are called the warped sample with deformation function $\Phi$. The sets $A, B, \Phi(A)$ are supposed to be given. The target function can be written as:

$$s(x) = g \circ \Phi(x) = g(\Phi(x)), \text{ with } g : \Phi(A) \to \mathbb{R}. \tag{5}$$

We first estimate the auxiliary function $g = s \circ \Phi^{-1}$ with $\Phi^{-1}$ the inverse function of $\Phi$. In the general case, $\Phi$ is unknown and we must estimate it also. Let $K$ be a function such that $\int_{\mathbb{R}} K(u)du = 1$ and set $K_h : u \mapsto K(u/h)/h$, for $h > 0$. We define, for $u \in \Phi(A)$,

$$\hat{g}_h(u) = \frac{1}{n} \sum_{i=1}^{n} \theta(Y_i) K_h \left( u - \hat{\Phi}(X_i) \right), \tag{6}$$

where $\theta : \mathbb{R} \to \mathbb{R}$ is a given function, $\hat{\Phi}$ is an empirical counterpart for $\Phi$, and for $x \in A$

$$\hat{s}_h(x) = \hat{g}_h \circ \hat{\Phi}(x) = \frac{1}{n} \sum_{i=1}^{n} \theta(Y_i) K_h \left( \hat{\Phi}(x) - \hat{\Phi}(X_i) \right). \tag{7}$$

Let us give examples covered by the above framework. They are sum up in Table 1, with the corresponding estimators.

**Example 1 (Additive random design regression):** we observe $(X_i, Y_i)$ with $Y_i = s(X_i) + \varepsilon_i$, $(\varepsilon_i)_{i=1,\dots,n}$ is independent of $(X_i)_{i=1,\dots,n}$, $\mathbb{E}[\varepsilon_i^2] < \infty$ and $\mathbb{E}[\varepsilon_i] = 0$. We choose $\Phi(x) = F_X(x)$, the cumulative distribution function (c.d.f. in the sequel) of $X$ and assume that $\Phi : A \to \Phi(A)$ is invertible.

| Example | $s$ | $\Phi$ | $\hat{s}_h$ |
|---|---|---|---|
| 1 $Y = s(X) + \varepsilon$ | $s$ | $F_X$ | $\frac{1}{n}\sum_{i=1}^{n} Y_i K_h(F_X(x) - F_X(X_i))$ |
| 2 $Y = \sigma(X)\varepsilon$ | $\sigma^2$ | $F_X$ | $\frac{1}{n}\sum_{i=1}^{n} Y_i^2 K_h(F_X(x) - F_X(X_i))$ |
| 3 $(X, \mathbf{1}_{Z\le X})$ | $F_Z$ | $F_X$ | $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Z_i\le X_i} K_h(F_X(x) - F_X(X_i))$ |
| 4 $(X = Z \wedge C, \mathbf{1}_{Z\le C})$ | $\frac{f_Z}{1-F_Z}$ | $\Phi(x) = \int_0^x (1 - F_X(t))dt$ | $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Z_i\le C_i} K_h(\Phi(x) - \Phi(X_i))$ |

TABLE 1. Summary of the studied examples and of the "warping" function used in each case.

**Example 2 (Heteroskedastic model):** $Y_i = \sigma(X_i)\varepsilon_i$, $(\varepsilon_i)_{i=1,\dots,n}$ independent of $(X_i)_{i=1,\dots,n}$, $\mathbb{E}[\varepsilon_i^2] = 1$, $\mathbb{E}[\varepsilon_i] = 0$, $\Phi(x) = F_X(x)$, with $\Phi : A \to \Phi(A)$ invertible. Here $s(x) = \sigma^2(x) = \mathbb{E}[Y_i^2 | X_i = x]$.

**Example 3 (Interval censoring, Case 1):** the observation is $(X_i, Y_i)$ where $Y_i = \mathbf{1}_{Z_i\le X_i}$, $Z_i, X_i \ge 0$ are independent survival times, $Y_i$ indicates whether $Z_i$ occurs before $X_i$ or not and $Z_i$ is not observed. The target function is $s(x) = \mathbb{P}(Z_i \le x) = \mathbb{E}[Y_i | X_i = x]$. We choose $\Phi = F_X$.

**Example 4 (Hazard rate estimation from right censored-data):** the observation is $X_i = Z_i \wedge C_i$, $Y_i = \mathbf{1}_{Z_i\le C_i}$, where $Z_i$ and $C_i$ are not observed and independent, $Z_i \ge 0$ is a lifetime and $C_i \ge 0$ is a censoring time. The function $s$ of interest is the hazard rate function $s(x) = f_Z(x)/(1 - F_Z(x))$, where $f_Z$ (resp. $F_Z$) is the density (resp. the c.d.f.) of $Z$. This function satisfies

$$s(x) = \frac{f_X(x)}{1 - F_X(x)} \mathbb{E}[Y | X = x], \tag{8}$$

a relation which is proved in Section 6.1. In this case, we assume $F_X(t) < 1$ for all $t \ge 0$, and take $\Phi(x) = \int_0^x (1 - F_X(t))dt$.

The following equality is the cornerstone of the method and justifies the introduction of (6):

**Proposition 1.** *Let $(X, Y)$ be a random couple with values in $A \times B \subset \mathbb{R}^2$. Assume that $X$ has a density $f_X$ and let $\Phi : A \to \Phi(A)$ be a one-to-one absolutely continuous function. Let $s : A \to \mathbb{R}$ defined by (4) and $g = s \circ \Phi^{-1}$. Then, if $\theta$ satisfies $\mathbb{E}|\theta(Y)K_h(u - \Phi(X)| < \infty$, for all $u \in \Phi(A)$,*

$$\mathbb{E}\left[\theta(Y)K_h(u - \Phi(X)\right] = K_h \star \left(g\mathbf{1}_{\Phi(A)}\right)(u) := g_h(u), \tag{9}$$

*where $\star$ is the convolution product.*

Equality (9) shows that $\hat{g}_h$, defined by (6) is an empirical version of $g_h$ and thus $\hat{s}_h$ in (7) suits well to estimate $s$.

Hereafter, for the sake of clarity, we assume that $\Phi$ is known. In Section 4, we discuss the case of an unknown deformation $\Phi$. The theoretical results are the same, up to further technicalities due to the plug-in of an empirical version for $\Phi$.

2.2. **Risk of the fixed bandwidth estimator.** In this section, we study the global properties of $\hat{s}_h$ as an estimate of $s$ on $A$, with a fixed bandwidth $h$. The quadratic risk weighted by the derivative $\phi$ of the warping function $\Phi$ is the natural criterion in our setting. Let us introduce, for a measurable function $t$ on $A$,

$$\|t\|_\phi^2 = \int_A t^2(x)\phi(x)dx, \tag{10}$$

and denote by $L^2(A,\phi)$ the space of functions $t$ for which the quantity (10) exists and is finite. We also use the corresponding scalar product $\langle .,. \rangle_\phi$. For $t_1, t_2$ belonging to $L^2(A,\phi)$, we have

$$\|t_1 \circ \Phi\|_\phi = \|t_1\|_{L^2(\Phi(A))}, \quad \langle t_1 \circ \Phi, t_2 \circ \Phi \rangle_\phi = \langle t_1, t_2 \rangle_{\Phi(A)},$$

where $\|t_1\|^2_{L^2(\Phi(A))} = \int_{\Phi(A)} t_1^2(x)dx$ and $\langle .,. \rangle_{\Phi(A)}$ denotes the usual scalar product on $L^2(\Phi(A))$. Therefore,

$$\|\hat{s}_h - s\|_\phi^2 = \|\hat{g}_h - g\|^2_{L^2(\Phi(A))}.$$

The following bias-variance decomposition of the risk holds:

**Proposition 2.** *Let $K$ belong to $L^2(\mathbb{R})$. Assume that $s$ belongs to $L^2(A,\phi)$, and that $\mathbb{E}[\theta^2(Y_1)] < \infty$. Then (recall that $g_h$ is defined in (9)),*

$$\mathbb{E}\left[\|\hat{s}_h - s\|_\phi^2\right] \leq \|g - g_h\|^2_{L^2(\Phi(A))} + \frac{1}{nh}\mathbb{E}\left[\theta^2(Y_1)\right] \|K\|^2_{L^2(\mathbb{R})}. \tag{11}$$

If $s$ is bounded on $A$, $s \in L^2(A,\phi)$. This is the case for Examples 1-3, as $\phi = f_X$. In Example 4, we can check that $s \in L^2(A,\phi)$ for all classical distributions for $C$ and $Z$ used in survival analysis (such as exponential, Weibull, Gamma...).

## 3. Adaptive estimation

3.1. **Data-driven choice of the bandwidth.** As usual, we must choose a bandwidth $h$ which realizes the best compromise between the squared-bias and the variance terms (see Proposition2). Moreover, we need define a data-driven choice of the bandwidth. For this, we use a method described in Goldenshluger and Lepski (2011). Let $\mathcal{H}_n$ be a finite collection of bandwidths, with cardinality depending on $n$ and properties precised below (Assumptions (H2)-(H3)). We introduce the auxiliary estimators, involving two kernels,

$$\hat{s}_{h,h'}(x) = \hat{g}_{h,h'}(\Phi(x)) \quad \text{with} \quad \hat{g}_{h,h'} = K_{h'} \star \left(\hat{g}_h \mathbf{1}_{\Phi(A)}\right).$$

For a constant $\kappa > 0$ to be precised later on, we define, for $h \in \mathcal{H}_n$,

$$V(h) = \kappa \left(1 + \|K\|^2_{L^1(\mathbb{R})}\right) \|K\|^2_{L^2(\mathbb{R})} \mathbb{E}\left[\theta^2(Y_1)\right] \frac{1}{nh}. \tag{12}$$

Next, we set

$$A(h) = \max_{h' \in \mathcal{H}_n} \left\{ \|\hat{s}_{h,h'} - \hat{s}_{h'}\|_\phi^2 - V(h') \right\}_+, \tag{13}$$

which is an estimation of the squared-bias term (see Lemma 4). Note that $\|\hat{s}_{h,h'} - \hat{s}_{h'}\|_\phi^2 = \|\hat{g}_{h,h'} - \hat{g}_{h'}\|^2$. Lastly, the adaptive estimator is defined in the following way:

$$\hat{s} = \hat{s}_{\hat{h}} \quad \text{with} \quad \hat{h} = \arg\min_{h\in\mathcal{H}_n} \{A(h) + V(h)\}. \tag{14}$$

The selected bandwidth $\hat{h}$ is data-driven. In $V(h)$, the expectation $\mathbb{E}[\theta^2(Y_1)]$ can be replaced by the corresponding empirical mean (see Brunel and Comte 2005, proof of Theorem 3.4 p.465). In Examples 3-4, it can be replaced by 1, its upper-bound.

3.2. **Results.** We consider the following assumptions:

(H1) The function $s$ is bounded. Denote by $\|s\|_{L^\infty(A)}$ its sup-norm.

(H2) There exist $\alpha_0 > 0$ and a constant $k_0 \geq 0$ such that $\displaystyle\sum_{h\in\mathcal{H}_n} \frac{1}{h} \leq k_0 n^{\alpha_0}$.

(H3) For all $\kappa_0 > 0$, there exists $C_0 > 0$, such that $\displaystyle\sum_{h\in\mathcal{H}_n} \exp\left(-\frac{\kappa_0}{h}\right) \leq C_0$.

(H4) The kernel $K$ is of order $l$, *i.e.* for all $j \in \{1, \dots, l+1\}$, the function $x \mapsto x^j K(x)$ is integrable, and for $1 \leq j \leq l$, $\int_{\mathbb{R}} x^j K(x)dx = 0$.

Assumption (H1) is required to obtain Theorem 3 below. Nevertheless the value $\|s\|_{L^\infty(A)}$ is not needed to compute the estimator (see (14)). This assumption holds in Example 3 ($s \leq 1$ in this case), and in Example 4, for instance when $Z$ has exponential or Gamma distribution. Assumptions (H2)-(H3) mean that the bandwidth collection should not be too large. For instance, the following classical collections satisfy these assumptions:

(1) $\mathcal{H}_{n,1} = \left\{k^{-1},\ k = 1, \dots, \chi(n)\right\}$ with $\alpha_0 = 2$, $\chi(n) = n$ or $\alpha_0 = 1$, $\chi(n) = \sqrt{n}$.

(2) $\mathcal{H}_{n,2} = \left\{2^{-k},\ k = 1, \dots, [\ln(n)/\ln(2)]\right\}$, with $\alpha_0 = 1$.

Assumption (H4) is required only to deduce convergence rate from the main nonasymptotic result. We need a moment assumption linked with (H2):

(H5) With $\alpha_0$ given by (H2), there exists $p > 2\alpha_0$, such that $\mathbb{E}[|\theta(Y) - \mathbb{E}[\theta(Y)|X]|^{2+p}] < \infty$.

If $\theta$ is bounded, (H5) evidently holds. In Examples 1 and 2, (H5) is a moment assumption on the noise which is usual in regression settings. Notice also that the smaller $\alpha_0$, the less restrictive the integrability constraint $p$ on the noise moments.
We prove the following oracle-type inequality:

**Theorem 3.** *Assume* (H1)-(H3) *and* (H5) *for Examples 1-2. Then there exist two constants* $c_1$ *and* $c_2$*, such that:*

$$\mathbb{E}\left[\|\hat{s} - s\|_\phi^2\right] \leq c_1 \min_{h\in\mathcal{H}_n} \left\{\|s - s_h\|_\phi^2 + \frac{\mathbb{E}\left[\theta^2(Y_1)\right] \|K\|_{L^2(\mathbb{R})}^2}{nh}\right\} + \frac{c_2}{n}, \tag{15}$$

*with* $s_h = g_h \circ \Phi$ *and* $\hat{s}$ *defined by (14). The constant* $c_1$ *depends only on* $\|K\|_{L^1(\mathbb{R})}$.

The constant $c_2$ depends on $\|s\|_{L^\infty(A)}$, $\|K\|_{L^1(\mathbb{R})}$ and $\|K\|_{L^2(\mathbb{R})}$ in Examples 3-4, and also on the moment of $\varepsilon$ and $\mathbb{E}[s^2(X_1)]$ for Examples 1-2. The adaptive estimator $\hat{s}$ automatically makes the squared-bias/variance compromise. The selected bandwidth $\hat{h}$ is performing as well as the unknown oracle:

$$\tilde{h} := \arg\min_{h\in\mathcal{H}_n} \mathbb{E}[\|s - \hat{s}_h\|_\phi^2].$$

up to the multiplicative constant $c_1$ and up to a remainding term of order $1/n$, which is negligible.

The interest of Inequality (15) is that it is nonasymptotic. Moreover, contrary to usual kernel estimation results, Assumption (H4) is not needed. This is one of the advantages of the bandwidth selection method.

To deduce convergence rates, smoothness classes must be defined to quantify the bias term. Define the Hölder class with order $\beta > 0$ and constant $L > 0$ by

$$\mathcal{H}(\beta, L) = \left\{ t : \mathbb{R} \to \mathbb{R},\ t^{(\lfloor \beta \rfloor)} \text{ exists}, \forall x, x' \in B,\ \left| t^{(\lfloor \beta \rfloor)}(x) - t^{(\lfloor \beta \rfloor)}(x') \right| \leq L|x - x'|^{\beta - \lfloor \beta \rfloor} \right\},$$

where $\lfloor \beta \rfloor$ is the largest integer less than $\beta$. We also need the Nikol'skii class of functions:

$$\mathcal{N}_2(\beta, L) = \left\{ t : \mathbb{R} \to \mathbb{R},\ t^{(\lfloor \beta \rfloor)} \text{ exists}, \forall x \in \mathbb{R},\ \int_{\mathbb{R}} \left( t^{(\lfloor \beta \rfloor)}(x' + x) - t^{(\lfloor \beta \rfloor)}(x') \right)^2 dx' \leq L^2 |x|^{2\beta - 2\lfloor \beta \rfloor} \right\}$$

We can now deduce from Theorem 3 the convergence rate of the risk, under regularity assumptions for the auxiliary function $g$.

**Corollary 1.** *Let $\tilde{g} = g\mathbf{1}_{\phi(A)}$ on $\mathbb{R}$. Assume that*

- *$\tilde{g}$ belongs to the Hölder class $\mathcal{H}(\beta, L)$, with $\tilde{g}(0) = \tilde{g}(1)$ in Examples 1-3,*
- *$\tilde{g}$ belongs to the Nikol'skii class $\mathcal{N}_2(\beta, L)$ in Example 4.*

*Assume* (H4) *with $l = \lfloor \beta \rfloor$. Then, under the assumptions of Theorem 3,*

$$\mathbb{E}\left[ \|\hat{s} - s\|_\phi^2 \right] \leq C n^{-\frac{2\beta}{2\beta+1}}, \tag{16}$$

*where $C$ is a constant which does not depend on $n$ and $\beta$.*

In Examples 1-3, $\Phi(A) = (0; 1)$ and the Hölder condition is enough. In Example 4, $\Phi(A) = \mathbb{R}_+$ and we need the Nikol'skii condition. Both spaces are standard in kernel estimation, see *e.g.* Tsybakov (2009) and Goldenshluger and Lepski (2011).
We recover the classical optimal rates in nonparametric estimation. Note however that our regularity assumptions are set on $g$ and not $s$, as long as we do not consider specific warped spaces defined in Kerkyacharian and Picard (2004).

## 4. The general case of unknown $\Phi$

Up to now we have considered the case of a known "warping" function $\Phi$. This is also the framework of *e.g.* Pham Ngoc (2009) or Chesneau (2007). It allows to derive the main result with few assumptions and short proofs. To deal with the general case, we use a plug-in device. Let $\hat{F}_n$ be the empirical c.d.f. of $X$. We estimate $\Phi$ by $\hat{\Phi}(x) = \hat{F}_n(x)$ for Examples 1-3, and by $\hat{\Phi}(x) = \int_0^x \left(1 - \hat{F}_n(t)\right) dt$ for Example 4. Now $\hat{g}_h$ is given by (6). To define $\hat{h}$, we replace $\|\hat{s}_{h,h'} - \hat{s}_{h'}\|_\phi^2$ by $\|\hat{g}_{h,h'} - \hat{g}_{h'}\|^2$ in $A(h)$ (see (13)). Theorem 3 holds under stronger assumptions on the bandwidth collection $\mathcal{H}_n$. However the proof requires lengthy and cumbersome technicalities. To deal with the difference $\hat{\Phi} - \Phi$, we use the deviation inequality of Dvoretzky et al. (1956): for any $\lambda > 0$,

$$\mathbb{P}\left( \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_X(x) \right| \geq \lambda \right) \leq K \exp\left(-2n\lambda^2\right),$$

with $K$ an universal constant. Details are given in Chagny (2011, 2012) for regression and conditional density estimation using warped bases, which require analogous arguments. Moreover, a non adaptive bound for the risk is proved in the appendix, Section 7, as an example of the required tools.

## 5. Illustration

To illustrate the procedure, we focus only on two of the four examples: the additive regression (Example 1), and the estimation of c.d.f. under interval censoring case I (Example 3). In each case, we compare the warped kernel strategy, denoted by WK in this section, with another adaptive method: a regression-type one, based on the minimization of a penalized least-squares contrast. We denote it by LS.

5.1. **Implementation of the warped-kernel estimators.** The theoretical study allows the choice of several kernels and bandwidth collections. For practical purpose, we consider the Gaussian kernel, $K : x \mapsto e^{-x^2/2}/\sqrt{2\pi}$, which satisfies Assumption $(K_1)$. It has the advantage of having simple convolution-products:

$$\forall h, h' > 0, \quad K_h \star K_{h'} = K_{\sqrt{h^2+h'^2}}. \tag{17}$$

The experiment is conducted with the dyadic collection $\mathcal{H}_{n,2}$ defined above. The larger collection $\mathcal{H}_{n,1}$ has also been tested: since it does not really improve the results but increases the computation time, we only keep the other collection. Besides, the simulations are performed in the case of unknown $\Phi$. Therefore in Examples 1 and 3, the estimator is

$$\hat{s} \; : \; x \; \mapsto \; \frac{1}{n}\sum_{i=1}^{n} \theta(Y_i) K_{\hat{h}}(\hat{F}_n(x) - \hat{F}_n(X_i)),$$

with $\hat{F}_n$ the empirical c.d.f. of the $X_i$'s. Then, the estimation procedure can be decomposed in some steps:

• Simulate a data sample $(X_i, Y_i)$, $i = 1, \ldots, n$, fitting Example 1 or 3.

• Compute $V(h)$ and $A(h)$ for each $h \in \mathcal{H}_{n,1}$. - For $V(h)$: its computation require a value for $\kappa$ (see (12)). A lower bound for its theoretical value is provided by the proof: it is very pessimistic due to rough upper-bounds (for the sake of clarity). A practical calibration is required, like in most model selection devices. Since classical techniques such as the slope heuristic are not currently well developed for the Goldenshluger-Lepki method, we adjust $\kappa$ on simulations, prior to the comparison with the other estimates. We set $\kappa = 0.05$ in Example 1, and $\kappa = 0.3$ in Example 3.

- For $A(h)$: thanks to (17), the auxiliary estimates are easily computed: $\hat{s}_{h,h'} = \hat{s}_{\sqrt{h^2+h'^2}}$. The $L^2-$norm is then approximated by a Riemann sum:

$$\|\hat{s}_{h,h'} - \hat{s}_{h'}\|^2_{\phi'_X} = \|\hat{g}_{h,h'} - \hat{g}_{h'}\|^2_{L^2(\phi_X(A))} \approx \frac{1}{K}\sum_{k=1}^{K} \left(\hat{g}_{h,h'}(u_k) - \hat{g}_{h'}(u_k)\right)^2,$$

where $K = 50$, and $(u_k)_k$ are grid points evenly distributed across $(0; 1)$.

• Select $\hat{h}$ such that $A(h) + V(h)$ is minimum.

• Compute $\hat{s}_{\hat{h}}$.

5.2. **Example 1: additive regression.** We compare the warped kernel method (WK) with the adaptive estimator studied in Baraud (2002). It is a projection estimator, developed in an orthogonal basis of $L^2(A)$, and built with a penalized least-squares contrast. The experiment is carried out with the Matlab toolbox FY3P, written by Yves Rozenholc, and available on his web page http://www.math-info.univ-paris5.fr/ rozen/. A regular piecewise polynomial basis is used, with degrees chosen in an adaptive way. Since the kernel we choose has only one vanishing moment, the comparison is fair if we consider polynomials with degrees equal to or less than 1. We denote by LS1 the resulting estimator. However, as shown below, we will see that the warped-kernel generally outperforms the least-square, even if we use polynomials with degree

at most 2 (LS2). We also experiment the Fourier basis, but the results are not as good as the polynomial basis for the least-squares estimator. Thus, we do not mention the values.

The procedure is applied for different regression functions, design and noise. We focus on the three following regression functions

$$\begin{aligned} s_1 &: x \mapsto x(x-1)(x-0.6) \\ s_2 &: x \mapsto -\exp(-200(x-0.1)^2) - \exp(-200(x-0.9)^2) + 1 \\ s_3 &: x \mapsto \cos(4\pi x) + \exp(-x^2) \end{aligned}$$

The influence of the design is explored through four distributions:

- $\mathcal{U}_{[0;1]}$, the uniform distribution on the interval $[0;1]$,
- $\gamma(4, 0.08)$, the Gamma distribution, with parameters 4 and 0.08 (0.08 is the scale parameter),
- $\mathcal{N}(0.5, 0.01)$, the Gaussian distribution with mean 0.5 and variance 0.01,
- $\mathcal{BN}$ a bimodal Gaussian distribution, with density $x \mapsto c(\exp(-200(x-0.05)^2) + \exp(-200(x-0.95)^2))$ ($c$ is a constant adjusted to obtain a density function).

We also test the sensibility of the method to the noise distribution: contrary to the underlying design distribution, it does not seem to affect the results. Thus, we present the simulation for a Gaussian centered noise, with variance $\sigma^2$. The value of $\sigma$ is chosen in such a way that the signal-to-noise ratio (the ratio of the variance of the signal $\mathrm{Var}(s(X_1))$ over the variance of the noise $\mathrm{Var}(\varepsilon_1)$) approximately equals 2.

Beams of estimators (WK, LS1, and LS2) are presented in Figures 1 and 2, with the generated data-sets and the function to estimate. Precisely, Figure 1 shows a regular case, while Figure 2 depicts the case where a hole occurs in the design density: the estimator built with warped kernel behaves still correctly, even if the data are very inhomogeneous.

A study of the risk is reported in Table 2, for the sample sizes $n = 60, 200, 500$ and 1000. The MISE is obtained by averaging the following approximations of the ISE values, for $j \in \{1, \dots, J = 200\}$, computed with $J$ sample replications:

$$ISE_j = \frac{b-a}{N} \sum_{k=1}^{N} \left(\tilde{s}(x_k) - s(x_k)\right)^2,$$

where $\tilde{s}$ stands for one of the estimators, $b$ is the quantile of order 95% of the $X_i$ and $a$ is the quantile of order 5%. The $(x_k)_{k=1,\dots,N}$ are the sample points falling in $[a;b]$. In 56% of the examples, the risks of the warped-kernel estimator are smaller than the ones of the least-squares estimator, in piecewise polynomials basis with degrees at most 2 (LS2). Besides, if we consider the comparison with LS1, which is more fair as explained above, the WK estimators give better results in 77% of the cases.

5.3. **Example 3: Interval censoring, case 1.** The same comparison is carried out for the estimation of the c.d.f. under interval censoring. The adaptive least-squares estimate is provided by Brunel and Comte (2009), and the same Matlab toolbox is used for its implementation: recall that the target function can be seen as a regression function: $s(x) = \mathbb{P}(Z \leq x) = \mathbb{E}[\mathbf{1}_{Z \leq x} | X = x]$. Different models are considered for generating the data. The estimation set $A$ is calibrated, such that most of the data belong to this interval, as it is done in Brunel and Comte (2009). We shorten "follows the distribution" by the symbol "$\sim$".

- M1: $X \sim \mathcal{U}_{[0;1]}$, and $Z \sim \mathcal{U}_{[0;1]}$, $A = (0;1)$ (for instance, the target function is $F_Z : x \mapsto x$),

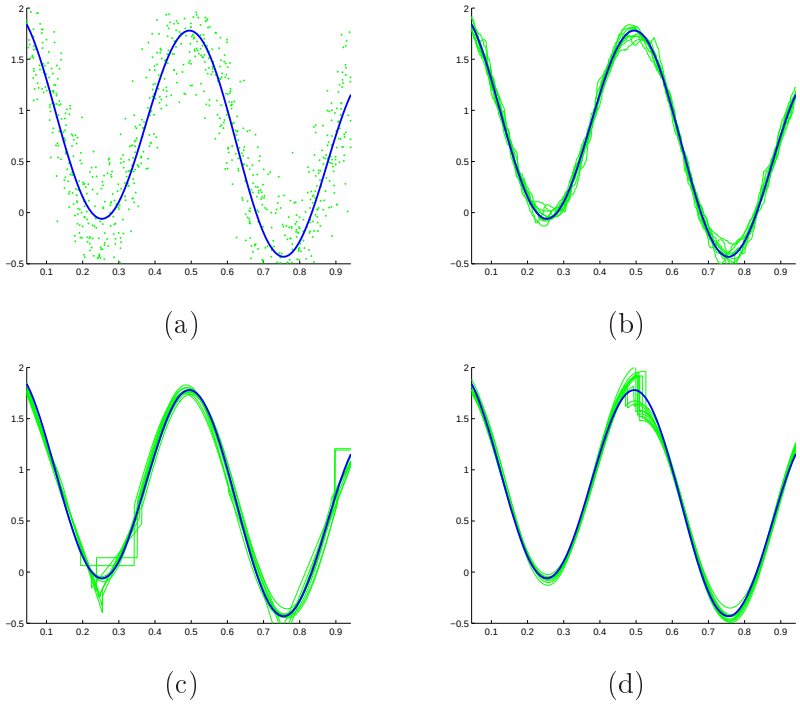


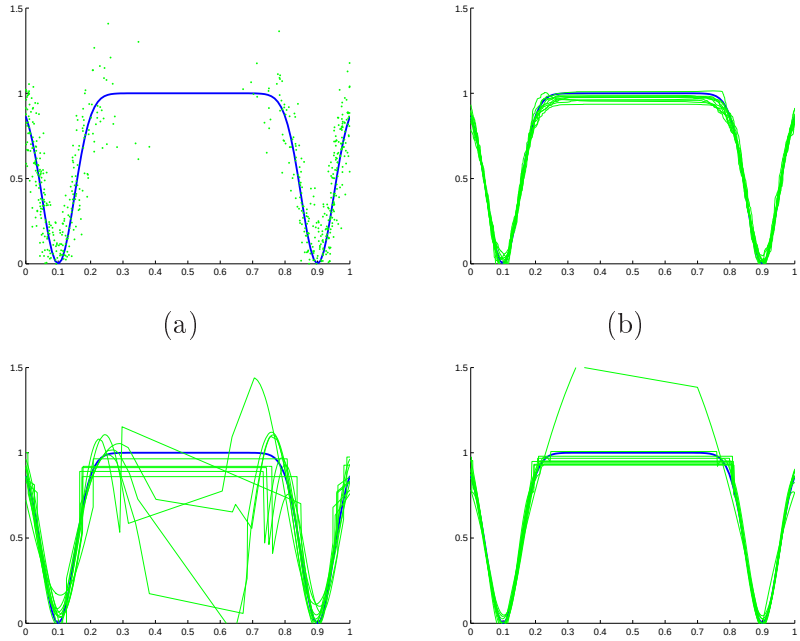
FIGURE 1. Estimation in Example 1, with true regression function $s_3$, design distribution $\gamma(4, 0.08)$, and $n = 1000$. (a) points: data $(X_i, Y_i)_i$, thick line: true function $s_3$. (b)-(c)-(d) beams of 20 estimators built from i.i.d. sample (thin lines) versus true function (thick line): warped kernel estimators (subplot (b)), least-squares estimator in piecewise polynomial bases with degree at most 1 (subplot (c)) or 2 (subplot (d)).

- M2: $X \sim \mathcal{U}_{[0;1]}$, and $Z \sim \chi_2(1)$ (Chi-squared distribution with 1 degree of freedom), $A = (0; 1)$,
- M3: $X \sim \mathcal{E}(1)$ (exponential distribution with mean 1), and $Z \sim \chi_2(1)$, $A = (0; 1.2)$,
- M4: $X \sim \beta(4, 6)$ (Beta distribution of parameter (4,6)), $Z \sim \beta(4, 8)$, $A = (0; 0.5)$,
- M5: $X \sim \beta(4, 6)$, $Z \sim \mathcal{E}(10)$ (exponential distribution with mean 0.1), $A = (0; 0.5)$,
- M6: $X \sim \gamma(4, 0.08)$, $Z \sim \mathcal{E}(10)$, $A = (0, 0.5)$,
- M7: $X \sim \mathcal{E}(0.1)$, $Z \sim \gamma(4, 3)$, $A = (1; 23)$.

The first two models, and the fourth, were also used by Brunel and Comte (2009). All these models allow to investigate thoroughly the sensibility of the method to the distribution of the examination time $X$, and to the range of the estimation interval.

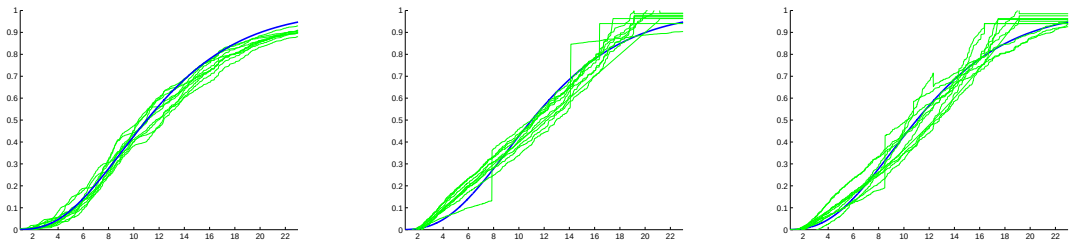
Figure 3 shows the smoothness of warped-kernel estimates. We also explore the difference between the estimators by computing the MISE for the different models. Table 3 reveals that the warped-kernel estimates can advantageously be used as soon as the design $X_i$ has not a uniform distribution: it always outperforms the least-squares estimators in these cases.

## 6. PROOFS



(a) (b) (c) (d)

FIGURE 2. Estimation in Example 1, with true regression function $s_2$, design distribution $\mathcal{BN}$, and $n = 1000$. (a) points: data $(X_i, Y_i)_i$, thick line: true function $s_2$. (b)-(c)-(d) beams of 20 estimators built from i.i.d. sample (thin lines) versus true function (thick line): warped kernel estimators (subplot (b)), least-squares estimator in piecewise polynomial bases with degree at most 1 (subplot (c)) or 2 (subplot (d)).



(a) (b) (c)

FIGURE 3. Estimation in Example 3, in model M7, and $n = 1000$. (a)-(b)-(c) beams of 20 estimators built from i.i.d. sample (thin lines) versus true function (thick line): warped kernel estimators (subplot (a)), least-squares estimator in piecewise polynomial bases with degree at most 1 (subplot (b)) or 2 (subplot (c)).

6.1. **Proof of Equality (8).** Equality (8) comes down to compute a conditional expectation: precisely, we prove that

$$\mathbb{E}\left[Y|X=x\right] = \frac{f_Z(x)}{f_X(x)}\frac{1-F_X(x)}{1-F_Z(x)}.$$

| $s$ | $X$ | $\sigma$ | $n=60$ | 200 | 500 | 1000 | Method |
|---|---|---|---|---|---|---|---|
| $s_1$ | $\mathcal{U}_{[0;1]}$ | $\sqrt{.0006}$ | 0.3719 | 0.1341 | 0.1957 | 0.2454 | WK |
| | | | 0.3892 | 0.1293 | 0.0681 | 0.0446 | LS2 |
| | $\gamma(4,0.08)$ | $5.10^{-5}$ | 0.0052 | 0.0033 | 0.0004 | 0.0003 | WK |
| | | | 0.0097 | 0.004 | 0.0017 | 0.0012 | LS2 |
| | $\mathcal{N}(0.5,0.01)$ | 0.011 | 0.0049 | 0.0020 | 0.0008 | 0.0005 | WK |
| | | | 0.0020 | 0.0012 | 0.0010 | 0.0008 | LS2 |
| | $B\mathcal{N}$ | 0.022 | 0.524 | 0.422 | 0.267 | 0.205 | WK |
| | | | 0.166 | 0.054 | 0.038 | 0.029 | LS2 |
| $s_2$ | $\mathcal{U}_{[0;1]}$ | 0.17 | 16.35 | 6.791 | 3.51 | 0.837 | WK |
| | | | 33.212 | 2.058 | 0.691 | 0.407 | LS2 |
| | $\gamma(4,0.08)$ | 0.08 | 1.885 | 0.354 | 0.204 | 0.147 | WK |
| | | | 4.047 | 0.801 | 0.552 | 0.429 | LS2 |
| | $\mathcal{N}(0.5,0.01)$ | 0.01 | 0.0619 | 0.0186 | 0.0079 | 0.0006 | WK |
| | | | 0.0078 | 0.0014 | 0.0001 | 0.0001 | LS2 |
| | $B\mathcal{N}$ | 0.18 | 12.052 | 5.279 | 1.698 | 1.041 | WK |
| | | | 52.668 | 11.009 | 5.817 | 1.215 | LS2 |
| $s_3$ | $\mathcal{U}_{[0;1]}$ | 0.35 | 28.03 | 10.55 | 4.63 | 2.747 | WK |
| | | | 125.055 | 45.298 | 12.607 | 5.713 | LS1 |
| | | | 31.073 | 7.477 | 4.199 | 3.319 | LS2 |
| | $\gamma(4,0.08)$ | 0.44 | 19.615 | 6.283 | 3.869 | 3.309 | WK |
| | | | 41.261 | 13.34 | 4.808 | 3.727 | LS1 |
| | | | 23.213 | 5.549 | 2.059 | 0.86 | LS2 |
| | $\mathcal{N}(0.5,0.01)$ | 0.44 | 6.341 | 2.452 | 1.28 | 0.861 | WK |
| | | | 10.453 | 3.961 | 2.098 | 1.078 | LS1 |
| | | | 3.753 | 1.386 | 1.028 | 0.644 | LS2 |
| | $B\mathcal{N}$ | 0.32 | 44.381 | 13.618 | 9.637 | 7.928 | WK |
| | | | 182.525 | 58.787 | 24.229 | 12.317 | LS1 |
| | | | 66.663 | 30.377 | 8.521 | 4.574 | LS2 |

TABLE 2. Values of MISE $\times 1000$ averaged over 200 samples, for the estimators of the regression function (Example 1), built with the warped kernel method (WK) or the least-squares methods, with piecewise polynomials of degree at most 1 or 2 (LS1 or LS2).

To do so, let $H$ be a test function. Recall that $X = Z \wedge C$, $Y = \mathbf{1}_{Z \leq C}$, and denote by $f_Z$ (resp. $f_C$) the density of $Z$ (resp. $C$). We compute

$$\begin{aligned}\mathbb{E}\left[YH(X)\right] &= \mathbb{E}\left[\mathbf{1}_{Z\leq C}H(Z)\right] = \int_{\mathbb{R}_+^2} \mathbf{1}_{z\leq c}H(z)f_Z(z)f_C(c)dzdc,\\ &= \int_{\mathbb{R}_+} \left(1-F_C(z)\right)H(z)f_Z(z)dz = \int_{\mathbb{R}_+} \frac{f_Z(z)}{f_X(z)}\frac{1-F_X(z)}{1-F_Z(z)}H(z)f_X(z)dz,\end{aligned}$$

| Model | $X$ | $Z$ | [a;b] | $n=60$ | 200 | 500 | 1000 | Method |
|---|---|---|---|---|---|---|---|---|
| 1 | $\mathcal{U}_{[0;1]}$ | $\mathcal{U}_{[0;1]}$ | $[0;1]$ | 2.41 | 1.125 | 0.975 | 0.533 | WK |
| | | | | 0.63 | 0.111 | 0.056 | 0.024 | LS2 |
| 2 | $\mathcal{U}_{[0;1]}$ | $\chi_2(1)$ | $[0;1]$ | 1.558 | 0.804 | 0.57 | 0.415 | WK |
| | | | | 1.602 | 0.44 | 0.244 | 0.13 | LS2 |
| 3 | $\mathcal{E}(1)$ | $\chi_2(1)$ | $[0;1.2]$ | 1.285 | 0.614 | 0.243 | 0.247 | WK |
| | | | | 2.385 | 0.893 | 0.651 | 0.365 | LS2 |
| 4 | $\mathcal{B}(4,6)$ | $\mathcal{B}(4,8)$ | $[0;0.5]$ | 0.423 | 0.236 | 0.09 | 0.094 | WK |
| | | | | 0.449 | 0.271 | 0.117 | 0.105 | LS2 |
| 5 | $\mathcal{B}(4,6)$ | $\mathcal{E}(10)$ | $[0;0.5]$ | 0.388 | 0.229 | 0.119 | 0.103 | WK |
| | | | | 0.467 | 0.261 | 0.13 | 0.095 | LS2 |
| 6 | $\gamma(4,0.08)$ | $\mathcal{E}(10)$ | $[0;0.5]$ | 0.424 | 0.166 | 0.102 | 0.069 | WK |
| | | | | 0.698 | 0.286 | 0.162 | 0.095 | LS2 |
| 7 | $\mathcal{E}(0.1)$ | $\gamma(4,3)$ | $[1;23]$ | 14.955 | 5.145 | 3.973 | 2.113 | WK |
| | | | | 19.825 | 11.797 | 9.738 | 5.898 | LS2 |

TABLE 3. Values of MISE $\times 100$ averaged over 100 samples, for the estimators of the c.d.f. from current status data (Example 3) built with the warped kernel method (WK) or the least-squares methods, with piecewise polynomials of degree at most 1 or 2 (LS1 or LS2).

taking into account the equality $1-F_X=(1-F_Z)(1-F_C)$. Thus, we identify $\mathbb{E}[Y|X]$.

□

6.2. **Proof of Proposition 1.** We have:

$$\begin{aligned}\mathbb{E}\left[\theta(Y)K_h(u-\Phi(X))\right] &= \mathbb{E}\left[\mathbb{E}\left[\theta(Y)|X\right]K_h(u-\Phi(X))\right],\\ &= \int_A K_h(u-\Phi(x))\mathbb{E}\left[\theta(Y)|X=x\right]f_X(x)dx.\end{aligned}$$

We set $u'=\Phi(x)$, thus $du'=\phi(x)dx$. Therefore,

$$\begin{aligned}\mathbb{E}\left[\theta(Y)K_h(u-\Phi(X))\right] &= \int_{\Phi(A)} K_h(u-u')\mathbb{E}\left[\theta(Y)|X=\Phi^{-1}(u)\right]f_X\left(\Phi^{-1}(u)\right)\frac{du}{\phi\circ\Phi^{-1}(u)},\\ &= \int_{\Phi(A)} K_h(u-u')s\circ\Phi^{-1}(u)du.\end{aligned}$$

□

6.3. **Proof of Proposition 2.** The following classical bias-variance decomposition holds:

$$\mathbb{E}\left[\|\hat{s}_h - s\|_\phi^2\right] = \|g - g_h\|^2_{L^2(\Phi(A))} + \mathbb{E}\left[\|g_h - \hat{g}_h\|^2_{L^2(\Phi(A))}\right],$$

since, thanks to (9), $\mathbb{E}[\hat{g}_h(u)] = g_h(u)$. We bound the variance term as follows:

$$\mathbb{E}\left[\|g_h - \hat{g}_h\|^2_{L^2(\Phi(A))}\right] = \mathbb{E}\left[\int_{\Phi(A)} \left(\hat{g}_h(u) - \mathbb{E}\left[\hat{g}_h(u)\right]\right)^2 du\right] = \int_{\Phi(A)} \mathrm{Var}\left(\hat{g}_h(u)\right) du,$$

and for each $u \in \Phi(A)$,

$$\mathrm{Var}\left(\hat{g}_h(u)\right) = \frac{1}{n}\mathrm{Var}\left(\theta(Y_1)K_h\left(u - \Phi(X_1)\right)\right) \leq \frac{1}{n}\mathbb{E}\left[\theta^2(Y_1)K_h^2\left(u - \Phi(X_1)\right)\right].$$

Therefore, by integrating with respect to $u$, we get

$$\mathbb{E}\left[\|g_h - \hat{g}_h\|^2_{L^2(\Phi(A))}\right] \leq \mathbb{E}\left[\theta^2(Y_1)\right]\|K\|^2_{L^2(\mathbb{R})}\frac{1}{nh}.$$

6.4. **Proof of Theorem 3.** Let $h \in \mathcal{H}_n$ be fixed. We start with the following decomposition for the loss of the estimator $\tilde{s} = \hat{s}_{\hat{h}}$:

$$\begin{aligned}\left\|\hat{s}_{\hat{h}} - s\right\|_\phi^2 &= \left\|\hat{g}_{\hat{h}} - g\right\|^2_{L^2(\Phi(A))}, \\ &\leq 3\left\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\right\|^2_{L^2(\Phi(A))} + 3\left\|\hat{g}_{h,\hat{h}} - \hat{g}_h\right\|^2_{L^2(\Phi(A))} + 3\left\|\hat{g}_h - g\right\|^2_{L^2(\Phi(A))}.\end{aligned}$$

The definitions of $A(h)$ and $A(\hat{h})$ enable us to write, using the definition of $\hat{h}$,

$$\begin{aligned}3\left\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\right\|^2_{L^2(\Phi(A))} + 3\left\|\hat{g}_{h,\hat{h}} - \hat{g}_h\right\|^2_{L^2(\Phi(A))} &\leq 3\left(A(h) + V\left(\hat{h}\right)\right) + 3\left(A\left(\hat{h}\right) + V(h)\right), \\ &\leq 6\left(A(h) + V(h)\right),\end{aligned}$$

Besides, applying also Proposition 2, we obtain

$$\mathbb{E}\left[\left\|\hat{s}_{\hat{h}} - s\right\|_\phi^2\right] \leq 6\mathbb{E}\left[A(h)\right] + 6V(h) + \frac{\mathbb{E}[\theta^2(Y_1)]\|K\|^2_{L^2(\mathbb{R})}}{nh} + 3\|g_h - g\|^2_{L^2(\Phi(A))}. \tag{18}$$

Therefore, the remainding part of the proof follows from the lemma hereafter.

**Lemma 4.** *Let $h \in \mathcal{H}_n$ be fixed. Under the assumptions of Theorem 3, there exist constants $C_1, C_2$ such that,*

$$\mathbb{E}\left[A(h)\right] \leq C_1\|g_h - g\|^2_{L^2(\Phi(A))} + \frac{C_2}{n}, \tag{19}$$

*where the constant $C_1$ only depends on $\|K\|_{L^1(\mathbb{R})}$.*

Applying Inequality (19) in (18) implies (15) by taking the infimum over $h \in \mathcal{H}_n$. This ends the proof of Theorem 3.

□

6.5. **Proof of Lemma 4.** To study $A(h)$, we introduce the auxiliary quantities $g_{h,h'} := K_{h'} \star (g_h \mathbf{1}_{\Phi(A)}) = K_{h'} \star ((K_h \star g\mathbf{1}_{\Phi(A)})\mathbf{1}_{\Phi(A)})$, for any $h' \in \mathcal{H}_n$, and we first split

$$\|\hat{s}_{h,h'} - \hat{s}_{h'}\|_{f_X}^2 = \|\hat{g}_{h,h'} - \hat{g}_{h'}\|_{L^2(\Phi(A))}^2 \leq 3\left(T_a + T_b + \|\hat{g}_{h'} - g_{h'}\|_{L^2(\Phi(A))}^2\right), \tag{20}$$

where

$$T_a = \|\hat{g}_{h,h'} - g_{h,h'}\|_{L^2(\Phi(A))}^2, \quad T_b = \|g_{h,h'} - g_{h'}\|_{L^2(\Phi(A))}^2.$$

The first term can be bounded as follows, using Lemma 8, with $p = 2$, $q = 1$, and $r = 2$:

$$\begin{aligned} T_a &\leq \left\|K_h \star \left(\hat{g}_{h'}\mathbf{1}_{\Phi(A)} - g_{h'}\mathbf{1}_{\Phi(A)}\right)\right\|_{L^2(\mathbb{R})}^2, \\ &\leq \|K\|_{L^1(\mathbb{R})}^2 \left\|\hat{g}_{h'}\mathbf{1}_{\Phi(A)} - g_{h'}\mathbf{1}_{\Phi(A)}\right\|_{L^2(\mathbb{R})}^2 = \|K\|_{L^1(\mathbb{R})}^2 \|\hat{g}_{h'} - g_{h'}\|_{L^2(\Phi(A))}^2. \end{aligned}$$

In the same way, $T_b \leq \|K_{h'}\|_{L^1(\mathbb{R})}^2 \|g_h - g\|_{L^2(\Phi(A))}^2$. Therefore, Decomposition (20) becomes:

$$\|\hat{s}_{h,h'} - \hat{s}_{h'}\|_{f_X}^2 \leq 3\|K\|_{L^1(\mathbb{R})}^2 \|g - g_h\|_{L^2(\Phi(A))}^2 + 3(1 + \|K\|_{L^1(\mathbb{R})}^2) \|\hat{g}_{h'} - g_{h'}\|_{L^2(\Phi(A))}^2.$$

Now, we get back to the definition of $A(h)$ given by (13):

$$\begin{aligned} A(h) \leq 3 &\|K\|_{L^1(\mathbb{R})}^2 \|g - g_h\|_{L^2(\Phi(A))}^2 \\ &+3(1 + \|K\|_{L^1(\mathbb{R})}^2) \max_{h' \in \mathcal{H}_n} \left(\|\hat{g}_{h'} - g_{h'}\|_{L^2(\Phi(A))}^2 - \frac{V(h')}{3(1 + \|K\|_{L^1(\mathbb{R})}^2)}\right)_+. \end{aligned} \tag{21}$$

We apply Lemma 7: $\|\hat{g}_{h'} - g_{h'}\|_{L^2(\Phi(A))} = \sup_{t \in \bar{S}(0,1)} \langle \hat{g}_{h'} - g_{h'}, t\rangle_{\Phi(A)}$, with $\bar{S}(0,1)$ a dense countable subset of $\tilde{S}(0,1) = \{t \in L^1(\Phi(A)) \cap L^2(\Phi(A)),\ \|t\|_{L^2(\Phi(A))} = 1\}$. Now,

$$\begin{aligned} \langle \hat{g}_{h'} - g_{h'}, t\rangle_{\Phi(A)} &= \frac{1}{n}\sum_{i=1}^n \int_{\Phi(A)} \{\theta(Y_i)K_{h'}(u - \Phi(X_i)) - \mathbb{E}\left[\theta(Y_i)K_{h'}(u - \Phi(X_i))\right]\} t(u)du \\ &= \nu_{n,h'}(t), \end{aligned}$$

where $\nu_{n,h'}$ is an empirical process. Thus, thanks to (21), it remains to bound the deviations of $\sup_{t \in \bar{S}(0,1)} \nu_{n,h'}^2(t)$. First, we have

$$\begin{aligned} &\mathbb{E}\left[\max_{h' \in \mathcal{H}_n} \left(\sup_{t \in \bar{S}(0,1)} \nu_{n,h'}^2(t) - \frac{V(h')}{3(1 + \|K\|_{L^1(\mathbb{R})}^2)}\right)_+\right] \\ &\leq \sum_{h' \in \mathcal{H}_n} \mathbb{E}\left[\left(\sup_{t \in \bar{S}(0,1)} \nu_{n,h'}^2(t) - \frac{V(h')}{3(1 + \|K\|_{L^1(\mathbb{R})}^2)}\right)_+\right]. \end{aligned}$$

Then, the conclusion results from the following lemma:

**Lemma 5.** *Under the assumptions of Theorem 3, there exists a constant $C$ such that,*

$$\sum_{h \in \mathcal{H}_n} \mathbb{E}\left[\left(\sup_{t \in \bar{S}(0,1)} \nu_{n,h}^2(t) - \tilde{V}(h)\right)_+\right] \leq \frac{C}{n},$$

*with $\tilde{V}(h) = \delta'\|K\|_{L^2(\mathbb{R})}\mathbb{E}[\theta(Y_1)^2]/(nh)$ for a numerical $\delta' > 0$.*

We choose the constant $\kappa$ involved in the definition of $V$ such that $\tilde{V}(h) \leq V(h)(1+\|K\|_{L^1(\mathbb{R})}^2)/3$. Thus, the proof is complete.

□

6.6. **Proof of Lemma 5.** We write the empirical process

$$\nu_{n,h}(t) = \frac{1}{n}\sum_{i=1}^{n}\psi_{t,h}(X_i,Y_i) - \mathbb{E}\left[\psi_{t,h}(X_i,Y_i)\right], \tag{22}$$
$$\text{with } \psi_{t,h}(X_i,Y_i) = \theta(Y_i)\int_{\Phi(A)} K_h\left(u - F_X(X_i)\right)t(u)du.$$

The guiding idea is to apply Talagrand's Inequality (Lemma 6). If $\theta$ is bounded, this inequality can be applied. Otherwise, we have to introduce a truncation.

6.6.1. *Example 1.* Recall that $\Phi = F_X$ and $\Phi(A) = [0;1]$. We split the process $\nu_{n,h}$ into three parts, writing $\nu_{n,h} = \nu_{n,h}^{(1)} + \nu_{n,h}^{(2,1)} + \nu_{n,h}^{(2,2)}$, with, for $l = 1, (2,1), (2,2)$,

$$\nu_{n,h}^{(l)} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{t,h}^{(l)}(Z_i) - \mathbb{E}\left[\varphi_{t,h}^{(l)}(Z_i)\right],$$

$Z_i = X_i$ or $(X_i,\varepsilon_i)$, and

$$\begin{array}{l}\varphi_{t,h}^{(1)} : x \mapsto s(x)\int_0^1 K_h(u - F_X(x))t(u)du,\\ \varphi_{t,h}^{(2,1)} : (x,\varepsilon) \mapsto \varepsilon\mathbf{1}_{|\varepsilon|\le\kappa_n}\int_0^1 K_h(u - F_X(x))t(u)du,\\ \varphi_{t,h}^{(2,2)} : (x,\varepsilon) \mapsto \varepsilon\mathbf{1}_{|\varepsilon|>\kappa_n}\int_0^1 K_h(u - F_X(x))t(u)du,\end{array}$$

where we define, for a constant $c$ which will be specified below,

$$\kappa_n = c\frac{\sqrt{n}}{\ln(n)}. \tag{23}$$

We apply Talagrand's Inequality to the first two bounded empirical processes, and bound roughly the last one. Thus, we split:

$$\begin{aligned}\sum_{h\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)}\nu_{n,h}^2(t) - \tilde{V}(h)\right)_+\right] \le\ & 3\sum_{h\in\mathcal{H}_n}\left\{\mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)}\left(\nu_{n,h}^{(1)}(t)\right)^2 - \frac{\tilde{V}_1(h)}{3}\right)_+\right]\right.\\ &+\mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)}\left(\nu_{n,h}^{(2,1)}(t)\right)^2 - \frac{\tilde{V}_2(h)}{3}\right)_+\right]\\ &\left.+\ \mathbb{E}\left[\sup_{t\in\bar{S}(0,1)}\left(\nu_{n,h}^{(2,2)}(t)\right)^2\right]\right\},\end{aligned} \tag{24}$$

with the decomposition $\tilde{V}(h) = \tilde{V}_1(h) + \tilde{V}_2(h)$, and, denoting by $\delta'' = \delta'/2$,

$$\tilde{V}_1(h) = 3\delta''\frac{\|K\|_{L^2(\mathbb{R})}^2\mathbb{E}\left[s^2(X_1)\right]}{nh},\ \text{and}\ \tilde{V}_2(h) = 3\delta''\frac{\|K\|_{L^2(\mathbb{R})}^2\mathbb{E}\left[\varepsilon_1^2\right]}{nh}.$$

Actually, recall that we have $\mathbb{E}[\theta^2(Y_1)] = \mathbb{E}[Y_1^2] = \mathbb{E}[s^2(X_1)] + \mathbb{E}[\varepsilon_1^2]$ here.

We now show that each of the three terms of the right hand-side of (24) is upper-bounded by a quantity of order $1/n$. This will end the proof.

- ***First term of (24).***

Let us begin with $\nu_{n,h}^{(1)}$. To do so, we compute $H^{(1)}$, $M^{(1)}$ and $v^{(1)}$, involved in Lemma 6.

- For $M^{(1)}$, let $t \in \bar{S}(0,1)$ and $x \in A$ be fixed:

$$\begin{aligned}\left|\varphi_{t,h}^{(1)}(x)\right| &\leq |s(x)| \int_0^1 |K_h(u - F_X(x))t(u)|\, du \leq |s(x)| \|K_h\|_{L^2(\mathbb{R})} \|t\|_{L^2(\Phi(A))}, \\ &= |s(x)| \frac{\|K\|_{L^2(\mathbb{R})}}{\sqrt{h}} \leq \|s\|_{L^\infty(A)} \frac{\|K\|_{L^2(\mathbb{R})}}{\sqrt{h}} := M^{(1)}.\end{aligned}$$

- For $H^{(1)}$, notice that

$$\nu_{n,h}^{(1)}(t) = \langle \hat{d}_h - g_h, t \rangle_{\Phi(A)}, \text{ with } \hat{d}_h = \frac{1}{n} \sum_{i=1}^n s(X_i) K_h \left(. - F_X(X_i)\right).$$

  Thus, thanks to Lemma 7, we obtain,

$$\begin{aligned}\mathbb{E}\left[\sup_{t \in \bar{S}(0,1)} \left(\nu_{n,h}^{(1)}(t)\right)^2\right] &= \mathbb{E}\left[\left\|\hat{d}_h - g_h\right\|^2_{L^2([0;1])}\right], \\ &= \int_0^1 \mathrm{Var}\left(\hat{d}_h(u)\right) du, \text{ since } g_h(u) = \mathbb{E}\left[\hat{d}_h(u)\right], \\ &\leq \int_0^1 \frac{1}{n} \mathbb{E}\left[s^2(X_1) K_h^2\left(u - F_X(X_1)\right)\right] du.\end{aligned}$$

  Then, we use the same computation as the one done to bound the variance term in the proof of Proposition 2, and set $(H^{(1)})^2 = \|K\|^2_{L^2(\mathbb{R})} \mathbb{E}[s^2(X_1)]/(nh)$.

- For $v^{(1)}$, we also fix $t \in \bar{S}(0,1)$. Hereafter, we set $\check{K}_h(u) = K_h(-u)$. First,

$$\mathrm{Var}\left(\varphi_{t,h}^{(1)}(X_1)\right) \leq \mathbb{E}\left[\left(\varphi_{t,h}^{(1)}(X_1)\right)^2\right] \leq \|s\|^2_{L^\infty(A)} \mathbb{E}\left[\left(\int_0^1 K_h(u - F_X(X_1))t(u)du\right)^2\right],$$

  and the expectation can be written

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^1 K_h(u - F_X(X_1))t(u)du\right)^2\right] &= \mathbb{E}\left[\left(\check{K}_h * \left(t\mathbf{1}_{[0;1]}\right)\right)^2 \left(F_X(X_1)\right)\right], \\ &= \int_0^1 \left(\check{K}_h * \left(t\mathbf{1}_{[0;1]}\right)\right)^2 (u)du \leq \left\|\check{K}_h * \left(t\mathbf{1}_{[0;1]}\right)\right\|^2_{L^2(\mathbb{R})}, \\ &\leq \left\|\check{K}_h\right\|^2_{L^1(\mathbb{R})} \|t\mathbf{1}_{[0;1]}\|^2_{L^2(\mathbb{R})} = \left\|\check{K}_h\right\|^2_{L^1(\mathbb{R})} \|t\|^2_{L^2([0;1])},\end{aligned}$$

  thanks to Lemma 8. Therefore,

$$\mathrm{Var}\left(\varphi_{t,h}^{(1)}(X_1)\right) \leq \|s\|_{L^\infty(A)} \|K\|^2_{L^1(\mathbb{R})} := v^{(1)}.$$

Then, Lemma 6 gives, for $\delta > 0$,

$$\mathbb{E}\left[\left(\sup_{t \in \bar{S}(0,1)} \left(\nu_{n,h}^{(1)}(t)\right)^2 - 2(1+2\delta)\left(H^{(1)}\right)^2\right)_+\right] \leq k_1 \left\{\frac{1}{n} \exp\left(-k_2 \frac{1}{h}\right) + \frac{1}{n^2 h} \exp\left(-k_3 \sqrt{n}\right)\right\},$$

where $k_1, k_2, k_3$ are three constants which depend on $\mathbb{E}[s^2(X_1)]$, $\|s\|_{L^\infty(A)}$, $\|K\|_{L^1(\mathbb{R})}$ and $\|K\|_{L^2(\mathbb{R})}$. Assumptions (H2)-(H3) lead to

$$\sum_{h \in \mathcal{H}_n} \mathbb{E}\left[\left(\sup_{t \in \bar{S}(0,1)} \left(\nu_{n,h}^{(1)}(t)\right)^2 - 2(1+2\delta)\|K\|^2_{L^2(\mathbb{R})} \mathbb{E}[s^2(X_1)] \frac{1}{nh}\right)_+\right] \leq \frac{C}{n},$$

with $C$ a constant (which also depends on the previous quantities).

• ***Second term of (24).***

For the second empirical process $\nu_{n,h}^{(2,1)}$, the sketch of the proof is the same: similarly, we compute the quantities involved in the Talagrand Inequality,

$$M^{(2)} = \kappa_n \|K\|_{L^2(\mathbb{R})} \frac{1}{\sqrt{h}}, \;\; H^{(2)} = \|K\|_{L^2(\mathbb{R})} \left(\mathbb{E}[\varepsilon_1^2]\right)^{1/2} \frac{1}{\sqrt{nh}}, \;\; v^{(2)} = \|K\|_{L^1(\mathbb{R})}^2 \mathbb{E}[\varepsilon_1^2],$$

and we obtain, by Lemma 6, for $\delta > 0$,

$$\mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)} \left(\nu_{n,h}^{(2,1)}(t)\right)^2 - 2(1+2\delta)\left(H^{(2)}\right)^2\right)_+\right] \leq k_1 \left\{\frac{1}{n}\exp\left(-k_2\frac{1}{h}\right) + \frac{\kappa_n^2}{n^2 h}\exp\left(-k_3\frac{\sqrt{n}}{\kappa_n}\right)\right\},$$

where $k_1, k_2, k_3$ are three constants which depend on $\mathbb{E}[\varepsilon_1^2]$, $\|K\|_{L^1(\mathbb{R})}$ and $\|K\|_{L^2(\mathbb{R})}$. The first term of the right hand-side is like above. With the definition (23) of $\kappa_n$, the sum over $h \in \mathcal{H}_n$ of the second term of the upper bound can be written

$$\sum_{h\in\mathcal{H}_n} \frac{\kappa_n^2}{n^2 h}\exp\left(-k_3\frac{\sqrt{n}}{\kappa_n}\right) = \frac{c^2}{n^{1+k_3/c}\ln^2(n)} \sum_{h\in\mathcal{H}_n} \frac{1}{h}.$$

Consequently, using Assumptions (H2)-(H3) and choosing $c$ in the definition of $\kappa_n$ such that $c \leq k_3/\alpha_0$, we also obtain for a constant $C$,

$$\sum_{h\in\mathcal{H}_n} \mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)} \left(\nu_{n,h}^{(2,1)}(t)\right)^2 - 2(1+2\delta)\|K\|_{L^2(\mathbb{R})}^2 \mathbb{E}[\varepsilon_1^2]\frac{1}{nh}\right)_+\right] \leq \frac{C}{n}.$$

• ***Third term of (24).***

The last empirical process is $\nu_{n,h}^{(2,2)}(t) = \int_0^1 t(u)\psi(u)du$, with

$$\psi(u) = \frac{1}{n}\sum_{i=1}^n \varepsilon_i \mathbf{1}_{\{|\varepsilon_i|>\kappa_n\}} K_h\left(u - F_X(X_i)\right) - \mathbb{E}\left[\varepsilon_i \mathbf{1}_{\{|\varepsilon_i|>\kappa_n\}} K_h\left(u - F_X(X_i)\right)\right].$$

It is not bounded. Nevertheless, we use the Cauchy-Schwarz Inequality, and the equality $\|t\|_{L^2(\Phi(A))} = 1$, for $t \in \bar{S}(0,1)$

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in\tilde{S}(0,1)} \left(\nu_{n,h}^{(2,2)}(t)\right)^2\right] &\leq \mathbb{E}\left[\int_0^1 \psi^2(u)du\right], \\
&\leq \frac{1}{n}\mathbb{E}\left[\varepsilon_1^2 \mathbf{1}_{\{|\varepsilon_1|>\kappa_n\}}\right] \mathbb{E}\left[\int_0^1 K_h^2\left(u - F_X(X_1)\right)du\right], \\
&\leq \frac{\|K\|_{L^2(\mathbb{R})}^2}{nh}\mathbb{E}\left[\varepsilon_1^2 \mathbf{1}_{\{|\varepsilon_1|>\kappa_n\}}\right] \leq \frac{\|K\|_{L^2(\mathbb{R})}^2 \kappa_n^{-p}}{nh}\mathbb{E}\left[\varepsilon_1^{2+p}\right].
\end{aligned}$$

Thus, there exists a constant $k_1$ which depends on $\|K\|_{L^2(\mathbb{R})}$ and $\mathbb{E}[\varepsilon_1^{2+p}]$,

$$\sum_{h\in\mathcal{H}_n} \mathbb{E}\left[\sup_{t\in\bar{S}(0,1)} \left(\nu_{n,h}^{(2,2)}(t)\right)^2\right] \leq k_1 \frac{\kappa_n^{-p}}{n}\sum_{h\in\mathcal{H}_n}\frac{1}{h} = c_1\kappa^{-p}\frac{\ln^p(n)}{n^{1+p/2}}\sum_{h\in\mathcal{H}_n}\frac{1}{h}.$$

The conclusion comes from Assumptions (H2)-(H3), and the choice of $p \geq 2\alpha_0$.

□

6.6.2. *Examples 2-4.* For the multiplicative regression model (Example 2), we split the process into two terms: $\nu_{n,h} = \nu_{n,h}^{(1)} + \nu_{n,h}^{(2)}$, with

$$\begin{aligned}\nu_{n,h}^{(1)}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\sigma^2(X_i)\varepsilon_i^2\mathbf{1}_{\{|\varepsilon_i|\le\kappa_n\}}\int_0^1 K_h\left(u - F_X(X_i)\right)t(u)du\\ &-\mathbb{E}\left[\sigma^2(X_i)\varepsilon_i^2\mathbf{1}_{\{|\varepsilon_i|\le\kappa_n\}}\int_0^1 K_h\left(u - F_X(X_i)\right)t(u)du\right]\Big\},\\ \nu_{n,h}^{(2)}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\sigma^2(X_i)\varepsilon_i^2\mathbf{1}_{\{|\varepsilon_i|>\kappa_n\}}\int_0^1 K_h\left(u - F_X(X_i)\right)t(u)du\\ &-\mathbb{E}\left[\sigma^2(X_i)\varepsilon_i^2\mathbf{1}_{\{|\varepsilon_i|>\kappa_n\}}\int_0^1 K_h\left(u - F_X(X_i)\right)t(u)du\right]\Big\},\end{aligned}$$

where $\kappa_n$ is still a constant for the proof, which equals $\sqrt{c\frac{\sqrt{n}}{\ln(n)}}$ and $c > 0$ is obtained by the computations, like in Example 1. We exactly recover the framework of this previous example: the deviations of the process $\nu_{n,h}^{(1)}$ are bounded thanks to Talagrand's Inequality of Lemma 6, and the second one is bounded in the same way as the process $\nu_{n,h}^{(2,2)}$ of the additive regression setting.

For Examples 3-4, there is no point in splitting the process (22), since it is already bounded (recall that $\theta(Y_1)$ is bounded by 1). Thus, we apply the concentration inequality.

Recall that $\Phi(A) = \mathbb{R}_+$. In both of these cases, the quantity $M_1$ involved in the assumptions of Lemma 6 equals $M_1 = \|K\|_{L^2(\mathbb{R})}/\sqrt{h}$. Moreover, $H^2$ can be chosen as the upper-bound of the variance term of the estimator $\hat{g}_h$, that is $H^2 = \|K\|_{L^2(\mathbb{R})}/nh$. Finally, $v$ equals $\|K\|_{L^1(\mathbb{R})}$ for Example 3, and $\|g\|_{L^\infty(\mathbb{R}_+)}\|K\|_{L^1(\mathbb{R})}$ for Example 4.

As an example, let us detail the computation of $v$ in Example 4. Recall that $X = C \wedge Z$, $Y = \mathbf{1}_{Z\le C}$, $s$ is the hazard rate, and the warping $\Phi$ is the function $x \mapsto \int_0^x (1 - F_X(t))dt$. Thus, denoting by $f_C$ (respectively $f_Z$) a density of the variable $C$ (respectively $Z$), and $F_C$ (respectively $F_Z$) its c.d.f.,

$$\begin{aligned}\mathrm{Var}\left(\psi_{t,h}(X_1,Y_1)\right) &\le \mathbb{E}\left[\left(\psi_{t,h}(X_1,Y_1)\right)^2\right] = \mathbb{E}\left[Y_1\left(\int_{\mathbb{R}_+}K_h(u' - \Phi(X_1))t(u')du'\right)^2\right],\\ &= \int_{\mathbb{R}_+\times\mathbb{R}}\mathbf{1}_{z\le c}\left(\int_{\mathbb{R}_+}K_h(u' - \Phi(z))t(u')du'\right)^2 f_C(c)f_Z(z)dzdc,\\ &= \int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+}K_h(u' - \Phi(z))t(u')du'\right)^2 f_Z(z)\left(1 - F_C\right)(z)dz.\end{aligned}$$

We set $z = \Phi^{-1}(u)$. The integral becomes

$$\begin{aligned}&\int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+}K_h(u' - \Phi(z))t(u')du'\right)^2 f_Z(z)\left(1 - F_C\right)(z)dz\\ &= \int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+}K_h(u' - u)t(u')du'\right)^2 f_Z\circ\Phi^{-1}(u)\left(1 - F_C\right)\circ\Phi^{-1}(u)\frac{du}{\left((1 - F_X)\circ\Phi^{-1}(u)\right)}.\end{aligned}$$

Thanks to the same arguments as the ones used to prove Proposition 1 in Section 6.2, we obtain:

$$\begin{aligned}\mathrm{Var}\left(\varphi_{t,h}(X_1,Y_1)\right) &\leq \int_{\mathbb{R}_+} g(u)\left(\int_{\mathbb{R}_+} K_h(u'-u)t(u')du'\right)^2 du,\\ &= \int_{\mathbb{R}_+} g(u)\left(K_h * \left(t\mathbf{1}_{\mathbb{R}_+}\right)(u)\right)^2 du \leq \|g\|_{L^\infty(\mathbb{R}_+)}\left\|\check{K}_h * \left(t\mathbf{1}_{\mathbb{R}_+}\right)\right\|_{L^2(\mathbb{R})},\\ &\leq \|g\|_{L^\infty(\mathbb{R}_+)}\left\|\check{K}_h\right\|_{L^1(\mathbb{R})}\left\|\left(t\mathbf{1}_{\mathbb{R}_+}\right)\right\|_{L^2(\mathbb{R})} = \|g\|_{L^\infty(\mathbb{R}_+)}\|K\|_{L^1(\mathbb{R})} := v.\end{aligned}$$

Once we have the three quantities, we easily apply Lemma 6 and the proof is complete by using Assumptions (H2)-(H3), like above (see the computations in Example 1).

□

6.7. **Proof of Corollary 1.** We must bound the bias term of the right hand-side of Inequality (15) (Theorem 3). Actually, if we prove that

$$\|s-s_h\|_\phi^2 \leq Ch^{2\beta},$$

where $C$ is a constant, then the proof of the Corollary will be completed by computing the minimum which is involved in (15). By definition,

$$\|s-s_h\|_\phi^2 = \|g-g_h\|^2_{L^2(\Phi(A))} = \int_{\Phi(A)}\left(g_h(u)-g(u)\right)^2 du.$$

We distinguish two cases in the sequel, depending on the considered examples.

Then we distinguish two cases:

6.7.1. *Examples 1-3.* Here, $\Phi(A)=(0;1)$. We start with the definition of $g_h$: for $u\in\Phi(A)$,

$$\begin{aligned}g_h(u) &= \frac{1}{h}\int_0^1 g(u')K\left(\frac{u-u'}{h}\right)du' = \int_{\frac{u-1}{h}}^{\frac{u}{h}} g(u-hz)K(z)dz,\\ &= \int_{\frac{u-1}{h}}^{\frac{u}{h}} \tilde{g}(u-hz)K(z)dz = \int_{\mathbb{R}} \tilde{g}(u-hz)K(z)dz.\end{aligned}$$

Thus, since $\int_{\mathbb{R}} K(u)du = 1$,

$$\tilde{g}_h(u)-g(u) = \int_{\mathbb{R}} K(z)\tilde{g}(u-hz)dz - \tilde{g}(u) = \int_{\mathbb{R}} K(z)\left[\tilde{g}(u-hz)-\tilde{g}(u)\right]dz. \tag{25}$$

We use a Taylor-Lagrange formula for $\tilde{g}$: for $u\in(0;1)$, and $z\in\mathbb{R}$, there exists $\theta\in(0;1)$ such that

$$\tilde{g}(u-hz)-g(u) = -hz\tilde{g}'(u)+\frac{(-hz)^2}{2!}\tilde{g}''(u)+\cdots+\frac{(-hz)^{l-1}}{(l-1)!}\tilde{g}^{(l-1)}(u)+\frac{(-hz)^l}{l!}\tilde{g}^{(l)}(u-\theta hz),$$

with $l=\lfloor\beta\rfloor$. With Assumption $(K_l)$, we obtain

$$\|s-s_h\|_\phi^2 \leq \left(\int_{z\in\mathbb{R}}|K(z)|\frac{|hz|^l}{l!}\left\{\int_{u=0}^1\left\{\tilde{g}^{(l)}(u-\theta hz)-\tilde{g}^{(l)}(u)\right\}^2 du\right\}^{1/2}dz\right)^2.$$

Since $\tilde{g}$ belongs to the Hölder space $\mathcal{H}(\beta,L)$,

$$\begin{aligned}\left[\int_{u=0}^1\left\{\tilde{g}^{(l)}(u-\theta hz)-\tilde{g}^{(l)}(u)\right\}^2 du\right]^{1/2} &\leq \left[\int_{u=0}^1 L^2(\theta hu)^{2(\beta-l)}du\right]^{1/2},\\ &= L|hz|^{\beta-l},\end{aligned}$$

which enables us to conclude.

□

6.7.2. *Example 4.* Here, $\Phi(A) = \mathbb{R}_+$. Similarly, we first obtain Equality (25). Then, the idea is the same as in Examples 1-3, but since we integrate over an unbounded subset, we choose an integrated remainding term in the Taylor formula:

$$\begin{aligned}\tilde{g}(u-hz)-\tilde{g}(u) &= -hz\tilde{g}'(u)+\frac{(-hz)^2}{2!}\tilde{g}''(u)+\cdots+\frac{(-hz)^{l-1}}{(l-1)!}\tilde{g}^{(l-1)}(u)\\ &\quad+\frac{(-hz)^l}{(l-1)!}\int_0^1(1-\theta)^{l-1}\tilde{g}^{(l)}(u-\theta hz)d\theta.\end{aligned}$$

The reasoning is then the same as in density estimation (see Tsybakov 2009 for details).

□

6.8. **Useful tools.** We recall classical results. The first one is a powerful concentration inequality, which permits to control the deviations of the supremum of an empirical process.

**Lemma 6.** *[Talagrand's Inequality] Let $\xi_1,\ldots,\xi_n$ be i.i.d. random variables, and define $\nu_n(r) = \frac{1}{n}\sum_{i=1}^n r(\xi_i)-\mathbb{E}[r(\xi_i)]$, for $r$ belonging to a countable class $\mathcal{R}$ of real-valued measurable functions. Then, for $\delta>0$, there exist three constants $c_l$, $l=1,2,3$, such that*

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}}\left(\nu_n\left(r\right)\right)^2-c(\delta)H^2\right)_+\right] &\leq c_1\left\{\frac{v}{n}\exp\left(-c_2\delta\frac{nH^2}{v}\right)\right.\\ &\quad\left.+\frac{M_1^2}{C^2(\delta)n^2}\exp\left(-c_3C(\delta)\sqrt{\delta}\frac{nH}{M_1}\right)\right\},\end{aligned}$$

*with, $C(\delta)=(\sqrt{1+\delta}-1)\wedge 1$, $c(\delta)=2(1+2\delta)$ and*

$$\sup_{r\in\mathcal{R}}\|r\|_\infty\leq M_1,\ \mathbb{E}\left[\sup_{r\in\mathcal{R}}|\nu_n(r)|\right]\leq H,\ \text{and}\ \sup_{r\in\mathcal{R}}Var\left(r\left(\xi_1\right)\right)\leq v.$$

Inequality (6) is a classical consequence of the Talagrand Inequality given in Klein and Rio (2005): see for example Lemma 5 (page 812) in Lacour (2008).

Then, we state a lemma which will allow us to replace a $L^2-$norm by the supremum of an empirical process.

**Lemma 7.** *Let $B$ be a borelian subset of $\mathbb{R}$. Denote by $\tilde{S}_B(0,1)$ the set of functions $t\in L^1(B)\cap L^2(B)$ such that $\|t\|_{L^2(B)}=1$. Then, for any function $v\in L^1(B)\cap L^2(B)$,*

$$\|v\|_{L^2(B)}=\sup_{t\in\tilde{S}_B(0,1)}\langle v,t\rangle_B.$$

*Moreover, the supremum over $\tilde{S}_B(0,1)$ equals the supremum over a countable subset $\bar{S}_B(0,1)$ of $\tilde{S}_B(0,1)$.*

*Proof of Lemma 7.* The Cauchy-Schwarz Inequality leads to

$$\sup_{t\in\tilde{S}_B(0,1)}\langle v,t\rangle_B\leq\sup_{t\in\tilde{S}_B(0,1)}\|v\|_{L^2(B)}\|t\|_{L^2(B)}=\|v\|_{L^2(B)}.$$

Besides, if we set $t=v/\|v\|_{L^2(B)}$, then $t$ belongs to $\tilde{S}_B(0,1)$, and $\langle t,v\rangle_B=\|v\|_{L^2(B)}$. This ends the proof of the equality. Finally, we can replace $\tilde{S}_B(0,1)$ by one of its dense countable subset: such a set exists thanks to the separability of $L^2(\mathbb{R})$.

□

Finally, we recall a useful and standard property of the convolution product.

**Lemma 8.** *[Young Inequality] Let $p, q \in [1;\infty[$ such that $1/p + 1/q \geq 1$. If $u \in L^p(\mathbb{R})$ and $v \in L^q(\mathbb{R})$, then the convolution product $u \star v$ exists. Moreover, if $r$ is defined by $1/r = 1/p + 1/q - 1$ then $u \star v \in L^r(\mathbb{R})$ and*

$$\|u \star v\|_{L^r(\mathbb{R})} \leq \|u\|_{L^p(\mathbb{R})} \|v\|_{L^q(\mathbb{R})}.$$

## 7. Appendix 1: additional materials and results for the general case of unknown $\Phi$

In this section, we give some details about the general case of an unknown transformation $\Phi$.

7.1. **Notations.** For the sake of clarity, we begin with an overview of the notations. The warping functions $\Phi$ and their estimator $\hat{\Phi}_n$ are defined in Table 4, for $x \in A$, with $\hat{F}_n$ is the empirical counterpart for $F_X$. But instead of estimating $F_X$ with the whole sample $(X_i)_{i=1,\dots,n}$, we assume that another sample $(X_{-i})_{i=1,\dots,n}$, independent of the $X_i$'s, but distributed like them, is available. Thus, we set

$$\hat{F}_n : \ x \ \mapsto \ \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_{-i} \leq x}.$$

The introduction of the second sample of variable $X$ is an artefact of the theory: it only allows to avoid dependency problems in the proof of the results, which are technical and cumbersome enough (see below). Using a single sample would have required totally different statistic and probabilistic tools. However, we have obviously used only one sample to compute the estimator in the simulation study, see Section 5 (otherwise the comparison with other methods would not have been fair).

| | $\Phi(x)$ | $\hat{\Phi}_n(x)$ |
|---|---|---|
| Examples 1-3 | $F_X(x)$ | $\hat{F}_n(x)$ |
| Example 4 | $\int_0^x (1 - F_X(t))\, dt$ | $\int_0^x \left(1 - \hat{F}_n(t)\right) dt$ |

Table 4. The warping functions and their estimators

Let us now recall the following definitions of the estimators we will studied in this section: for $h > 0$,

$$\hat{s}_h = \hat{g}_h \circ \hat{\Phi}_n, \quad \text{with } \hat{g}_h \ : \ u \ \mapsto \ \frac{1}{n} \sum_{i=1}^{n} \theta(Y_i) K_h \left(u - \hat{\Phi}_n(X_i)\right). \tag{26}$$

We aim at providing an upper bound for the risk of this estimator. The main challenge of the plug-in device is to bound the difference $\hat{\Phi}_n - \Phi$, in order to come down to the risk of the estimator with known $\Phi$, bounded in Proposition 2.

7.2. **Properties of $\Phi$ and $\hat{\Phi}_n$.** We introduce in this section properties and deviation inequalities which will be repeatedly used in the proof of Proposition 11.

First, we set $U_i = F_X(X_{-i})$, for $i \in \{1, \dots, n\}$ and denote by $\hat{U}_n$ the empirical c.d.f related to the sample $(U_i)_{1\leq i\leq n}$. Recall first that $U_i$ has a uniform distribution on the interval $(0;1)$ and that $\hat{F}_n \circ F_X^{-1} = \hat{U}_n$. The variable $\sup_{u\in\mathbb{R}} \left|\hat{U}_n(u) - u\right|$, which will be denoted by $\|\hat{U}_n - id\|_{L^\infty(\mathbb{R})}$, has the same distribution as $\sup_{x\in A} \left|\hat{F}_n(x) - F_X(x)\right| := \|\hat{F}_n - F_X\|_{L^\infty(\mathbb{R})}$. The following deviation result is known as the Dvoretzky-Kiefer-Wolfowitz Inequality (see Dvoretzky et al. (1956)):

**Proposition 9.** *There exists a constant $C > 0$, such that, for any integer $n \geq 1$, and all $\lambda > 0$,*

$$\mathbb{P}\left(\|\hat{U}_n - id\|_{L^\infty(\mathbb{R})} \geq \lambda\right) \leq C \exp\left(-2n\lambda^2\right).$$

By integration, we deduce the following bounds:

**Corollary 2.** *For any $p \in \mathbb{N}\backslash\{0\}$, there exists a constant $C_p > 0$ such that*

$$\mathbb{E}\left[\left\|\hat{U}_n - id\right\|^p_{L^\infty(\mathbb{R})}\right] \leq \frac{C_p}{n^{p/2}}.$$

Proposition 9 and Corollary 2 are sufficient to handle the case $\Phi = F_X$, $\hat{\Phi}_n = \hat{F}_n$ (Examples 1-3). In Example 4, recall that the deformation is $\Phi(x) = \int_0^x (1 - F_X(t))dt$. Before studying the deviations of its empirical counterpart, we first state the following equalities, which will be useful, even though simple:

**Lemma 10.** *Denote by $\Phi(x) = \int_0^x (1 - F_X(t))dt$, $x \in (0;B)$, and $\Phi' = \phi$. The function $\Phi$ satisfies*

(1) $\Phi(x) = \displaystyle\int_0^B (y \wedge x) f_X(y)dy$, $x \in (0;B)$,
(2) $\mathbb{E}[X] = \int_0^B \phi(x)dx$.

***Proof of Lemma 10***

(1) Fix $x \in (0;B)$, and compute

$$\begin{aligned}\Phi(x) &= \int_0^x \mathbb{P}(X > t)dt = \int_0^x \left(\int_0^B \mathbf{1}_{y>t} f_X(y)dy\right) dt,\\ &= \int_0^B f_X(y) \left(\int_0^x \mathbf{1}_{y>t} dt\right) dy = \int_0^B f_X(y)\,(y \wedge x)\, dy.\end{aligned}$$

(2) To recover the expectation of $X$, we compute

$$\begin{aligned}\int_0^B \phi(x)dx &= \int_0^B (1 - F_X(x))\, dx = \int_0^B \left(\int_x^\infty f_X(t)dt\right) dx,\\ &= \int_0^\infty f_X(t) \int_0^B \mathbf{1}_{t>x} dxdt = \int_0^\infty f_X(t)tdt = \mathbb{E}[X].\end{aligned}$$

□

Now, consider the estimator $\hat{\Phi}_n = \int_0^x (1-\hat{F}_n(t))dt$. It can also be written $\hat{\Phi}_n(x) = \frac{1}{n}\sum_{i=1}^n X_{-i} \wedge x$. Thus it has no bias for the estimation of $\Phi$ (see the first point of Lemma 10). Moreover, assuming that $A = (0;B)$ with finite $B > 0$ (still in Example 4) permits to write,

$$\sup_{x\in(0;B)} \left|\hat{\Phi}_n(x) - \Phi(x)\right| \leq B \sup_{x\in\mathbb{R}_+} \left|\hat{F}_n(x) - F_X(x)\right|. \tag{27}$$

Thus, Proposition 9 and Corollary 2 are also useful to bound $\hat{\Phi}_n - \Phi$ in Example 4.
To standardize the notations in the proofs below, we finally set $B = 1$ when $\Phi = F_X$ and $\hat{\Phi}_n = \hat{F}_n$, that is in Examples 1-3. Therefore, Inequality (27) holds whatever $\Phi$ is.

7.3. **Result.** Before stating the result, we introduce new assumptions:

(H1') the function $s$ is continuously derivable on $A$.
(H2') the kernel $K$ is twice continuously derivable, with bounded derivatives $K'$ and $K''$ on $\mathbb{R}$.

Assumption (H1') is somehow restrictive but required for integration by parts (see Section 7.4.3). Assumption (H2') permits to use Taylor formulas to deal with the difference $K_h(u - \hat{\Phi}_n(X_i)) - K_h(u - \Phi(X_i))$. This is not a problem as we choose the kernel.

We now illustrate how the plug-in device suits well to recover the function $s$ by providing an upper bound for the risk of $\hat{s}_h$.

**Proposition 11.** *Assume* (H1') *and* (H2'). *Assume also that* $A = (0; B)$ *with* $B < \infty$ *for Example 4. Then, there exist three constants* $c_1$, $c_2$ *and* $c_3$ *such that*

$$\mathbb{E}\left[\|\hat{s}_h - s\|_\phi^2\right] \leq 5\,\|g - g_h\|^2_{L^2(\Phi(A))} + c_1\frac{1}{nh} + c_2\frac{1}{n^2h^4} + c_3\frac{1}{n^2h^6}. \tag{28}$$

*If moreover* $h > n^{-1/4}$, *there exists* $c > 0$ *such that*

$$\mathbb{E}\left[\|\hat{s}_h - s\|_\phi^2\right] \leq 5\,\|g - g_h\|^2_{L^2(\Phi(A))} + c\frac{1}{nh}. \tag{29}$$

Notice that the additional assumption $A = (0; B)$ with $B < \infty$ is needed to control the difference $\hat{\Phi}_n - \hat{\Phi}$ in Example 4 (see Section above). Inequality (29) is immediatly deduced from (28) with the additional assumption $h > n^{-1/4}$. It shows that the same result as Proposition 2 holds when the warping function $\Phi$ is unknown, under mild assumptions. The main adaptive result (Theorem 3) in this general framework can then be deduced from this bound.

7.4. **Proof of Proposition 11.**

7.4.1. *Main part of the proof of Proposition 11.* Let us specify the notations of this section. Our goal is to study the risk of $\hat{s}$ defined by (26): we denote it by $\hat{s}^{\hat{\Phi}_n,\hat{\Phi}_n}$. We have

$$\hat{s}^{\hat{\Phi}_n,\hat{\Phi}_n} = \hat{g}_h^{\hat{\Phi}_n} \circ \hat{\Phi}_n, \text{ with } \hat{g}_h^{\hat{\Phi}_n}(u) = \frac{1}{n}\sum_{i=1}^n \theta(Y_i) K_h\left(u - \hat{\Phi}_n(X_i)\right)$$

Moreover, $\hat{s}^{\Phi,\Phi} = \hat{g}_h^{\Phi} \circ \Phi$ with $\hat{g}_h^{\Phi}(u) = (1/n)\sum_{i=1}^n \theta(Y_i) K_h(u - \Phi(X_i))$ is the estimator studied in the main part of the paper. Coherently, we also introduce $\hat{s}^{\hat{\Phi}_n,\Phi} = \hat{g}_h^{\hat{\Phi}_n} \circ \Phi$. The following decomposition is the key of the proof:

$$\|\hat{s}_h - s\|_\phi^2 \leq 5\sum_{l=0}^3 T_l^h,$$

with

$$\begin{aligned}
T_0^h &= \left\|\hat{s}_h^{\Phi,\Phi} - s_h\right\|_\phi^2 + \left\|s_h^{\Phi} - s\right\|_\phi^2, \\
T_1^h &= \left\|\hat{s}_h^{\hat{\Phi}_n,\Phi} - \hat{s}_h^{\Phi,\Phi} - \mathbb{E}\left[\hat{s}_h^{\hat{\Phi}_n,\Phi} - \hat{s}_h^{\Phi,\Phi}\,|(X_{-i})\right]\right\|_\phi^2, \\
T_2^h &= \left\|\hat{s}_h^{\hat{\Phi}_n,\hat{\Phi}_n} - \hat{s}_h^{\hat{\Phi}_n,\Phi} - \mathbb{E}\left[\hat{s}_h^{\hat{\Phi}_n,\hat{\Phi}_n} - \hat{s}_h^{\hat{\Phi}_n,\Phi}\,|(X_{-i})\right]\right\|_\phi^2, \\
T_3^h &= \left\|\mathbb{E}\left[\hat{s}_h^{\hat{\Phi}_n,\hat{\Phi}_n} - \hat{s}_h^{\Phi,\Phi}\,|(X_{-i})\right]\right\|_\phi^2,
\end{aligned}$$

where $\mathbb{E}[Z|(X_{-i})]$ is the conditional expectation of a variable $Z$ given the sample $(X_{-i})_{i=1,\dots,n}$. The term $T_0^h$ has been bounded in Proposition 2. For the three others, we set the following lemmas, which end the proof.

**Lemma 12.** *Under the assumptions of Proposition 11,*

$$\mathbb{E}\left[T_l^h\right] \leq \kappa_1 \mathbb{E}\left[\theta^2(Y_1)\right] \|K'\|_{L^\infty(\mathbb{R})} C_2 \frac{1}{n^2 h^4},$$

*with $\kappa_1 = 1$ in Examples 1-3, $\kappa_1 = B^2\mathbb{E}[X]$ in Example 4, and $C_2$ defined by Proposition 2.*

**Lemma 13.** *Under the assumptions of Proposition 11,*

$$\mathbb{E}\left[T_3^h\right] \leq \frac{\kappa}{nh} + 16\kappa_1 \mathbb{E}\left[\theta^2(Y_1)\right] \left\|K''\right\|^2_{L^\infty(\mathbb{R})} C_4 \frac{1}{n^2 h^6}.$$

*with $\kappa_1 = B^2\mathbb{E}[X]$ and*

$$\kappa = \begin{cases} 12\|K\|^2_{L^2(\mathbb{R})}\left(2C_2\|s\|^2_{L^\infty(A)} + 2\|s\|^2_\phi + \left(\frac{1}{4} + 3C_2\right)\|s'\|^2_{L^2(A)}\right) & \textit{(Examples 1-3),} \\ \|K\|^2_{L^2(\mathbb{R})}\Big(\|s\|^2_{L^\infty(A)}(\mathbb{E}[X_{-1}^2] + 2C_2B^2) & \\ \quad + \|s'\|^2_{L^1(A)}(\mathbb{E}[X_{-1}^2] + C_2B^2) + \mathbb{E}[X_{-1}]\|s\|^2_{L^2(A)}\Big) & \textit{(Example 4).} \end{cases}$$

**Remark 1.** To prove Lemma 12, Assumption (H1') is not required and Assumption (H2') can be weakened: it is sufficient to assume that the kernel $K$ is continuously derivable, with bounded derivative $K'$.

□

7.4.2. *Proof of Lemma 12.* The first term to bound is

$$T_1^h = \int_A \left(\hat{g}_h^{\hat{\Phi}_n} \circ \Phi(x) - \hat{g}_h^{\Phi} \circ \Phi(x) - \mathbb{E}\left[\hat{g}_h^{\hat{\Phi}_n} \circ \Phi(x) - \hat{g}_h^{\Phi} \circ \Phi(x) \,|(X_{-i})\right]\right)^2 \phi(x)dx,$$

and its conditional expectation is

$$\mathbb{E}\left[T_1^h \,|(X_{-i})\right] = \int_A \mathrm{Var}\left(\hat{g}_h^{\hat{\Phi}_n}(\Phi(x)) - \hat{g}_h^{\Phi}(\Phi(x)) \,|(X_{-i})\right) \phi(x)dx.$$

For any $x \in A$, we compute

$$\begin{aligned} &\mathrm{Var}\left(\hat{g}_h^{\hat{\Phi}_n}(\Phi(x)) - \hat{g}_h^{\Phi}(\Phi(x)) \,|(X_{-i})\right) \\ &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \theta(Y_i)\left\{K_h\left(\Phi(x) - \hat{\Phi}_n(X_i)\right) - K_h\left(\Phi(x) - \Phi(X_i)\right)\right\}|(X_{-i})\right), \\ &= \frac{1}{n}\mathrm{Var}\left(\theta(Y_1)\left\{K_h\left(\Phi(x) - \hat{\Phi}_n(X_1)\right) - K_h\left(\Phi(x) - \Phi(X_1)\right)\right\}|(X_{-i})\right), \\ &\leq \frac{1}{nh^2}\mathbb{E}\left[\theta^2(Y_1)\left\{K\left(\frac{\Phi(x) - \hat{\Phi}_n(X_1)}{h}\right) - K\left(\frac{\Phi(x) - \Phi(X_1)}{h}\right)\right\}^2 |(X_{-i})\right]. \end{aligned}$$

The mean value theorem for the kernel $K$ between to real numbers $a$ and $b$ leads to:

$$|K(b) - K(a)| \leq \|K'\|_{L^\infty(\mathbb{R})}|b - a|.$$

By choosing $a = (\Phi(x) - \hat{\Phi}_n(X_1))/h$ and $b = (\Phi(x) - \Phi(X_1))/h$, we obtain

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{g}_h^{\hat{\Phi}_n}(\Phi(x)) - \hat{g}_h^{\Phi}(\Phi(x))\,|(X_{-i})\right) \\
&\qquad \leq \frac{1}{nh^2}\mathbb{E}\left[\theta^2(Y_1)\|K'\|^2_{L^\infty(\mathbb{R})}\frac{\left\{\Phi(X_i) - \hat{\Phi}_n(X_i)\right\}^2}{h^2}\,|(X_{-i})\right], \\
&\qquad \leq \frac{1}{nh^4}\|K'\|^2_{L^\infty(\mathbb{R})}\left\|\hat{\Phi}_n - \Phi\right\|^2_{L^\infty(A)}\mathbb{E}\left[\theta^2(Y_1)\,|(X_{-i})\right], \\
&\qquad = \frac{1}{nh^4}\|K'\|^2_{L^\infty(\mathbb{R})}\left\|\hat{\Phi}_n - \Phi\right\|^2_{L^\infty(A)}\mathbb{E}\left[\theta^2(Y_1)\right].
\end{aligned}
$$

Thus,

$$
\mathbb{E}\left[T_1^h\,|(X_{-i})\right] \leq \frac{1}{nh^4}\int_A \phi(x)dx\|K'\|^2_{L^\infty(\mathbb{R})}\left\|\hat{\Phi}_n - \Phi\right\|^2_{L^\infty(A)}\mathbb{E}\left[\theta^2(Y_1)\right],
$$

and consequently

$$
\mathbb{E}\left[T_1^h\right] \leq \frac{1}{nh^4}\int_A \phi(x)dx\|K'\|^2_{L^\infty(\mathbb{R})}\mathbb{E}\left[\left\|\hat{\Phi}_n - \Phi\right\|^2_{L^\infty(A)}\right]\mathbb{E}\left[\theta^2(Y_1)\right].
$$

With Inequalities (27) and (2) successively, we obtain

$$
\mathbb{E}\left[T_1^h\right] \leq \frac{B^2C_2}{n^2h^4}\mathbb{E}\left[\theta^2(Y_1)\right]\int_A \phi(x)dx\|K'\|^2_{L^\infty(\mathbb{R})}.
$$

To conclude, it remains to compute $\int_A \phi(x)dx$: it equals 1 in Examples 1-3 and $\mathbb{E}[X]$ in Example 4 (see Lemma 10).

To deal with the second term, we first write

$$
T_2^h = \int_A \left(\hat{g}_h^{\hat{\Phi}_n}\circ\hat{\Phi}_n(x) - \hat{g}_h^{\hat{\Phi}_n}\circ\Phi(x) - \mathbb{E}\left[\hat{g}_h^{\hat{\Phi}_n}\circ\hat{\Phi}_n(x) - \hat{g}_h^{\hat{\Phi}_n}\circ\Phi(x)\,|(X_{-i})\right]\right)^2\phi(x)dx.
$$

We now argue as for $T_1^h$:

$$
\begin{aligned}
\mathbb{E}\left[T_2^h\,|(X_{-i})\right] &= \int_A \mathrm{Var}\left(\hat{g}_h^{\hat{\Phi}_n}\circ\hat{\Phi}_n(x) - \hat{g}_h^{\hat{\Phi}_n}(\Phi(x))\,|(X_{-i})\right)\phi(x)dx, \\
&\leq \int_A \frac{1}{nh^2}\mathbb{E}\left[\theta^2(Y_1)\left\{K\left(\frac{\hat{\Phi}_n(x) - \hat{\Phi}_n(X_1)}{h}\right)\right.\right. \\
&\qquad \left.\left. -K\left(\frac{\Phi(x) - \hat{\Phi}_n(X_1)}{h}\right)\right\}^2\,|(X_{-i})\right]\phi(x)dx, \\
&\leq \int_A \frac{1}{nh^2}\mathbb{E}\left[\theta^2(Y_1)\|K'\|^2_{L^\infty(\mathbb{R})}\frac{\left\{\hat{\Phi}_n(x) - \Phi(x)\right\}^2}{h^2}\,|(X_{-i})\right]\phi(x)dx, \\
&\leq \int_A \phi(x)dx\frac{1}{nh^4}\mathbb{E}\left[\theta^2(Y_1)\right]\|K'\|_{L^\infty(\mathbb{R})}\left\|\hat{\Phi}_n - \Phi\right\|^2_{L^\infty(A)}.
\end{aligned}
$$

The conclusion is the same as for the first term.

□

7.4.3. *Proof of Lemma 13.* We have

$$\begin{aligned}
T_3^h &= \mathbb{E}\left[\int_A \left(\hat{s}^{\hat{\Phi}_n,\hat{\Phi}_n}(x) - \hat{s}^{\Phi,\Phi}(x)\right)^2 \phi(x)dx\,|(X_{-i})\right], \\
&= \int_A \left(\mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^n \theta(Y_i)\left\{K\left(\frac{\hat{\Phi}_n(x)-\hat{\Phi}_n(X_i)}{h}\right) - K\left(\frac{\Phi(x)-\Phi(X_i)}{h}\right)\right\}|(X_{-i})\right]\right)^2 \phi(x)dx, \\
&= \int_A \left(\mathbb{E}\left[\frac{1}{h}\theta(Y_1)\left\{K\left(\frac{\hat{\Phi}_n(x)-\hat{\Phi}_n(X_1)}{h}\right) - K\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)\right\}|(X_{-i})\right]\right)^2 \phi(x)dx,
\end{aligned}$$

since $(X_i, Y_i)_i$ are *i.i.d.* given the sample $(X_{-i})_i$. We apply the Taylor formula with Lagrange form for the remainder term $K$: for $b = (\hat{\Phi}_n(x) - \hat{\Phi}_n(X_1))/h$ and $a = (\Phi(x) - \Phi(X_1))/h$, there exists $\hat{\alpha}_{n,x,h}$ between $a$ and $b$ such that

$$K(b) - K(a) = (b-a)K'(a) + (b-a)^2\frac{K''(\hat{\alpha}_{n,x,h})}{2}.$$

This leads to the decomposition $T_3^h \le 2T_{3,1}^h + 2T_{3,2}^h$, with

$$\begin{aligned}
T_{3,1}^h = \int_A \Bigg(\mathbb{E}\Bigg[\frac{1}{h}\theta(Y_1)\Bigg\{&\frac{\hat{\Phi}_n(x) - \hat{\Phi}_n(X_1) - \Phi(x) + \Phi(X_1)}{h} \\
&\times K'\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)\Bigg\}|(X_{-i})\Bigg]\Bigg)^2 \phi(x)dx, \\
T_{3,2}^h = \int_A \left(\mathbb{E}\left[\frac{1}{h}\theta(Y_1)\left\{\frac{\left(\hat{\Phi}_n(x) - \hat{\Phi}_n(X_1) - \Phi(x) + \Phi(X_1)\right)^2}{h^2}K''(\hat{\alpha}_{n,x,h})\right\}|(X_{-i})\right]\right)^2 &\phi(x)dx.
\end{aligned}$$

We now bound each of the terms.

• **Upper-bound for $\mathbf{T_{3,1}^h}$.**

Let us begin with the splitting $T_{3,1}^h \le 2T_{3,1,1}^h + 2T_{3,1,2}^h$, where

$$\begin{aligned}
T_{3,1,1}^h &= \int_A \left(\mathbb{E}\left[\frac{1}{h}\theta(Y_1)\left\{\frac{\left(\hat{\Phi}_n(x) - \Phi(x)\right)}{h}K'\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)\right\}|(X_{-i})\right]\right)^2 \phi(x)dx, \\
T_{3,1,2}^h &= \int_A \left(\mathbb{E}\left[\frac{1}{h}\theta(Y_1)\left\{\frac{\left(\hat{\Phi}_n(X_1) - \Phi(X_1)\right)}{h}K'\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)\right\}|(X_{-i})\right]\right)^2 \phi(x)dx.
\end{aligned}$$

⋆ **Upper-bound for $\mathbf{T_{3,1,1}^h}$.**

We have

$$\begin{aligned}
T_{3,1,1}^h &= \frac{1}{h^2}\int_A \left(\frac{\left(\hat{\Phi}_n(x) - \Phi(x)\right)}{h^2}\mathbb{E}\left[\theta(Y_1)K'\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)|(X_{-i})\right]\right)^2 \phi(x)dx, \\
&= \frac{1}{h^2}\int_A \frac{\left(\hat{\Phi}_n(x) - \Phi(x)\right)^2}{h^2}\left(\mathbb{E}\left[\theta(Y_1)K'\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)\right]\right)^2 \phi(x)dx.
\end{aligned}$$

We now need a slightly different version of Proposition 1. We replace $K_h(u-\Phi(X))$ by any function $t\circ\Phi(X)$ such that the expectation exists, in the proof of Section 6.2, to obtain:

$$\mathbb{E}\left[\theta(Y_1)t\circ\Phi(X_1)\right]=\int_{\Phi(A)}t(u')g(u')du'=\int_A t(\Phi(x'))s(x')\phi(x')dx', \tag{30}$$

where the last inequality is the consequence of the change of variable $x=\Phi^{-1}(x')$. Here, with $t=K'((\Phi(x)-\Phi(X_1))/h)$, (30) leads to

$$\begin{aligned} T^h_{3,1,1} &= \frac{1}{h^2}\int_A\left(\hat\Phi_n(x)-\Phi(x)\right)^2\left(\int_A\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)^2\phi(x)dx,\\ &\le \frac{1}{h^2}\int_A\left\|\hat\Phi_n-\Phi\right\|^2_{L^\infty(A)}\left(\int_A\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)^2\phi(x)dx, \end{aligned}$$

We anew apply Inequalities (27) and (2), which give

$$\mathbb{E}\left[T^h_{3,1,1}\right]\le\frac{B^2C_2}{nh^2}\int_A\left(J_x\right)^2\phi(x)dx,\ \text{with } J_x=\int_A\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'.$$

We now discuss two cases, depending on the value of $\Phi(A)$ (see Table 4).

- **Example 1-3 ($\Phi(A)=(0;1)$).** By the changes of variables $u=\Phi(x)$ and $u'=\Phi(x')$, the last inequality can be written

$$\mathbb{E}\left[T^h_{3,1,1}\right]\le\frac{C_2}{nh^2}\int_{(0;1)}\left(I_u\right)^2du,\ \text{avec } I_u=\int_{(0;1)}\frac{1}{h}K'\left(\frac{u-u'}{h}\right)g(u')du'.$$

  An integration by parts permits to compute $I_u$ ($s$, and thus $g$ are assumed to be continuously derivable):

$$\begin{aligned} I_u &= \left[-g(u')K\left(\frac{u-u'}{h}\right)\right]_0^1+\int_{(0;1)}K\left(\frac{u-u'}{h}\right)g'(u')du',\\ &= -g(1)K\left(\frac{u-1}{h}\right)+g(0)K\left(\frac{u}{h}\right)+\int_{(0;1)}K\left(\frac{u-u'}{h}\right)g'(u')du'. \end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[T^h_{3,1,1}\right] &\leq \frac{C_2}{nh^2}\int_{(0;1)}\left(-g(1)K\left(\frac{u-1}{h}\right)+g(0)K\left(\frac{u}{h}\right)+\int_{(0;1)}K\left(\frac{u-u'}{h}\right)g'(u')du'\right)^2 du,\\
&= \frac{C_2}{nh}\int_{(0;h)}\left(-g(1)K\left(v-\frac{1}{h}\right)+g(0)K\left(v\right)+\int_{(0;1)}K\left(v-\frac{u'}{h}\right)g'(u')du'\right)^2 du,\\
&\leq \frac{3C_2}{nh}\left\{g^2(1)\int_{(0;1)}K^2\left(v-\frac{1}{h}\right)dv+g^2(0)\int_{(0;1)}K^2(v)dv\right.\\
&\quad \left.+\int_{(0;1)}\left(g'(u')\right)^2du'\int_{(0;1)}\left(\int_{(0;1)}K^2\left(v-\frac{u'}{h}\right)du'\right)dv\right\},\\
&\leq \frac{3}{C_2nh}\left\{g^2(1)\|K\|^2_{L^2(\mathbb{R})}+g^2(0)\|K\|^2_{L^2(\mathbb{R})}+\int_{(0;1)}\left(g'(u')\right)^2du'\|K\|^2_{L^2(\mathbb{R})}\right\},\\
&\leq \frac{3C_2}{4nh}\left(2\|g\|^2_{L^\infty((0;1))}+\|g'\|^2_{L^2((0;1))}\right)\|K\|^2_{L^2(\mathbb{R})},\\
&\leq \frac{3C_2}{4nh}\left(2\|s\|^2_{L^\infty(A)}+\|s'\|^2_{L^2(A)}\right)\|K\|^2_{L^2(\mathbb{R})}
\end{aligned}$$

- **Example 4** ($A=(0;B)$). We also integrate by parts:

$$\begin{aligned}
J_x &= \left[-s(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\right]_0^B+\int_0^B s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx',\\
&= -s(B)K\left(\frac{\Phi(x)-\Phi(B)}{h}\right)+s(0)K\left(\frac{\Phi(x)}{h}\right)+\int_0^B s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'.
\end{aligned}$$

With similar computations as in the previous examples, but with variables $x$ and $x'$ in the integrals,

$$\begin{aligned}
\mathbb{E}\left[T^h_{3,1,1}\right] &\leq \frac{3C_2B^2}{nh^2}\left\{s^2(B)\int_0^B K^2\left(\frac{\Phi(x)-\Phi(B)}{h}\right)\phi(x)dx\right.\\
&\quad \left.+s^2(0)\int_0^B K^2\left(\frac{\Phi(x)}{h}\right)\phi(x)dx+\int_0^B\left(\int_0^B s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\phi(x)dx\right\}.
\end{aligned}$$

There are three terms to bound. First,

$$\begin{aligned}
&s^2(B)\int_0^B K^2\left(\frac{\Phi(x)-\Phi(B)}{h}\right)\phi(x)dx\leq\|s\|^2_{L^\infty(A)}\|K\|^2_{L^2(\mathbb{R})},\\
&s^2(0)\int_0^B K^2\left(\frac{\Phi(x)}{h}\right)\phi(x)dx\leq\|s\|^2_{L^\infty(A)}\|K\|^2_{L^2(\mathbb{R})}.
\end{aligned}$$

For the most complicated term, we apply the generalized Minkowski Inequality (see Lemma 1.1 p.13 in Tsybakov 2009):

$$\begin{aligned}
&\int_0^B\left(\int_0^B s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\phi(x)dx\\
&\qquad\leq \left[\int_0^B|s'(x')|\left(\int_0^B K^2\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\phi(x)dx\right)^{1/2}dx'\right]^2.
\end{aligned}$$

Moreover,

$$\begin{aligned}\left(\int_0^B K^2\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\phi(x)dx\right)^{1/2} &= \left(\int_0^{\Phi(B)} K^2\left(\frac{u-\Phi(x')}{h}\right)du\right)^{1/2},\\ &= \left(\int_0^{\Phi(B)/h} K^2\left(v-\frac{\Phi(x')}{h}\right)hdv\right)^{1/2},\\ &\leq \sqrt{h}\|K\|_{L^2(\mathbb{R})},\end{aligned}$$

and therefore

$$\begin{aligned}\int_0^B\left(\int_0^B s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\phi(x)dx &\leq h\|K\|^2_{L^2(\mathbb{R})}\left[\int_0^B s'(x')dx'\right]^2,\\ &\leq h\|K\|^2_{L^2(\mathbb{R})}\|s'\|^2_{L^1(A)}.\end{aligned}$$

At the end,

$$\mathbb{E}\left[T^h_{3,1,1}\right] \leq \frac{3C_2B^2}{nh}\|K\|^2_{L^2(\mathbb{R})}\left\{2\|s\|^2_{L^\infty(A)}+\|s'\|^2_{L^1(A)}\right\}.$$

Finally, whatever the example we consider, we have shown that

$$\mathbb{E}\left[T^h_{3,1,1}\right] \leq \frac{3C_2B^2}{nh}\|K\|^2_{L^2(\mathbb{R})}\left\{2\|s\|^2_{L^\infty(A)}+C_s\right\}, \tag{31}$$

with $C_s=\|s'\|^2_{L^2(A)}$ (Examples 1-3), and $C_s=\|s'\|^2_{L^1(A)}$ (Example 4).

⋆ **Upper-bound for $\mathbf{T^h_{3,1,2}}$.**
Recall first the definition of this term:

$$T^h_{3,1,2}=\int_A\left(\mathbb{E}\left[\frac{1}{h}\theta(Y_1)\left\{\frac{\left(\hat\Phi_n(X_1)-\Phi(X_1)\right)}{h}K'\left(\frac{\Phi(x)-\Phi(X_1)}{h}\right)\right\}|(X_{-i})\right]\right)^2\phi(x)dx.$$

By using that $\hat\Phi_n$ is measurable with respect to the sample $(X_{-i})_i$, and by noticing that the $X_i$'s and $Y_i$'s are independent of $(X_{-i})_i$, we can derive a conditional version of (30): for any function $t$ for which the expectation exists,

$$\mathbb{E}\left[\theta(Y_1)t\left(\hat\Phi_n(X_1),\Phi(X_1)\right)|(X_{-i})\right]=\int_A t\left(\hat\Phi_n(x'),\Phi(x')\right)s(x')\phi(x')dx'.$$

This leads to

$$T^h_{3,1,2}=\frac{1}{h^2}\int_A\left(\int_A\frac{\left(\hat\Phi_n(x')-\Phi(x')\right)}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)^2\phi(x)dx.$$

Now, $\hat\Phi_n(x')$ is an empirical mean of variables with expectation $\Phi(x')$:

- If $\hat\Phi_n$ is the empirical c.d.f, it is the mean of $\mathbf{1}_{X_{-i}\leq x'}$, with expectation $F_X(x')$,
- If $\hat\Phi_n(x)=\int_0^x(1-\hat F_n(t))dt$, it is the mean of $X_{-i}\wedge x'$, with expectation $\Phi(x')$ (see Lemma 10).

Thus, we have

$$\mathbb{E}\left[T^h_{3,1,2}\right] = \frac{1}{h^2}\int_A \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \int_A T_{i,x'}\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)\phi(x)dx,$$

where $T_{i,x'} = \mathbf{1}_{X_{-i}\leq x'}$ for Examples 1-3, and $T_{i,x'} = X_{-i}\wedge x'$ for Example 4 (see Table 4 and Lemma 10). Then,

$$\begin{aligned}\mathbb{E}\left[T^h_{3,1,2}\right] &= \frac{1}{nh^2}\int_A \mathrm{Var}\left(\int_A T_{1,x'}\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)\phi(x)dx,\\ &= \frac{1}{nh^2}\int_A \mathbb{E}\left[\left(\int_A \left(T_{1,x'}-\mathbb{E}\left[T_{1,x}\right]\right)\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)^2\right]\phi(x)dx.\end{aligned}$$

We must now separate the two cases, depending on the definition of $T_{1,x}$.

- **Examples 1-3**.

We set $u=\Phi(x)$ and $u'=\Phi(x')$ in the integrals. With $F_X(X_{-1})=U_1$, the last inequality becomes

$$\mathbb{E}\left[T^h_{3,1,2}\right] \leq \frac{1}{nh^2}\int_{(0;1)}\mathbb{E}\left[\left(I_{U_1,u}\right)^2\right]du,$$

$$\text{with } I_{U_1,u} = \int_0^1 \left(\mathbf{1}_{U_1\leq u'}-u\right)g(u')\frac{1}{h}K'\left(\frac{u-u'}{h}\right)du'.$$

An integration by parts leads to:

$$\begin{aligned}I_{U_1,u} &= \int_{U_1}^1 g(u')\frac{1}{h}K'\left(\frac{u-u'}{h}\right)du' - \int_0^1 u'g(u')\frac{1}{h}K'\left(\frac{u-u'}{h}\right)du',\\ &= \left[-g(u')K\left(\frac{u-u'}{h}\right)\right]_{U_1}^1 + \int_{U_1}^1 g'(u')K\left(\frac{u-u'}{h}\right)du' + \left[u'g(u')K\left(\frac{u-u'}{h}\right)\right]_0^1\\ &\quad - \int_0^1 \left[g(u')+u'g'(u')\right]K\left(\frac{u-u'}{h}\right)du',\\ &= g(U_1)K\left(\frac{u-U_1}{h}\right) + \int_0^1 g'(u')K\left(\frac{u-u'}{h}\right)\mathbf{1}_{U_1\leq u'}du'\\ &\quad - \int_0^1 \left[g(u')+u'g'(u')\right]K\left(\frac{u-u'}{h}\right)du',\\ &= g(U_1)K\left(\frac{u-U_1}{h}\right) + \int_0^1 g'(u')K\left(\frac{u-u'}{h}\right)\left(\mathbf{1}_{U_1\leq u'}-u'\right)du'\\ &\quad - \int_0^1 g(u')K\left(\frac{u-u'}{h}\right)du'.\end{aligned}$$

It follows that $\mathbb{E}[T^h_{3,1,2}] \leq 3/(nh^2)\{\mathbb{E}[T^h_{3,1,2,1}]+\mathbb{E}[T^h_{3,1,2,2}]+\mathbb{E}[T^h_{3,1,2,3}]\}$, with

$$\begin{aligned}T^h_{3,1,2,1} &= \int_0^1\left(g(U_1)K\left(\frac{u-U_1}{h}\right)\right)^2du,\\ T^h_{3,1,2,2} &= \int_0^1\left(\int_0^1 g'(u')K\left(\frac{u-u'}{h}\right)\left(\mathbf{1}_{U_1\leq u'}-u'\right)du'\right)^2du,\\ T^h_{3,1,2,3} &= \int_0^1\left(\int_0^1 g(u')K\left(\frac{u-u'}{h}\right)du'\right)^2du.\end{aligned}$$

We now show that the expectations of these three terms are roughly of size $h$. The main argument is that $h\|K_h\|_{L^2(\mathbb{R})} = \|K\|_{L^2(\mathbb{R})}$. Precisely, for the first term,

$$\begin{aligned}\mathbb{E}\left[T^h_{3,1,2,1}\right] &= \int_0^1\int_0^1 g^2(u')K^2\left(\frac{u-u'}{h}\right)du'du,\\ &\leq \int_0^1 g^2(u')h\,\|K\|^2_{L^2(\mathbb{R})}\,du' = h\,\|K\|^2_{L^2(\mathbb{R})}\,\|g\|^2_{L^2((0;1))}.\end{aligned}$$

The second term is bounded by,

$$\begin{aligned}\mathbb{E}\left[T^h_{3,1,2,2}\right] &\leq \int_0^1 \mathbb{E}\left[\int_0^1 K^2\left(\frac{u-u'}{h}\right)\left(\mathbf{1}_{U_1\leq u'}-u'\right)^2\left(g'(u')\right)^2du'\right]du,\\ &= \int_0^1\int_0^1 K^2\left(\frac{u-u'}{h}\right)u'(1-u')\left(g'(u')\right)^2du'du,\\ &\leq \frac{1}{4}\int_0^1\left(\int_0^1 K^2\left(\frac{u-u'}{h}\right)du\right)\left(g'(u')\right)^2du',\\ &\leq h\frac{1}{4}\,\|K\|^2_{L^2(\mathbb{R})}\,\left\|g'\right\|^2_{L^2((0;1))}.\end{aligned}$$

Finally, the same method leads to

$$\mathbb{E}\left[T^h_{3,1,2,2}\right] \leq \int_0^1\int_0^1 g^2(u')K^2\left(\frac{u-u'}{h}\right)du'du \leq h\,\|K\|^2_{L^2(\mathbb{R})}\,\|g\|^2_{L^2((0;1))}.$$

We have proved that

$$\begin{aligned}\mathbb{E}\left[T^h_{3,1,2}\right] &\leq \frac{3}{nh}\,\|K\|^2_{L^2(\mathbb{R})}\left(2\,\|g\|^2_{L^2((0;1))}+\frac{1}{4}\left\|g'\right\|^2_{L^2((0;1))}\right),\\ &= \frac{3}{nh}\,\|K\|^2_{L^2(\mathbb{R})}\left(2\,\|s\|^2_{\phi}+\frac{1}{4}\left\|s'\right\|^2_{L^2(A)}\right).\end{aligned}$$

We gather this inequality with (31):

$$\mathbb{E}\left[T^h_{3,1}\right] \leq \frac{6}{nh}\,\|K\|^2_{L^2(\mathbb{R})}\left(2C_2\|s\|^2_{L^\infty(A)}+2\,\|s\|^2_{\phi}+\left(\frac{1}{4}+3C_2\right)\left\|s'\right\|^2_{L^2(A)}\right).$$

- **Example 4**.

The term to bound is

$$\begin{aligned}\mathbb{E}\left[T^h_{3,1,2}\right] &= \frac{1}{nh^2}\int_0^B \mathrm{Var}\left(\int_0^B X_{-1}\wedge x'\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'\right)\phi(x)dx,\\ &\leq \frac{1}{nh^2}\int_0^B \mathbb{E}\left[\left(\underbrace{\int_0^B X_{-1}\wedge x'\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'}_{J_x}\right)^2\right]\phi(x)dx.\end{aligned}$$

We can split $J_x$ into two terms: $J_x = J_{x,1} + J_{x,2}$, with

$$\begin{aligned}J_{x,1} &= \int_0^B x'\mathbf{1}_{x'\leq X_{-1}}\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx',\\ J_{x,2} &= X_{-1}\int_0^B \mathbf{1}_{x'>X_{-1}}\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)s(x')\phi(x')dx'.\end{aligned}$$

We integrate $J_{x,1}$ by parts:

$$\begin{aligned}
J_{x,1} &= \int_0^{X_{-1}} x's(x')\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\phi(x')dx', \\
&= \left[-x's(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\right]_0^{X_{-1}} + \int_0^{X_{-1}} \left[x's(x')\right]' K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx', \\
&= -X_{-1}s(X_{-1})K\left(\frac{\Phi(x)-\Phi(X_{-1})}{h}\right) + \int_0^{X_{-1}} \left[s(x')+x's'(x')\right] K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx',
\end{aligned}$$

and compute similarly $J_{x,2}$,

$$\begin{aligned}
J_{x,2} &= X_{-1}\int_{X_{-1}}^{B} s(x')\frac{1}{h}K'\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\phi(x')dx', \\
&= X_{-1}\left[-s(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\right]_{X_{-1}}^{B} + X_{-1}\int_{X_{-1}}^{B} s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx', \\
&= X_{-1}s(X_{-1})K\left(\frac{\Phi(x)-\Phi(X_{-1})}{h}\right) - X_{-1}s(B)K\left(\frac{\Phi(x)-\Phi(B)}{h}\right) \\
&\quad + X_{-1}\int_{X_{-1}}^{B} s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'.
\end{aligned}$$

We add the two results:

$$\begin{aligned}
J_x &= -X_{-1}s(B)K\left(\frac{\Phi(x)-\Phi(B)}{h}\right) \\
&\quad + \int_0^{B} \left(\mathbf{1}_{x'>X_{-1}}X_{-1} + x'\mathbf{1}_{x'\le X_{-1}}\right) s'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx' \\
&\quad + \int_0^{X_{-1}} s(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx', \\
&= -X_{-1}s(B)K\left(\frac{\Phi(x)-\Phi(B)}{h}\right) + \int_0^{B} X_{-1}\wedge x's'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx' \\
&\quad + \int_0^{X_{-1}} s(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'.
\end{aligned}$$

It follows that

$$\mathbb{E}\left[T_{3,1,2}^h\right] = \frac{1}{nh^2}\int_0^B \mathbb{E}\left[J_x^2\right]\phi(x)dx \le \frac{3}{nh^2}\left(\mathbb{E}\left[T_{3,1,2,1}^h\right] + \mathbb{E}\left[T_{3,1,2,2}^h\right] + \mathbb{E}\left[T_{3,1,2,3}^h\right]\right)$$

with

$$\begin{aligned}
T_{3,1,2,1}^h &= \int_0^B X_{-1}^2 s^2(B)K^2\left(\frac{\Phi(x)-\Phi(B)}{h}\right)\phi(x)dx, \\
T_{3,1,2,2}^h &= \int_0^B \left(\int_0^B X_{-1}\wedge x's'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\phi(x)dx, \\
T_{3,1,2,3}^h &= \int_0^B \left(\int_0^{X_{-1}} s(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\phi(x)dx.
\end{aligned}$$

For the first term

$$\begin{aligned}
\mathbb{E}\left[T^h_{3,1,2,1}\right] &= \mathbb{E}\left[X^2_{-1}\right] s^2(B) \int_0^B K^2\left(\frac{\Phi(x)-\Phi(B)}{h}\right)\phi(x)dx,\\
&= \mathbb{E}\left[X^2_{-1}\right] s^2(B) \int_0^{\Phi(B)} K^2\left(\frac{u-\Phi(B)}{h}\right)du,\\
&= h\mathbb{E}\left[X^2_{-1}\right] s^2(B) \int_0^{\Phi(B)/h} K^2\left(v-\frac{\Phi(B)}{h}\right)dv,\\
&\leq h\mathbb{E}\left[X^2_{-1}\right] \|s\|^2_{L^\infty(A)}\|K\|^2_{L^2(\mathbb{R})}.
\end{aligned}$$

For the second, we apply the generalized Minkowski Inequality (see Lemma 1.1, p.13 in Tsybakov 2009),

$$\begin{aligned}
\mathbb{E}\left[T^h_{3,1,2,2}\right] &= \mathbb{E}\left[\int_0^B \left(\int_0^B X_{-1}\wedge x's'(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\phi(x)dx\right],\\
&\leq \mathbb{E}\left[\left\{\int_0^B \left(\int_0^B \left(X_{-1}\wedge x's'(x')\right)^2 K^2\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\phi(x)dx\right)^{1/2}dx\right\}^2\right],\\
&= \mathbb{E}\left[\left\{\int_0^B X_{-1}\wedge x'|s'(x')|\left(\int_0^B K^2\left(\frac{\Phi(x)-\Phi(x')}{h}\right)\phi(x)dx\right)^{1/2}dx\right\}^2\right],\\
&\leq \mathbb{E}\left[X^2_{-1}\right]\left\{\int_0^B |s'(x')|\sqrt{h}\|K\|_{L^2(\mathbb{R})}dx'\right\}^2,\\
&= h\mathbb{E}\left[X^2_{-1}\right]\|K\|_{L^2(\mathbb{R})}\|s'\|^2_{L^1(A)}.
\end{aligned}$$

And finally,

$$\begin{aligned}
\mathbb{E}\left[T^h_{3,1,2,3}\right] &= \int_0^B \mathbb{E}\left[\left(\int_0^{X_{-1}} s(x')K\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\right)^2\right]\phi(x)dx,\\
&\leq \mathbb{E}\left[X_{-1}\int_0^B\int_0^{X_{-1}} s^2(x')K^2\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\phi(x)dx\right],\\
&\leq \mathbb{E}\left[X_{-1}\right]\int_0^B\int_0^B s^2(x')K^2\left(\frac{\Phi(x)-\Phi(x')}{h}\right)dx'\phi(x)dx,\\
&\leq h\mathbb{E}\left[X_{-1}\right]\|K\|^2_{L^2(\mathbb{R})}\|s\|^2_{L^2(A)}.
\end{aligned}$$

At the end,

$$\mathbb{E}\left[T^h_{3,1,2}\right] \leq \frac{3}{nh}\|K\|^2_{L^2(\mathbb{R})}\left(\mathbb{E}\left[X^2_{-1}\right]\|s\|^2_{L^\infty(A)} + \mathbb{E}\left[X^2_{-1}\right]\|s'\|^2_{L^1(A)} + \mathbb{E}\left[X_{-1}\right]\|s\|^2_{L^2(A)}\right).$$

We gather this inequality with (31), like for the first three examples:

$$\begin{aligned}
\mathbb{E}\left[T^h_{3,1}\right] \leq \frac{6}{nh}\|K\|^2_{L^2(\mathbb{R})}\Big(&\|s\|^2_{L^\infty(A)}(\mathbb{E}[X^2_{-1}]+2C_2B^2)\\
&+\|s'\|^2_{L^1(A)}(\mathbb{E}[X^2_{-1}]+C_2B^2)+\mathbb{E}[X_{-1}]\|s\|^2_{L^2(A)}\Big).
\end{aligned}$$

The upper-bound of $T^h_{3,1}$ is thus completed in each of the examples.

- **Upper-bound for $\mathbf{T^h_{3,2}}$.**

We again split the term: $T_{3,2}^h \leq 4T_{3,2,1}^h + 4T_{3,2,2}^h$, with

$$T_{3,2,1}^h = \int_A \left( \mathbb{E}\left[ \frac{1}{h}\theta(Y_1) \left\{ \frac{\left(\hat{\Phi}_n(x) - \Phi(x)\right)^2}{h^2} K''\left(\hat{\alpha}_{n,x,h}\right) \right\} |(X_{-i}) \right] \right)^2 \phi(x)dx,$$

$$T_{3,2,2}^h = \int_A \left( \mathbb{E}\left[ \frac{1}{h}\theta(Y_1) \left\{ \frac{\left(\hat{\Phi}_n(X_1) - \Phi(X_1)\right)^2}{h^2} K''\left(\hat{\alpha}_{n,x,h}\right) \right\} |(X_{-i}) \right] \right)^2 \phi(x)dx.$$

The two terms can be bounded in the same way. We detail the computation for the first one:

$$\begin{aligned} T_{3,2,1}^h &\leq \frac{1}{h^6}\int_A \mathbb{E}\left[\theta^2(Y_1)\left\{\left(\hat{\Phi}_n(x) - \Phi(x)\right)^4 \left(K''\left(\hat{\alpha}_{n,x,h}\right)\right)^2\right\}|(X_{-i})\right]\phi(x)dx, \\ &\leq \frac{1}{h^6}\int_A \mathbb{E}\left[\theta^2(Y_1)\,|(X_{-i})\right] \left\|\hat{\Phi}_n - \Phi\right\|_{L^\infty(A)}^4 \left\|K''\right\|_{L^\infty(\mathbb{R})}^2 \phi(x)dx, \\ &= \frac{1}{h^6}\int_A \phi(x)dx\mathbb{E}\left[\theta^2(Y_1)\right] \left\|\hat{\Phi}_n - \Phi\right\|_{L^\infty(A)}^4 \left\|K''\right\|_{L^\infty(\mathbb{R})}^2. \end{aligned}$$

The conclusion is analogous as the one of the proof of Lemma 12 (see Section 7.4.2). We obtain

$$\mathbb{E}\left[T_{3,2,1}^h\right] \leq \kappa_1 \mathbb{E}\left[\theta^2(Y_1)\right] \left\|K''\right\|_{L^\infty(\mathbb{R})}^2 C_4 \frac{1}{n^2h^6},$$

where $\kappa_1 = 1$ for Examples 1-3, and $\kappa_1 = B^4\mathbb{E}[X]$ in Example 4. The same bound holds for $T_{3,2,2}^h$. At the end,

$$\mathbb{E}\left[T_{3,2}^h\right] \leq 8\kappa_1 \mathbb{E}\left[\theta^2(Y_1)\right] \left\|K''\right\|_{L^\infty(\mathbb{R})}^2 C_4 \frac{1}{n^2h^6}.$$

□

## 8. Appendix 2: generalization of the proof of Lemma 5

8.1. **Objective.** Recall that Lemma 5 gives a bound of size $1/n$ for

$$\mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)} \nu_{n,h}^2(t) - \tilde{V}(h)\right)_+\right],$$

with, for any $t \in \bar{S}(0,1)$ (a subset of $\tilde{S}(0,1) = \{t \in L^1(\Phi(A)) \cap L^2(\Phi(A)),\ \|t\|_{L^2(\Phi(A))} = 1\}$)

$$\nu_{n,h}(t) = \frac{1}{n}\sum_{i=1}^n \int_{\Phi(A)} \left\{\theta(Y_i)K_{h'}\left(u - \Phi(X_i)\right) - \mathbb{E}\left[\theta(Y_i)K_{h'}\left(u - \Phi(X_i)\right)\right]\right\} t(u)du,$$

and where $\tilde{V}(h) = \delta'\|K\|_{L^2(\mathbb{R})}\mathbb{E}[\theta(Y_1)^2]/(nh)$ for a numerical $\delta' > 0$. We have provided in Section 6.6.1 a proof which depends on the different considered examples, mainly under Assumption (H1) ($s$ is bounded). The aim of this section is to show that a general proof is possible with (H1) replaced by the following slightly stronger assumption:

(H1$_{bis}$) (i) The conditional expectation $\mathbb{E}[\theta(Y)|X]$ is bounded,
(ii) The ratio $f_X/\Phi$ is bounded.

To see that (H1$_{bis}$) is more restrictive than (H1), one must recall that the target function $s$ can be defined by (4): $s = (f_X/\phi)\mathbb{E}[\theta(Y)|X]$.
Notice that Assumption (i) is automatically fulfilled in Example 3-4, since $\theta(Y) = Y$, with $Y$ bounded by 1. Moreover, if $\Phi = F_X$, the ratio $f_X/\phi$ equals 1 and (ii) is satisfied.

Under ($\mathrm{H1}_{bis}$), we give a proof which embraces Examples 1-4 and moreover, permits to handle all the statistical settings covered by the formula (4). In the sequel, $s_0(X) := \mathbb{E}[\theta(Y)|X]$.

8.2. **Generalized proof.** We begin with the following splitting of the empirical process: $\nu_{n,h} = \nu_{n,h}^{(1b)} + \nu_{n,h}^{(2,1b)} + \nu_{n,h}^{(2,2b)}$, where $\nu_{n,h}^{(l)}$ has the form

$$\nu_{n,h}^{(l)}(t) = \frac{1}{n}\sum_{i=1}^{n} \xi_{t,h}^{(l)} - \mathbb{E}\left[\xi_{t,h}^{(l)}\right], \quad l \in \{(1b), (2,1b), (2,2b)\},$$

with, for $\kappa_n^b = c^b\sqrt{n}/\ln(n)$ ($c^b > 0$),

$$\begin{aligned}
\xi_{t,h}^{(1b)}(X_i) &= s_0(X_i)\int_{\Phi(A)} K_{h'}\left(u - \Phi(X_i)\right) t(u)du, \\
\xi_{t,h}^{(2,1b)}(X_i, Y_i) &= \left(\theta(Y_i) - s_0(X_i)\right)\mathbf{1}_{|\theta(Y_i)-s_0(X_i)|\leq\kappa_n^b}\int_{\Phi(A)} K_{h'}\left(u - \Phi(X_i)\right) t(u)du, \\
\xi_{t,h}^{(2,2b)}(X_i, Y_i) &= \left(\theta(Y_i) - s_0(X_i)\right)\mathbf{1}_{|\theta(Y_i)-s_0(X_i)|>\kappa_n^b}\int_{\Phi(A)} K_{h'}\left(u - \Phi(X_i)\right) t(u)du.
\end{aligned}$$

The guideline is to apply the Talagrand Inequality (Lemma 6) to bound the first two empirical processes $\nu_{n,h}^{(1b)}$ and $\nu_{n,h}^{(2,1b)}$, and bound the third one roughly. Details can be found in the three following sections, in which we often refer to Section 6.6.1, since the computations are similar. Finally, gathering the three upper bounds of size $1/n$ (up to constant) permits to complete the proof of Lemma 5.

**Remark 2.** Clearly, if $\theta$ is bounded, the proof can be greatly simplified, since there is no use in splitting the empirical process in this case. However, we choose to give the most general formulation.

8.2.1. *Upper-bound for* $\nu_{n,h}^{(1b)}$. To apply the concentration results of Lemma 6, we first need to compute appropriate values for the bounds $M_1^{(1b)}$, $H^{(1b)}$ and $v^{(1b)}$. The computation for the first quantity is totally similar to the one done for $M_1^{(1)}$ in Section 6.6.1, and leads to $M_1^{(1b)} = \|s_0\|_{L^\infty(A)}\|K\|_{L^2(\mathbb{R})}\sqrt{h}$. In an analogous way, we first obtain

$$\mathbb{E}\left[\sup_{t\in\bar{S}(0,1)}\left(\nu_{n,h}^{(1b)}(t)\right)^2\right] \quad \leq \quad \|K\|_{L^2(\mathbb{R})}^2\frac{\mathbb{E}[s_0^2(X_1)]}{nh}.$$

But, $\mathbb{E}[s_0^2(X_1)] = \mathbb{E}[\mathbb{E}[\theta(Y_1)|X_1]^2] \leq \mathbb{E}[\theta^2(Y_1)]$. Thus, we set $(H_1^{(1b)})^2 = \mathbb{E}[\theta^2(Y_1)]\|K\|_{L^2(\mathbb{R})}^2/nh$. Finally, $v^{(1b)}$ is a bound for the following variance:

$$\begin{aligned}
\mathrm{Var}\left(\xi_{t,h}^{(1)}(X_1)\right) &\leq \mathbb{E}\left[s_0^2(X_1)\left(\int_{\Phi(A)} K_h(u - \Phi(X_1))t(u)du\right)^2\right], \\
&\leq \|s_0\|_{L^\infty(A)}^2\mathbb{E}\left[\left(\int_{\Phi(A)} K_h(u - \Phi(X_1))t(u)du\right)^2\right].
\end{aligned}$$

The difference between this computation and the one of $v^{(1)}$ in Section 6.6.1 is the way to deal with the expectation:

$$\begin{aligned}\mathbb{E}\left[\left(\int_{\Phi(A)} K_h(u-\Phi(X_1))t(u)du\right)^2\right] &= \mathbb{E}\left[\left(\check{K}_h \star t\mathbf{1}_{\Phi(A)}\right)^2(\Phi(X_1))\right],\\ &= \int_A \left(\check{K}_h \star t\mathbf{1}_{\Phi(A)}\right)^2(\Phi(x))\, f_X(x)dx,\\ &\le \left\|\frac{f_X}{\phi}\right\|_{L^\infty(A)}\int_A \left(\check{K}_h \star t\mathbf{1}_{\Phi(A)}\right)^2(\Phi(x))\,\phi(x)dx,\end{aligned}$$

thanks to Assumption (H1$_{bis}$) (ii). The change of variables $u=\Phi(x)$ leads to

$$\begin{aligned}\mathbb{E}\left[\left(\int_{\Phi(A)} K_h(u-\Phi(X_i))t(u)du\right)^2\right] &\le \left\|\frac{f_X}{\phi}\right\|_{L^\infty(A)}\int_{\Phi(A)} \left(\check{K}_h \star t\mathbf{1}_{\Phi(A)}\right)^2(u)du,\\ &= \left\|\frac{f_X}{\phi}\right\|_{L^\infty(A)}\left\|\check{K}_h \star t\mathbf{1}_{\Phi(A)}\right\|^2_{L^2(\mathbb{R})}. \qquad (32)\end{aligned}$$

We end the computation in the same way as previously (see $v^{(1)}$, Section 6.6.1), and then $v^{(1b)}=\|f_X/\phi\|_{L^\infty(A)}\|K\|_{L^1(\mathbb{R})}$. Now, applying the Talagrand Inequality is straightforward, and, like in Section 6.6.1, thanks to Assumptions (H2)-(H3), we obtain

$$\sum_{h\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)}\left(\nu^{(1b)}_{n,h}(t)\right)^2-2(1+2\delta)\|K\|^2_{L^2(\mathbb{R})}\mathbb{E}[\theta^2(X_1)]\frac{1}{nh}\right)_+\right]\le\frac{C}{n}.$$

8.2.2. *Upper-bound for $\nu^{(2,1b)}_{n,h}$.* We also precise the quantities which permit to apply the concentration result. Their computations are also inspired by Section 6.6.1, and precisely by the bound for $\nu^{(2,1)}_{n,h}$. To begin, we choose $M_1^{(2b)}=\kappa_n^b\|K\|_{L^2(\mathbb{R})}\sqrt{h}$. For $H^{(2b)}$, similarly to $H^{(2)}$, we first obtain

$$\mathbb{E}\left[\sup_{t\in\bar{S}(0,1)}\left(\nu^{(1)}_{n,h}(t)\right)^2\right] \le \frac{1}{n}\mathbb{E}\left[(\theta(Y_1)-s_0(X_1))^2\int_{\Phi(A)}K_h^2(u-\Phi(X_1))du\right]$$

But, $\int_{\Phi(A)}K_h^2(u-\Phi(X_1))du\le\|K_h\|^2_{L^2(\mathbb{R})}$, and

$$\mathbb{E}\left[(\theta(Y_1)-s_0(X_1))^2\right]=\mathbb{E}\left[\mathrm{Var}(\theta(Y_1)|X_1)\right]\le\mathbb{E}[\theta^2(Y_1)]. \qquad (33)$$

Therefore, we set $(H_1^{(2b)})^2=\mathbb{E}[\theta^2(Y_1)]\|K\|^2_{L^2(\mathbb{R})}/nh$. Finally,

$$\begin{aligned}\mathrm{Var}\left(\xi^{(2,1b)}_{t,h}(X_1,Y_1)\right) &\le \mathbb{E}\left[(\theta(Y_1)-s_0(X_1))^2\left(\int_{\Phi(A)}K_h(u-\Phi(X_i))t(u)du\right)^2\right],\\ &= \mathbb{E}\left[\mathbb{E}\left[(\theta(Y_1)-s_0(X_1))^2|X_1\right]\left(\int_{\Phi(A)}K_h(u-\Phi(X_i))t(u)du\right)^2\right],\\ &\le \mathbb{E}\left[\mathbb{E}[\theta^2(Y_1)]\left(\int_{\Phi(A)}K_h(u-\Phi(X_i))t(u)du\right)^2\right],\end{aligned}$$

by using (33). It remains to recall that (32) holds, to set $v^{(2b)} = \mathbb{E}[\theta^2(Y_1)]\|K\|_{L^1(\mathbb{R})}$.

Once again, the existence and size of these three quantities justify that we obtain

$$\sum_{h\in\mathcal{H}_n} \mathbb{E}\left[\left(\sup_{t\in\bar{S}(0,1)}\left(\nu_{n,h}^{(2,1b)}(t)\right)^2 - 2(1+2\delta)\|K\|^2_{L^2(\mathbb{R})}\mathbb{E}[\theta^2(X_1)]\frac{1}{nh}\right)_+\right] \leq \frac{C}{n}.$$

8.2.3. *Upper-bound for $\nu_{n,h}^{(2,2b)}$.* We exactly follow the same line as for $\nu_{n,h}^{(2,2)}$ (see Section 6.6.1) to write

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in\tilde{S}(0,1)}\left(\nu_{n,h}^{(2,2b)}(t)\right)^2\right] &\leq \frac{1}{n}\mathbb{E}\left[(\theta(Y_1)-s_0(X_1))^2\,\mathbf{1}_{\{|\theta(Y_1)-s_0(X_1)|>\kappa_n^b\}}\int_{\Phi(A)} K_h^2\left(u-\Phi(X_1)\right)du\right],\\
&\leq \frac{\|K\|_{L^2(\mathbb{R})}}{nh}\mathbb{E}\left[(\theta(Y_1)-s_0(X_1))^2\,\mathbf{1}_{\{|\theta(Y_1)-s_0(X_1)|>\kappa_n^b\}}\right],\\
&\leq \|K\|_{L^2(\mathbb{R})}\frac{(\kappa_n^b)^{-p}}{nh}\mathbb{E}\left[|\theta(Y_1)-s_0(X_1)|^{2+p}\right],
\end{aligned}$$

with $p$ defined by (H5). As previously, the way to conclude can be found in Section 6.6.1,

$$\sum_{h\in\mathcal{H}_n} \mathbb{E}\left[\sup_{t\in\bar{S}(0,1)}\left(\nu_{n,h}^{(2,2b)}(t)\right)^2\right] \leq \frac{C}{n}.$$

□

## Acknowledgements

I would like to thank Fabienne Comte for a wealth of smart advice and carefully readings along this work. I am also grateful to Valentine Genon-Catalot for proofreading earlier version of this paper.

## References

Nathalie Akakpo and Cécile Durot. Histogram selection for possibly censored data. *Math. Methods Statist.*, 19(3):189–218, 2010.

Yannick Baraud. Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6: 127–146 (electronic), 2002. ISSN 1292-8100.

Elodie Brunel and Fabienne Comte. Penalized contrast estimation of density and hazard rate with censored data. *Sankhyā*, 67(3):441–475, 2005.

Elodie Brunel and Fabienne Comte. Adaptive estimation of hazard rate with censored data. *Comm. Statist. Theory Methods*, 37(8-10):1284–1305, 2008.

Elodie Brunel and Fabienne Comte. Cumulative distribution function estimation under interval censoring case 1. *Electron. J. Stat.*, 3:1–24, 2009.

Gaëlle Chagny. Penalization versus goldenshluger-lepski strategies in warped bases regression. *ESAIM Probab. Statist.*, to appear (available online):1–24, 2011.

Gaëlle Chagny. Warped bases for conditional density estimation. *Submitted, hal-00641560, v2*, 3:1–24, 2012.

Christophe Chesneau. A maxiset approach of a Gaussian noise model. *TEST*, 16(3):523–546, 2007.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, 27:642–669, 1956.

Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011.

Nicholas P. Jewell and Mark van der Laan. Current status data: review, recent developments and open problems. In *Advances in survival analysis*, volume 23 of *Handbook of Statist.*, pages 625–642. Elsevier, Amsterdam, 2004.

Gérard Kerkyacharian and Dominique Picard. Regression in random design and warped wavelets. *Bernoulli*, 10(6):1053–1105, 2004.

Thierry Klein and Emmanuel Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.

Claire Lacour. Adaptive estimation of the transition density of a particular hidden Markov chain. *J. Multivariate Anal.*, 99(5):787–814, 2008.

Shuangge Ma and Michael R. Kosorok. Adaptive penalized $M$-estimation with current status data. *Ann. Inst. Statist. Math.*, 58(3):511–526, 2006.

K. L. Mehra, Y. S. Ramakrishnaiah, and P. Sashikala. Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode. *Ann. Inst. Statist. Math.*, 52(4): 630–645, 2000.

Hans-Georg Müller and Jane-Ling Wang. Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50(1):61–76, 1994.

Elizbar Nadaraya. On estimating regression. *Theory of Probability and its Application*, 9(4): 141–142, 1964.

Prakash N. Patil. Bandwidth choice for nonparametric hazard rate estimation. *J. Statist. Plann. Inference*, 35(1):15–30, 1993.

Marianna Penskaya. Mean square consistent estimation of a ratio. *Scand. J. Statist.*, 22(1): 129–137, 1995.

Thanh Mai Pham Ngoc. Regression in random design and Bayesian warped wavelets estimators. *Electron. J. Stat.*, 3:1084–1112, 2009.

Sandra Plancade. Estimation of conditional cumulative distribution function from current status data. *Submitted arXiv:1110.5927*, 69(4):734–763, 2011. ISSN 0972-7671.

Patricia Reynaud-Bouret. Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, 12(4):633–661, 2006.

Winfried Stute. Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.*, 12(3):917–926, 1984.

Winfried Stute. Conditional empirical processes. *Ann. Statist.*, 14(2):638–647, 1986.

Martin A. Tanner and Wing Hung Wong. The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.*, 11(3):989–993, 1983.

Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Geoffrey S. Watson. Smooth regression analysis. *Sankhyā Ser. A.*, 26(1):359–372, 1964.

Shie-Shien Yang. Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function. *J. Amer. Statist. Assoc.*, 76(375):658–662, 1981.